Picking Funds with Confidence

Niels S. Grønborg
Aarhus BSS and CREATES

Asger Lunde
Aarhus BSS and CREATES

Allan Timmermann
UCSD

Russ Wermers
University of Maryland

December 2, 2015

Abstract

We present a new approach to select the set of mutual funds with superior performance, i.e., funds whose performance is not dominated by that of any other funds. The approach eliminates funds with unpredictable or inferior performance through a sequence of pair-wise comparisons that determines both the identity and number of superior mutual funds. Empirically, we find that funds identified as being superior go on to earn substantially higher risk-adjusted returns than top funds identified by conventional ranking methods. Moreover, the size of the set of superior funds fluctuates across economic states (being wider during economic expansions) and can, at times, be very narrow, suggesting that the approach has the ability to discriminate between the funds at the top end of the cross-sectional performance distribution.

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

JEL codes: G2, G11, G17

1. Introduction

The ability to identify skill among mutual funds poses an important economic challenge: As of 2014, $16 trillion was invested in U.S. mutual funds and a large industry of investment advisors and consultants were engaged in advising retail and institutional clients on how to select funds (Blake et al. (2013) and Jenkinson et al. (2014)). While there is broad consensus in the academic literature that some funds consistently underperform due to high trading costs and fees, empirical evidence suggests that few mutual funds manage to outperform their benchmarks on a consistent basis (Carhart (1997)). There are good economic reasons why the ability of individual funds to outperform has proven difficult to predict: First, estimates of individual funds' risk-adjusted returns tend to be surrounded by large sampling errors and so standard approaches to identifying superior performance ("alpha") tend to have weak power. Second, fund managers' ability to outperform may be short-lived because evidence of successful strategies is bound to attract more competition (Hoberg et al. (2015)). Third, active fund management has been found to be subject to diseconomies of scale and good past performance attracts higher inflows which in turn leads to deteriorating performance.[1] Finally, the nature of a fund manager's information–and the manager's strategies for acting on such information–could depend on the state of the economy which itself evolves and so leads to changes in the set of funds that can outperform.[2]

This paper introduces a new approach to address whether we can (ex-ante) identify a set of funds with superior performance and, if so, how wide the selected set of funds is, which funds get included, what types of investment strategies they adopt, and how this set of funds (along with their risk-adjusted performance) varies over time. Superior funds are funds whose performance is not dominated by any other funds. Identifying the set of "best" (or superior) funds therefore 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 conduct a large set of pairwise comparisons of

[1] Glode et al. (2011) present a simple 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. See also Berk and Green (2004) for a theoretical model that implies vanishing fund alphas.

[2] Mamaysky et al. (2008) develop a model in which managers observe private information signals which revert towards being uninformative.

all funds in existence to eliminate any funds whose performance is dominated by at least one other fund.

Conventional approaches in the finance literature are not well designed to handle such comparisons, nor do they control the "size" of the test, i.e., the probability of wrongly eliminating truly superior funds. To deal with such issues, our analysis adopts a new approach for selection of mutual funds that makes use of the Model Confidence Set (MCS) methodology of Hansen et al. (2011) which is designed to select the most accurate prediction models from a large set of candidate models. Hansen et al. (2011) show that a step-wise bootstrap approach can be used to determine critical values for elimination of models from the MCS in a way that controls the probability of correctly identifying models with superior predictive performance and, conversely, eliminates truly inferior models for sure.

In our context the set of candidate models is the list of mutual funds in existence at a given point in time and their performance is measured through the funds' risk-adjusted returns.[3] The approach undertakes a series of pair-wise tests to sequentially eliminate funds with inferior performance. If at least one fund with significantly inferior performance can be identified, the fund with the "worst" performance is eliminated and the elimination process is repeated on the reduced set of funds. The procedure continues until no further funds with inferior performance can be identified and eliminated. We label the set of funds remaining at the end, i.e., the funds identified to have superior performance, as the Fund Confidence Set (FCS).

The FCS approach uses estimates of individual funds' risk-adjusted returns as a way to rank and compare their performance and thus the results from applying this methodology will depend on how good the underlying performance model is at extracting information about fund performance. We apply the FCS approach to three different performance models which assume (i) constant (or slowly evolving) alphas; (ii) time-varying alphas using the latent skill approach of Mamaysky et al. (2007); (iii) time-varying alphas with alpha estimates extracted using both fund returns and holdings data. For each performance model we find that the FCS approach can be used to select funds and form portfolios of funds with considerably higher performance than portfolios containing

[3]Another important difference between our approach and the MCS is that we propose an elimination rule that excludes funds with poor or unpredictable performance—either is detrimental from an investment perspective.

a fixed proportion (e.g., 5% or 10%) of top ranked funds. The performance results are particularly strong for the performance measurement model that combines data on returns and holdings to obtain a sharper estimate of fund alphas. Using this model, an equal-weighted portfolio of funds included in the top FCS generates a four-factor alpha exceeding 50 bps/month which is highly statistically significant; for comparison, the four-factor alpha estimate of the top decile of funds is around 2 bps/month.

Empirically, we find that the FCS approach can be used to identify a narrow set of funds with superior performance. In fact, because many funds' performance is estimated with large sampling errors, the approach is found to work best if we use a relative stringent criterion for inclusion of funds in the FCS, resulting in a reduced probability of wrongly including funds with inferior performance. The set of funds selected by the approach fluctuates considerably over time. Sometimes only a single fund gets selected for a long stretch of time; at other times the approach identifies a much broader set of funds as being superior. Moreover, the breadth of the set of funds with superior performance is found to be significantly correlated with a range of macroeconomic state variables as the fraction of superior funds is larger in expansions and smaller in recessions. This suggests that funds' ability to outperform is state dependent, consistent with recent findings of Kacperczyk et al. (2014).

Our analysis of holdings data for funds in the top FCS suggests that superior funds change their industry concentration and shift their risk loadings significantly over time. For example, the top FCS funds tilt towards growth stocks from 1994 to 2003 and, again from 2009 to 2011, overweighting instead value stocks from 2003 to 2009. Superior funds overweight small stocks from 2001 to 2003 and from 2009-2012 but over-weighted large caps from 2005 to 2008. Conversely, superior funds have above-average exposures to the momentum factor only during brief spells. Funds in the top FCS overweight computer and electronic equipment stocks after 2005 and also overweight business services and machinery during shorter spells. Conversely, the top-rated funds underweight retail stocks and, in particular, banking stocks throughout most of the sample.

The FCS methodology can also be used to successfully identify funds with inferior performance. When applied to select inferior funds, we find that the set of "worst" funds is somewhat wider than the equivalent set of superior funds. This reflects the greater persistence of factors giving rise to underperformance such as high trading costs and management fees. Again, we find that the funds in the bottom FCS portfolio produce substantially worse performance than a portfolio consisting of a fixed proportion of alpha-ranked funds such as the bottom 5-10% of funds.

Our empirical results are related to several findings in recent studies on mutual fund performance. Carhart (1997) finds that the performance of top-ranked funds reverts towards the mean after about one year. Using daily mutual fund returns, Bollen and Busse (2004) find abnormal performance that lasts for one quarter, but disappears at longer horizons. Glode et al. (2011) find evidence of predictability of mutual fund returns following periods of high market returns, while such predictability is weaker after periods with low market returns. Avramov and Wermers (2006) and Banegas et al. (2013) find evidence of predictability of mutual fund manager skills that depend on persistent variables tracking the state of the macroeconomy. Kacperczyk et al. (2014) find that a subset of managers possess stock picking (but not market timing) skills in economic booms, while conversely possessing market timing (but not stock picking) skills in recessions. These results are consistent with our finding that the set of funds with the ability to outperform ("positive alpha") is highly time varying and that the type of skill that is associated with superior performance depends on the state of the economy.

Our analysis differs from previous studies in several important dimensions. Kosowski et al. (2006) ask whether there exists star fund managers, i.e., if the single best manager can outperform some benchmark, such as the four-factor model of Carhart (1997), or, alternatively, if some predetermined fraction of funds, such as the top 10% of funds, can outperform. However, their methodology cannot be used to endogenously determine the size of the set of funds with superior performance or the identify of the individual funds. In fact, a strategy of only investing in the fund whose alpha is deemed highest can sometimes backfire because such a fund might have been "lucky" and its performance could reflect very high idiosyncratic risk taking. Barras et al. (2010) develop an approach for controlling for funds that have high alpha estimates due to "luck" and so identify the set of funds which truly have positive alphas. However, unlike us, they do not address whether funds deemed to have positive alphas are equally good. This is a highly relevant question from an investor's perspective because not all funds, even those with positive alphas, need to perform equally well.[4] Interestingly, Barras et al. (2010) find evidence that the set of funds with positive alphas has been shrinking over time, making it more difficult to identify truly superior fund managers. In fact, our results show that the set of funds with superior performance at times only consists of a single fund.

It is also worth contrasting our approach to identifying superior funds with the conventional "portfolio sorting" approach used in most finance studies. The latter approach ranks funds by their expected alphas, assigns the funds into decile portfolios, repeats the sorting at regular rebalancing points, followed by inspection of the portfolios' subsequent risk-adjusted return performance. Several limitations restrict the usefulness of this decile sorting approach. First, due to competitive pressures (e.g., Hoberg et al. (2015)) and state-dependence in skills (e.g., Ferson and Schadt (1996)), we would expect the proportion of funds that can outperform to vary over time and across economic states.[5] Imagine that a method for ranking mutual funds has the ability to correctly identify a set of mutual funds with superior performance, but that less than 10% of funds on average are identified as having positive alphas. By focusing on the top 10% of alpha-ranked funds, the traditional decile sorting approach is likely to mix truly superior funds with inferior funds and thus add noise. This problem is only exacerbated if the proportion of funds identified as outperformers is time-varying and sometimes exceeds 10% (in which case the top decile portfolio is too narrow), at other times is smaller (in which case it is too wide). As we show empirically, an approach (such as ours) that endogenously determines the set of funds expected to produce superior performance can be far better at identifying future outperformance than the conventional portfolio sorting approach.

The outline of the paper is as follows. Section 2 introduces the fund confidence set methodology, while Section 3 describes our data and the models 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

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

[5]Pastor, Stambaugh and Taylor (2015) find strong evidence of decreasing returns to scale for active mutual fund managers at the industry level. Their estimates suggest that active managers have become more skilled over time, although this has not translated into better fund performance due to the increase in the size of the fund management industry.

through time. Section 6 performs an attribution results which decomposes the performance of the FCS funds using holdings data on industry concentrations and stock-level estimates of individual funds' exposures to common risk factors. Section 7 concludes.

2. The Fund Confidence Set

A large empirical literature in finance explores whether it is possible to *ex ante* identify funds with superior risk-adjusted performance. The most widespread practice used in the literature is to, first, rank individual funds based on their expected (predicted) performance then, second, form decile portfolios based on such rankings and, finally, track the portfolios' subsequent risk-adjusted performance.[6] While the practice of allocating individual funds into decile portfolios is simple to perform and intuitive to interpret, it only addresses whether the *average* risk-adjusted performance of the top 10% of funds (or a similar proportion) is positive. This question is very different from the more relevant and interesting question of whether we can identify a set of funds with superior performance. For example, if the set of funds capable of producing positive risk-adjusted performance varies over time and is sometimes far narrower than 10%, then we may well find empirically that the top 10% of funds do not generate positive performance on average even though there exists a set of funds with positive (ex ante) risk-adjusted performance.

This section introduces an alternative approach that does not fix the proportion of funds deemed capable of delivering superior performance but, rather, determines this endogenously as part of the process used to estimate individual funds' performance. We first describe the approach in broad terms and characterize its properties, before providing details on how we implement the approach on our mutual fund data.

2.1. Methodology

For a fund to be attractive to investors it must have a high expected risk-adjusted performance. This requires that the fund's performance is at least modestly predictable. To see the importance

[6]The practice of studying the performance of individual assets grouped into portfolios can be viewed as an alternative to using rank correlations or other measures of performance based on the returns of individual assets or funds. Patton and Timmermann (2010) propose nonparametric ranking tests based on the time series of returns on single- or double-sorted portfolios.

of this point, consider a fund that has produced a high average risk-adjusted performance because it generated a very high return during a single period. This would not instill much confidence in the fund's ability to produce high future returns. Contrast this with another fund that consistently performs well; this fund might be attractive to investors, particularly if the periods when it outperforms can be predicted ahead of time.

Our objective is to identify funds which we can predict with some confidence will produce positive future risk-adjusted returns. Following common practice, we compute a fund's risk-adjusted return by adjusting the fund's returns, net of the T-bill rate, $R_{i,t}$, for its exposure to a set of risk factors, $\mathbf{z}_t$:

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

Here i refers to the fund and t refers to the time period; $\varepsilon_{i,t}$ is the fund's idiosyncratic return, $\beta_{i,t}$ measures the fund's exposure to the common risk factors, while $\alpha_{i,t}$ measures its risk-adjusted (abnormal) performance, often referred to as the fund's 'alpha'.

The model in equation (1) is quite general as it allows both α_i and β_i to vary over time. If a fund's alpha is constant over time, i.e., $\alpha_{i,t} = \alpha_i$, 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, $\alpha_{i,t} \neq \alpha_{i,s}$ for $s \neq t$, we need to model how $\alpha_{i,t}$ changes over time. In both cases, let $\hat{\alpha}_{i,t+1|t}$ denote the expected value of fund i's alpha in period $t+1$ based on observations available at time t.

Given such an estimate, we next need to measure if $\hat{\alpha}_{i,t+1|t}$ has been good at predicting whether the fund subsequently outperformed. To this end, we consider the product of $Max(\hat{\alpha}_{i,t+1|t}, 0)$ and the sign of the fund's actual risk-adjusted performance, $sign(R_{i,t+1} - \hat{\beta}'_{i,t+1} z_{t+1})$:

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

Here the sign function $sign(\bullet)$ equals +1 if the argument is positive and it is -1 if the argument is negative or zero. To motivate this objective function, note that the expression in equation (2) will be large for funds with large, positive predicted alphas whose subsequent risk-adjusted returns were positive. Conversely, the objective in (2) penalizes funds for which we predicted a positive alpha

($\hat{\alpha}_{j,t+1|t} > 0$), but whose subsequent risk-adjusted returns were negative ($R_{i,t+1} - \beta'_{i,t+1}\mathbf{z}_{t+1} < 0$). Thus, the "predictive alpha" measure in (2) accounts for both the magnitude (and sign) of the predicted performance (through $Max(\hat{\alpha}_{i,t+1|t}, 0)$) and for the success of the forecast (through the product of the two terms). Finally, note that funds with negative predicted alphas ($\hat{\alpha}_{i,t+1|t} < 0$) get excluded from consideration since $P_{i,t+1} = 0$ for such funds and so (2) is useful for identifying superior performance.[7] We explain how to identify inferior performance below.

To help with selecting funds with the highest expected value of the predictive alpha measure we compute the sample estimate of the average value of (2):

$$\bar{P}_{i,t} = \frac{1}{t-t_0}\sum_{\tau=t_0}^{t} P_{i,\tau} = \frac{1}{t-t_0}\sum_{\tau=t_0}^{t} Max(\hat{\alpha}_{i,\tau|\tau-1}, 0) sign(R_{i,\tau} - \hat{\beta}'_{i,\tau}\mathbf{z}_{\tau}), \tag{3}$$

where $\hat{\alpha}_{i,\tau|\tau-1}$ is the forecast of $\alpha_{i,\tau}$ based on information available in period $\tau-1$ and $\hat{\beta}'_{i,\tau}$ are least-squares estimates of β'_i, using only data up to time τ, so that the estimate $\bar{P}_{i,t}$ can be computed at time t. t_0 is the starting point of the sample used to estimate $\bar{P}_{i,t}$.

Our data contain more than 2,000 funds whose performance needs to be pair-wise compared at each point in time. This introduces a complicated multiple hypothesis testing problem which we address by applying the model confidence set (MCS) approach of Hansen et al. (2011).

The Model Confidence Set (MCS) of Hansen et al. (2011) is designed to choose the set of "best" forecasting models from a larger set of candidate models. Because the approach is developed for selection of forecasting models, we need to modify it to our setting. Most obviously, the object of interest in our analysis is not a model, but a fund and so we label our approach the Fund Confidence Set (FCS). We next describe how the approach works.

Our goal is to select a set of funds which, at a certain level of confidence, contains the best fund–or set of funds if multiple funds are believed to have identical performance. The approach relies on an equivalence test and an elimination rule. Let $\mathcal{F}_t^0 = \{F_{1t}, ..., F_{nt}\}$ be the initial set of funds under consideration at time t and let

$$\hat{P}_{i,t} = Max(\hat{\alpha}_{i,t|t-1}, 0) sign(R_{i,t} - \hat{\beta}'_{i,t-1}\mathbf{z}_t) \tag{4}$$

[7]This assumes that at least one fund has a positive expected value $\hat{\alpha}_{i,t+1|t} > 0$.

be the estimated performance of fund i in period t. The difference between the performance of funds i and j at time t is then

$$d_{ij,t} = P_{i,t} - P_{j,t}\,, \quad i, j \in \mathcal{F}_t^0. \tag{5}$$

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

$$H_{0,\mathcal{F}_t} : \mu_{ij} = 0, \text{ for all } i, j \in \mathcal{F}_t \subset \mathcal{F}_t^0, \tag{6}$$

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

$$H_{A,\mathcal{F}_t} : \mu_{ij} \neq 0 \text{ for some } i, j \in \mathcal{F}_t. \tag{7}$$

Following Hansen et al. (2011), we define the Fund Confidence Set (FCS) as any subset of $\mathcal{F}_t^0$ that contains $\mathcal{F}_t^*$ with a certain probability, $1 - \lambda$.

With these definitions in place we next explain how the algorithm for constructing the FCS works. The first step sets $\mathcal{F}_t = \mathcal{F}_t^0$, the full list of funds under consideration at time t. The second step uses an equivalence test to test $H_{0,\mathcal{F}_t} : E[d_{ij}] = 0$ for all $i, j \in \mathcal{F}_t^0$ at a critical level λ. If $H_{0,\mathcal{F}_t}$ is accepted, the FCS is $\hat{\mathcal{F}}^*_{1-\lambda,t} = \mathcal{F}_t$. If, instead, $H_{0,\mathcal{F}_t}$ gets rejected, the elimination rule ejects one fund from $\mathcal{F}_t$, and the procedure is repeated on the reduced set of funds, $\mathcal{F}_t$. The procedure continues until the equivalence test does not reject and so no additional funds need to be eliminated. The remaining set of funds in this final step is $\hat{\mathcal{F}}^*_{1-\lambda,t}$.

Because a random number of possibly dependent tests are carried out, it is far from trivial to control the coverage probability of this step-wise 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 Hansen et al. (2011) is to design a sequential procedure that can be used to control the coverage probability, $1-\lambda$, of the FCS. Specifically, Theorem 1 in Hansen et al. (2011) establishes conditions under which the probability that truly superior funds are included in the estimated FCS is greater than or equal to $1-\lambda$, while the probability of wrongly including an inferior fund in $\hat{\mathcal{F}}^*_{1-\lambda,t}$ asymptotically goes to zero.[8]

The elimination of individual funds is based on the funds' relative sample performance, using measures such as (3). Specifically, we can estimate the performance of fund i relative to fund j as $\bar{d}_{ij} = t^{-1}\sum_{\tau=1}^{t} d_{ij,\tau}$. To obtain a better behaved test statistic, we divide this measure by its standard error to obtain

$$t_{ij} = \frac{\bar{d}_{ij}}{\sqrt{\widehat{var}(\bar{d}_{ij})}}. \tag{8}$$

As in Hansen et al. (2011) we can base a test of $H_{0,\mathcal{F}_t}$ on the smallest ("worst") t-statistic chosen among the many pairwise $t-$tests in (8):

$$T_{R,\mathcal{F}_t} = \min_{i,j\in\mathcal{F}_t} |t_{ij}|. \tag{9}$$

Under assumptions listed in Hansen et al. (2011), the set of pair-wise $t-$tests are joint asymptotically normally distributed with unknown covariance matrix, Ω. Because so many pairwise test statistics are being compared and Ω is unknown, the resulting test statistic has a non-standard asymptotic distribution whose critical values can be bootstrapped using the approach of White (2000). Using these draws, the sequential elimination rule is used to purge any fund whose performance looks sufficiently poor relative to that of at least one other fund currently included in the FCS.[9]

[8]The high-level assumptions which ensure this result are that, asymptotically, as the sample size, $T\to\infty$, (i) the probability of wrongly eliminating a fund does not exceed λ (the size of the test); (ii) the power of the test goes to one; and (iii) superior funds are not eliminated from a set containing inferior funds.

[9]Specifically, if the FCS p-value for the fund identified by (9) is smaller than the $\lambda-$quantile of the bootstrapped distribution, then this fund is deemed inferior to at least one other fund and gets eliminated.

2.2. Choosing λ

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

If the estimated performance of many of the funds is quite noisy, then the equivalence test may not be very powerful and the algorithm will eliminate too few funds, resulting in a bloated set of funds that includes many inferior funds. This would simply reflect that the data are not sufficiently informative to distinguish between the performance of different funds. We can easily imagine economic environments or volatility states for which this would plausibly be the case. Conversely, when the data are informative and allows for sharper inference, the equivalence tests first eliminate the poor funds before questioning the superior funds.

We opt for a relatively high value of λ, choosing $\lambda = 0.90$ as our benchmark value. This choice is based on the large sampling errors surrounding individual funds' alpha estimates which means that the power of the test based on (9) can be expected to be quite low, increasing the risk of wrongly including a large set of funds with inferior performance simply because their performance is imprecisely estimated.[10] However, to illustrate the sensitivity of our results to this particular choice of λ, we also consider two alternative values ($\lambda = 0.50, 0.10$) which result in fewer funds being eliminated. We refer to the three sets of λ-values as tight ($\lambda = 0.90$), medium ($\lambda = 0.50$),

[10] Note that funds can avoid being eliminated from the FCS either if they have a high average performance which is precisely estimated, or, alternatively, if their performance is imprecisely estimated (e.g., if their alpha is surrounded by large standard errors). The hope is that the procedure avoids including too many funds in the second category.

and wide ($\lambda = 0.10$).

2.3. Choice of Candidate Set of Funds

We implement the FCS approach as follows. First, because there are more than 2,000 funds in our sample, it is not feasible to conduct all possible pairwise performance comparisons.[11] To handle this issue, our baseline results restrict the set of funds being considered for inclusion in the initial round of the FCS ($\mathcal{F}_t^0$) to the top 15% of funds with positive alpha estimates, $\hat{\alpha}_{i,t}$. This greatly reduces the set of funds under consideration and makes it feasible to implement the approach. Because this cutoff is somewhat arbitrarily chosen, for robustness we also consider cutoffs that use either 25% or 5% of the funds ranked by alpha estimates in the initial round.[12]

An alternative approach to constructing the initial set of funds under consideration ($\mathcal{F}_t^0$) is to identify the set of funds with alphas significantly higher than zero. This can be accomplished using the step-wise procedure of Romano and Wolf (2005). The objective of the Romano-Wolf approach is to identify as many of the truly superior funds as possible while controlling the familywise error rate (FWE). To explain this approach, let α_i be the parameter of interest for fund $i \in \mathcal{F}_t$. Then the null and alternative hypotheses that we test using the Romano-Wolf approach are $H_{0i} : \alpha_i \leq 0$ vs. $H_{1i} : \alpha_i > 0$. The familywise error rate is then defined as the probability (under the true data generating process) of wrongly rejecting the null for at least one fund, i.e., $FWE = prob$(rejecting at least one H_{0i} for which $\alpha_i > 0$). The challenge is to design an approach whose asymptotic FWE is no greater than some critical level while accounting for the multiple hypothesis testing problem arising from comparing so many mutual funds.

To implement the Romano-Wolf procedure, define a test statistic, $test_i$, for testing H_{0i} vs. H_{1i}. Suppose we have renumbered the funds $i = 1, ..., n$ by the magnitude of their individual test statistics, $test_1 \leq test_2 \leq ... \leq test_n$. A critical value, c_1, is then determined such that the set of R_1 funds with test statistics $test_{R_1} \geq c_1$ has a coverage probability of $1 - \lambda$. Funds with lower test statistics (i.e., funds numbered $1, ...,\ R_1 - 1$) are eliminated in this step. Next, the procedure is repeated on the remaining $n - R_1$ funds, resulting in a new critical value, c_2, and elimination of

[11] Even with just 100 funds, $9.33{\times}10^{15}$ pairwise comparisons need to be conducted.

[12] We estimate the fund confidence set using the MulCom 3.0 package for Ox, see Hansen and Lunde (2010) and Doornik (2006).

funds with smaller or less significant positive alpha estimates. The procedure is repeated until no additional fund gets eliminated, at which point it stops. Compared with a single-step procedure, this multistep approach will be more powerful in the sense that it can eliminate additional funds in subsequent rounds, while asymptotically controlling the FWE. We use the Basic StepM method (Algorithm 3.1) of Romano and Wolf (2005) to determine the initial set of funds ($\mathcal{F}_t^0$).[13]

3. Performance Measurement Models

This section introduces the models used to estimate individual mutual funds' risk-adjusted performance. The performance measurement model plays an important role for the results of the FCS approach. A conditional alpha approach that uses both return and holdings data such as (20) might provide sharper inference about alphas than an unconditional approach such as (10) which uses only returns data. Sharper inference on alphas should translate into an improved ability to discriminate between funds with superior performance and funds with inferior performance.

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

3.1. Benchmark model

Our benchmark specification is a four-factor model with constant alpha and constant loadings on the risk factors. Specifically, let R_{it} be the monthly excess return on fund i, measured in excess of a 1-month T-bill rate. Similarly, let $\mathbf{z}_t = (R_{mt}, HML_t, SMB_t, MOM_t)'$ denote the values of the four risk factors, where R_{mt} is the return on the market portfolio in excess of a 1-month T-bill rate,

[13] Like the MCS approach, the Romano-Wolf method uses the White (2000) bootstrap approach to calculate the critical values used for eliminating funds.

HML_t and SMB_t are the value-minus-growth and small-minus-big size factors of Fama and French (1992) and MOM_t is the momentum factor of Carhart (1997) constructed as the return differential on portfolios comprising winner versus loser stocks, tracked over the previous 12 months. The benchmark model takes the form

$$R_{it} = \alpha_i + \beta_i' \mathbf{z}_t + \varepsilon_{it}. \tag{10}$$

Following common practice in the finance literature, we obtain estimates of $\theta_i = (\alpha_i \ \beta_i')'$ using a rolling 60-month estimation window. Such estimates account for slowly-evolving shifts in mutual fund performance and risk exposures.[14]

3.2. Time-varying skills

Many recent studies find evidence suggesting that mutual funds' ability to outperform varies over time. For example, Kacperczyk et al. (2014) find that mutual funds' investment strategies–as well as the ability of mutual funds to outperform–depend on whether the economy is in an expansion or in a recession state; Ferson and Schadt (1996), Avramov and Wermers (2006), and Banegas et al. (2013) show that macroeconomic state variables can be used to track and predict the performance of individual equity mutual funds. Mamaysky et al. (2007, 2008) model fund performance as driven by an unobserved, mean-reverting process. We follow this latter approach and show how it can be generalized–and improved–to take advantage of information from mutual fund holdings data.

To understand what induces time-varying investment performance, our analysis starts with individual stocks' performance. Specifically, we decompose the excess return of each stock, r_{jt}, into a risk-adjusted return component, α_{jt}, a systematic return component obtained as the product between a set of risk exposures, β_j and factor returns, $\mathbf{z}_t$, and an idiosyncratic return component, ε_{jt}. We stack these return components into $N_t \times 1$ vectors α_t and ε_t and an $N_t \times 4$ matrix of betas, β, where N_t is the number of stocks in existence at time t. Notice that the individual stock alphas are allowed to vary over time, reflecting that any abnormal returns are likely to be temporary.[15]

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

[15] Liu and Timmermann (2013) develop a theoretical model with temporary abnormal returns in the context of convergence trading.

Following Mamaysky et al. (2008), for simplicity we assume that stock betas are constant, although this assumption can be relaxed.

Mutual fund returns can be computed by summing across individual stock returns, $\mathbf{r}_t$, weighted by the fund's ex-ante portfolio weights at the end of the previous period, ω'_{t-1}. Using the decomposition of individual stock returns described above, the excess return of an individual mutual fund (i) net of the risk-free rate, R_{it}, can be expressed as follows:

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

where $\alpha_{it} = \omega'_{it-1}\alpha_t - k_i$, $\beta_{it} = \omega'_{it-1}\beta$, and k_i captures the fund's transaction costs.[16] It follows from (11) that an individual fund's alpha is a value-weighted average of its stock-level alphas.

Different approaches have been suggested for capturing time variation in fund alphas. Mamaysky et al. (2008) view manager skills as a latent process driven by an unobserved and potentially persistent process which reflects the fund's ability to process and act on private information.[17] We next describe this approach and show how to generalize it to incorporate information on fund holdings, ω_{it}.

Suppose that the manager of fund i receives a private signal, F_{it}, which follows a stationary autoregressive process,

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

The innovations ε_{it} in (11) and η_{it} in (12) are assumed to be independent of each other and normally distributed. Following Mamaysky et al. (2008) we assume that fund portfolio weights are linear in the private signal

$$\omega_{it-1} = \bar{\omega}_i + \gamma_i F_{it-1}, \tag{13}$$

[16] As in Mamaysky et al. (2008) these are assumed to be proportional to the fund's assets under management.

[17] A related approach, proposed by Kosowski (2011), models manager skills as a latent variable driven by a regime-switching process. As in Mamaysky et al. (2008), this gives rise to a filtering problem, although the filter is non-linear in this case.

Moreover, assuming that funds' private signals have predictive power over subsequent stock-level alphas, we have

$$\alpha_{it} = \bar{\alpha}_i F_{it-1}. \tag{14}$$

Equations (11) and (14) imply that a fund's alpha and beta depend on the signal, F_{it-1} :

$$\begin{aligned}
\alpha_{it} &= \bar{\omega}_i' \bar{\alpha} F_{it-1} + \gamma_i' \bar{\alpha} F_{it-1}^2 - k_i \\
&\equiv \bar{\alpha}_i F_{it-1} + b_i F_{it-1}^2 - k_i, \qquad (15) \\
\beta_{it} &= \bar{\omega}_i' \beta + \gamma_i' \beta F_{it-1} \\
&\equiv \bar{\beta}_i + \mathbf{c}_i F_{it-1}. \qquad (16)
\end{aligned}$$

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

$$\begin{aligned}
R_{it} &= \bar{\alpha}_i F_{it-1} + b_i F_{it-1}^2 - k_i + (\bar{\beta}_i' + \mathbf{c}_i' F_{it-1}) \mathbf{z}_t + \varepsilon_{it} \qquad (17) \\
F_{it} &= \nu_i F_{it-1} + \eta_{it}.
\end{aligned}$$

We refer to this as the latent skill (LS) model. As explained by Mamaysky et al. (2008), the parameters of this model can be estimated fund-by-fund using an extended Kalman Filter that accounts for the presence of the squared value of the underlying state variable, F_{it-1}^2, in equation (17).[18]

3.3. Introducing Information from Fund Holdings

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

One way to address this issue is by making use of additional information. Specifically, as is clear from equations (11) and (14), information on funds' portfolio holdings can potentially be used

[18] It is necessary to normalize one of the elements of $\mathbf{c}_i$. Given such a normalization, the four-factor model requires 13 parameters to be estimated.

to capture how a fund's alpha evolves through time. Building on this idea we next generalize the methodology in Mamaysky et al. (2008), and show that holdings-based information can be added in the form of an additional measurement equation in the state space representation of the model, and that this model can be estimated by means of an extended Kalman filter.

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

$$CS_{it} = \omega'_{it-1}(\mathbf{r}_t - \mathbf{r}_{bt}), \tag{18}$$

Here $\mathbf{r}_t$ and $\mathbf{r}_{bt}$ are vectors of excess returns on stocks ($\mathbf{r}_t$) and benchmark portfolios ($\mathbf{r}_{bt}$), respectively. These are chosen to match, as closely as possible, the characteristics of the individual stocks. Because the characteristic-matched stocks are chosen mechanically and the average stock can be expected to have zero alpha, $\alpha_{bt} = \mathbf{0}$. Moreover, $\beta_{it} = \beta_{bt}$ because the benchmark stocks are chosen to match the individual stocks' factor exposures at time t. Using (17), (18), and (15) we get

$$\begin{aligned} CS_{it} &= \omega'_{it-1}\left(\alpha_t + \beta'\mathbf{z}_t + \varepsilon_t - \left(\alpha_{bt} + \beta'_b\mathbf{z}_t + \varepsilon_{bt}\right)\right) \\ &= \omega'_{it-1}(\alpha_t - \alpha_{bt}) + \omega'_{it-1}(\beta - \beta_b)'\mathbf{z}_t + \omega'_{it-1}(\varepsilon_t - \varepsilon_{bt}) \\ &= \alpha_{it} + k_i - \omega'_{it-1}\alpha_{bt} + (\beta_{it} - \beta_{bt})'\mathbf{z}_t + \varepsilon_{it} - \varepsilon_{bt} \\ &= \bar{\alpha}_i F_{it-1} + b_i F^2_{it-1} + \varepsilon_{it} - \varepsilon_{bt}. \end{aligned} \tag{19}$$

Since the CS measure does not depend on estimated risk factor loadings obtained over some prior historical period, it has the potential to generate a more accurate estimate of fund performance and thus improving on the performance of return-based models.[19]

[19] Alternatively, the CS and return-based measures can be viewed as different estimates of the same underlying fund performance, observed with different estimation errors. When presented with different estimates of the same

The generalized latent skill, holding-based (LSH) model can now be written in state space form as

$$\begin{pmatrix} R_{it} \\ CS_{it} \end{pmatrix} = \begin{pmatrix} \bar{\alpha}_i F_{it-1} + b_i F_{it-1}^2 - k_i \\ \bar{\alpha}_i F_{it-1} + b_i F_{it-1}^2 \end{pmatrix} + \begin{pmatrix} \bar{\beta}_i + \mathbf{c}_i F_{it-1} \\ 0 \end{pmatrix} \mathbf{z}_t + \begin{pmatrix} \varepsilon_{it} \\ \varepsilon_{it} - \varepsilon_{bt} \end{pmatrix} \tag{20}$$

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

Compared to the model in Mamaysky et al. (2008), this model has the additional information contained in CS_{it}, which has the potential to make estimation and extraction of funds' private information component, F_{it}, more precise. We can again estimate the parameters of (20) using the extended Kalman filter, now using two measurement equations.

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

4. Performance Results

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

object, statistical theory suggests that there can be gains from combining such estimates.

4.1. Data

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

To construct our estimate of the CS measure we use quarterly holdings data from Thomson Financial CDS/Spectrum.[20] We use these quarterly holdings data to construct a three-month estimate of CS. We merge the returns and holdings data using the MFLINKS files of Wermers (2000) which have been updated by the Wharton Research Data Services and allows us to map the Thomson holdings data to CRSP returns using the funds' WFICN identifier. We require each fund to have at least six months of data and also require funds to have contiguous returns data. In total we have returns and holdings data on 2,480 funds, but we exclude 255 sector funds and so end up with 2,225 funds.[21] The number of funds included in the analysis peaks at above 1700 in 2009 before declining to around 1500 in 2012.

For each fund we obtain an alpha estimate using time-series data on the fund's historical returns. Funds with a very short return record tend to generate noisy alpha estimates. To avoid that our analysis gets dominated by such funds, we require funds to have a return record of at least five years. Table 1 provides summary statistics for the cross-sectional distribution of individual fund alphas, using each of the three performance models described in the previous section. The median fund has a negative alpha ranging from -61 bps/year to -74 bps/year across the three models. The finding that the median fund underperforms on a risk-adjusted basis is consistent with previous academic studies. The bottom 5% of funds ranked by alpha performance have a negative alpha estimate around -35 bps/month or just under -4%/year—a number that again does not vary much across the three model specifications. The top 5% of funds have alpha estimates ranging from 21 bps/month to 25 bps/month, approximately 3%/year; these estimates are again quite similar across the three models.

[20] In the early part of the sample funds were only required to report holdings every six months.

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

4.2. *Performance of Decile-sorted portfolios*

To establish a reference point for our FCS results, we first follow the common practice of ranking individual funds' alphas and forming decile portfolios. This approach can be used to see if funds that are expected to have the highest alphas do indeed produce better subsequent performance than lower-ranked funds. Specifically, each month, t, we rank funds based on their expected alphas $\hat{\alpha}_{i,t+1|t}$. We then form ten equal-weighted decile portfolios with the first portfolio (P1) containing the bottom 10% alpha-ranked funds, the next decile containing funds ranked in the second-lowest 10%, continuing up to the top 10% of alpha-ranked funds (P10). To obtain more detailed classification results for the bottom and top funds, we also divide P1 into the bottom 5% alpha-ranked funds and funds ranked between the bottom 5% and 10% (labeled P1A and P1B, respectively); we use a similar split for the P10 portfolio (labeled P10A and P10B). Finally, we record the returns on each of these portfolios over the subsequent month. Each month, as new data arrive, we repeat this sorting routine and, again, form equal-weighted portfolios based on the funds' updated alpha estimates and record their returns. We use five years of data to initiate the portfolio sorts and another five years to obtain an estimate of the predictive alpha in (3) and thus generate a time series of portfolio excess returns, R_{pt}, over the 21-year period from 1990:07 to 2012:12.

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

$$R_{pt} = \alpha_p + \beta_p' \mathbf{z}_t + \varepsilon_{pt}, \quad t = 1, \ldots, T. \tag{21}$$

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

Table 2 presents alpha estimates for the decile portfolios. We find strong evidence of negative and statistically significant alpha estimates for the bottom three ranked decile portfolios (P1-P3). The underperformance of these decile portfolios ranges from -8 bps/month to -21 bps/month and are quite similar across the three different models used to rank funds. The alpha estimate of the top-ranked decile portfolio (P10) is smaller (between 2 and 6 bps/month) and statistically insignificant, suggesting that the conventional portfolio sorting approach fails to identify funds with abnormal positive performance. This conclusion carries over from the top 10% to the top 5% of funds (P10B) which only perform marginally better than the P10 portfolio.

Despite these shortcomings, the portfolio sorting approach does succeed in differentiating between the best and worst performing funds as the estimated differential in alphas between the P10 and the P1 portfolios is large and positive (18-26 bps/month) and highly statistically significant. It is clear from the previous results, however that this finding is driven mostly by the ability to identify funds with inferior performance. As an alternative test of the value of the ranking information in the portfolio sorts, the last line in Table 2 reports the MR test for a monotonic pattern in the alphas proposed by Patton and Timmermann (2010). The null is that there is a flat or declining pattern in the alphas while the alternative is that there is a monotonically increasing pattern, so a small p-value for this test is evidence that a model succeeds in ranking the future risk-adjusted performance of the funds. The test statistic generates a p-value of 0.07 when applied to the portfolios ranked by the constant-alpha benchmark model and is statistically significant for the LS and LS-CS models. Hence, the performance models do contain valuable ranking information despite their failure to identify funds with large positive alphas.

We conclude from these findings that the conventional portfolio sorting approach can be used to identify a broad set of funds with inferior performance, but is less well suited for identifying funds with superior performance.

4.3. Performance of top FCS funds

Figure 1 shows how the FCS approach helps select funds that stand out even among funds with positive forecasts of alpha. The figure plots the distribution of predictive alpha estimates obtained using the latent factor-holdings model at a single point in time (July 2006). The black curve shows the distribution of predictive alphas for the full set of funds in existence at that point, i.e., the population of funds. This curve is centered a little to the left of zero and has a wide dispersion. The green and red lines show the distributions of predictive alphas for funds with positive predicted alphas (green line) and funds whose alpha estimates are positive and statistically significant using the Romano Wolf approach (red line). Finally, the blue line captures the distribution of predictive alphas for funds included in the FCS. The FCS curve is much further to the right than the distribution curve for the set of funds with positive predictive alphas, highlighting that the FCS methodology is much more discriminating than an approach that simply selects funds with (significantly) positive alphas.

We next analyze the performance of our alternative FCS approach which forms portfolios from the set of funds identified as top performers. Table 3 presents alpha estimates for a set of portfolios formed by equal- or value-weighting the funds selected each month by applying the FCS approach to the current set of candidate funds. As in Table 2, the (out-of-sample) period goes from 1990:07 to 2012:12. As explained in Section 2, we consider a range of values for the significance level of the test, $\lambda = \{0.10, 0.50, 0.90\}$, corresponding to a wide, medium, and tight FCS, and present results for the three performance models introduced in Section 3. Our baseline analysis (shown in Panel A of Table 3) assumes a tight FCS and uses the top 15% of funds as the initial set of candidate funds. We consider both equal- and value-weighted portfolios of funds, but focus our discussion on the former.

For the constant alpha benchmark model the results show that moving from the top decile portfolio to the tight FCS portfolio increases the average alpha from 6 bps/month to 30 bps/month, although the latter estimate is not significantly different from zero at conventional critical levels. Similar, if somewhat stronger, results hold for the LS model estimated only on returns data. For this model the alpha estimate is 25 bps/month for the FCS portfolio. This performance is 20 bps/month better than that of the top decile and the alpha estimate is statistically significant with a p-value below 5% using a one-sided test.

We obtain very strong performance results for the latent skill model that combines return and holdings data (LSH). For this model the tight FCS portfolio produces an alpha estimate that averages 54 bps/month.

As we reduce λ from 0.90 to 0.50 or 0.10, fewer funds get eliminated, the set of funds included in the FCS portfolio increases, and the performance of the FCS portfolios gets markedly reduced. This suggests that the additionally included funds tend to perform worse than the funds identified by the more selective approach. This finding is unsurprising because many funds have alphas whose estimates are surrounded by large standard errors. For these funds, choosing a small value of λ

means that the equivalence test used to eliminate inferior funds has insufficient power to reject the null that the funds' alphas are identical; we analyze this finding in more detail below.

Recall that we need to restrict the initial set of funds ($\mathcal{F}_t^0$) used in the FCS analysis in order to limit the number of pair-wise comparisons that has to be conducted by our approach. The results in Panel A assume that we apply the FCS approach to the top 15% of funds ranked on their alpha estimate. To explore the sensitivity of our results to starting either with a wider (25%) or narrower (5%) set of funds, panels B and C in Table 3 report results for these alternative initial sets of funds. Again we see that the performance is better for the FCS portfolios than for the decile portfolios and that the performance tends to get better as we narrow the set of funds in the FCS, i.e., apply a higher value of λ.

Panel D in Table 3 considers the performance of the strategy that first uses the step-wise bootstrap method of Romano and Wolf (2005) to identify funds with significantly positive alphas followed by the use of the FCS approach to select the set of top performers among these funds. Again, we see that the FCS portfolios generate higher alpha estimates than the top decile portfolio. Improvements are particularly large for the constant alpha (benchmark) and latent skill-holdings models for which the performance of the tight FCS portfolio is similar to that in Panel A.

We obtain very similar results if we use value weights rather than equal weights to form the top FCS portfolios. For example, for the tight FCS portfolio in Panel A of Table 3, the alpha estimates generated by the three performance models change to 36, 25, and 52 bps/month (previously 30, 25, and 54 bps/month using equal weights).

We conclude the following from these results. First, the FCS approach is capable of selecting funds whose performance, when combined into equal- or value-weighted portfolios, is far better than that achieved by funds included in the conventional rank-based top-decile portfolio. Second, the performance of the FCS approach is best when we allow the approach to eliminate more funds and become more selective. This is consistent with our finding from the analysis of the decile portfolio's performance that the ability to outperform is not very widespread among the funds in our sample. Third, the FCS approach works particularly well for the latent skill-holdings model which allows for time-varying fund performance and combines returns and holdings data to provide sharper inference about alpha estimates.

4.4. Time variation in top-rated funds' performance

The alpha estimates reported in tables 2 and 3 show the average performance of a set of decile-sorted and FCS portfolios, respectively. Such estimates do not reveal if the performance of difference portfolios is concentrated in certain states or occur over certain periods of time. Addressing this point can help provide us with insights into the nature of the portfolios' performance. We next explore this issue using a new non-parametric methodology designed to capture "local" time variation in portfolio alphas.

To track different portfolios' performance over time, we introduce 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:

$$R_{it} = \alpha_i(t/T) + \beta_i'(t/T)\mathbf{z}_t + \varepsilon_{it}, \quad t = 1, \ldots, T. \tag{22}$$

Here $\alpha_i(\cdot)$ and $\beta_i(\cdot)$ are unknown smooth functions that are allowed to depend on the sample "fraction", t/T, and thus can vary over time.[22] To see how we can nonparametrically estimate the parameters, define the vector of regressors $\mathbf{X}_t = (1\ \mathbf{z}_t)$ and parameters $\theta_i(t/T) = (\alpha_i(t/T)\ \beta_i'(t/T))$ and rewrite equation (22) as

$$R_{it} = \mathbf{X}_t'\theta_i(t/T) + \varepsilon_{it}. \tag{23}$$

The parameters $\theta_i(t/T)$ can be estimated by means of a two-step procedure that first considers the OLS estimator $\hat{\gamma}_i = (\hat{\gamma}_{i0}, \hat{\gamma}_{i1})'$ of the transformed model

$$k_{st}^{1/2} R_{is} = k_{st}^{1/2}\mathbf{X}_s'\gamma_{i0} + k_{st}^{1/2}\left(\frac{s-t}{T}\right)\mathbf{X}_s'\gamma_{i1} + \varepsilon_{is}, \quad s = 1, \ldots, T, \tag{24}$$

where $k_{st} = k(\frac{s-t}{Th})$ is a kernel function, and h is the bandwidth. In the second step we construct an estimator of $\alpha_{it} = \alpha_i(t/T)$ as

$$\hat{\alpha}_{it} = (\mathbf{e} \otimes \mathbf{I})\hat{\gamma}_i, \tag{25}$$

[22] Technically, the $\alpha_i(\cdot)$ and $\beta_i(\cdot)$ functions allow for a finite number of discontinuities.

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 $\hat{\alpha}_{i,t}$ from (22) portrays the evolution in the performance of the different portfolios in a way that does not "average out" potentially interesting time variation in α_i. This can be contrasted with a more conventional full-sample approach or even a rolling-window procedure which does not take into account how much the data varies over time.

Figure 2 shows the evolution through time in the risk-adjusted performance (alpha) of the superior funds included in the FCS portfolio (in red) along with the performance of the traditional top decile portfolio (P10, in black). The blue bands on the graph show the pointwise standard error of $\hat{\alpha}_{i,t}$.[23] The FCS portfolio in the top panel assumes $\lambda = 0.90$ while the FCS portfolio in the middle and bottom panels sets $\lambda = 0.50$ and $\lambda = 0.10$, respectively. First consider the top panel which tracks the performance of the tight FCS portfolio. From 1990-1995 the local alpha estimates of the FCS portfolio and the P10 decile portfolio are close to zero. However, starting in 1996 the performance of the FCS portfolio increases dramatically before peaking around 200 bps/month in 1999-2001. Over the same five-year period the performance of the P10 decile portfolio only improves marginally. The FCS portfolio only performs poorly during brief spells in 1990 and around 2006. After 2008, the FCS portfolio produces positive alpha performance around 50 bps/year. The time-series evolution in the alpha estimates of the FCS portfolios with $\lambda = 0.5$ or $\lambda = 0.10$ (middle and bottom panels) is notably more subdued than the estimates for the FCS portfolio with $\lambda = 0.90$ and are closer to the performance of the P10 decile portfolio.

4.5. Performance of worst FCS funds

Next, we analyze whether the FCS approach can be used to identify funds with inferior performance. To this end we need to redefine the objective function. We do so by maximizing a "predictive alpha"

[23] Standard error bands are constructed using the stationary bootstrapping methodology of Politis and Romano (1994) to resample the time series of returns associated with each portfolio.

loss function which focuses on funds in the left tail of the performance distribution:

$$L_{i,t} = Min(\hat{\alpha}_{i,t|t-1}, 0) sign(R_{i,t} - \beta'_{i,t} F_t). \tag{26}$$

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

Again we base our empirical analysis on a sample estimate of the mean loss

$$\bar{L}_{i,t} = \frac{1}{t} \sum_{\tau=1}^{t} Min(\hat{\alpha}_{i,\tau|t}, 0) sign(R_{i,\tau} - \hat{\beta}'_{i,t} F_\tau). \tag{27}$$

The bottom panel of Figure 1 shows the distribution of predictive alphas for the funds that were in existence (and had a minimum of five years of data) in 2006:07. The distribution of predictive alphas for funds in the FCS (blue line) is centered around -0.5%/year and is much further to the left than the distribution curve for funds with negative alpha estimates (green line), funds with significantly negative alphas (red line) or for the population of funds as a whole (black curve). As in the case of funds with superior performance (top panel), the FCS approach takes a more selective approach to identifying funds with inferior performance than an approach that simply considers funds with negative expected performance.

Table 4 reports the four-factor risk-adjusted performance of the set of worst funds identified by the FCS approach. First consider the performance of the bottom funds as ranked by the constant-alpha benchmark model. Assuming again that the initial set of funds is chosen as the top 15% of funds in existence ranked by their alpha estimates (Panel A) and equal weighting is applied to form portfolios of funds, this approach yields a negative alpha estimate of -19 bps/month for the bottom decile portfolio. The alpha estimate for the FCS portfolio with $\lambda = 0.90$ more than doubles in magnitude to -43 bps/month which is highly statistically significant. Moreover, we find similarly large and statistically significant (negative) alpha estimates for the worst funds across a broad range of values for the λ parameter.

These results show that the performance estimates for the funds with inferior performance are not as sensitive to the choice of λ as are our estimates for funds with superior performance. It is easy to explain this finding. As we reduce λ from 0.90 to 0.50 or 0.10, we widen the set of funds included in the FCS portfolios. Because inferior performance is so widespread—Table 2 suggests that more than 30% of funds underperform—this has little effect on the portfolio comprising the worst funds. In contrast, superior performance is much less widespread and so requires constructing a more concentrated portfolio (high λ) to be exploited by the approach.

Turning to the fund rankings based on the latent skill model we find equally strong evidence that the FCS methodology succeeds in identifying funds with strongly negative performance. Specifically, the average alpha goes from -21 bps/month for the bottom decile of funds to -36 bps/month for the FCS portfolio with $\lambda = 0.90$. In all cases the t-statistics are greater than three in absolute value and so the results are both economically and statistically significant. A similar finding holds when we apply our approach to the latent skill-holdings model. As shown in the last columns of Table 4, the average alpha performance drops from -16 bps/month for the bottom decile of funds to -32 bps/month for the FCS portfolio with $\lambda = 0.90$. Again the results are quite robust to using value-weighting as opposed to equal-weighting the funds in the FCS portfolio; using value weights and $\lambda = 0.90$ the alpha estimates for the three performance models are -41 bps/month, -27 bps/month and -26 bps/month, respectively (-43, -36, and -32 bps/month under equal weights).

Figure 3 plots the time-series evolution in the local alpha estimates for the FCS portfolios comprising the bottom-ranked funds. For the FCS portfolio with $\lambda = 0.90$ (top panel) we see substantially worse performance during the first five years of the sample followed by slightly better performance during 1998-2002 and, again, in 2012. The FCS portfolio with $\lambda = 0.50$ (bottom panel) generally produces very similar or slightly smaller alpha estimates than the P1 portfolio of bottom-decile funds.

4.6. Performance in expansions and recessions

Studies such as Kacperczyk et al. (2014) suggest that fund performance and fund manager skill are related to the business cycle. To explore if the performance of the funds in our sample varies with the state of the economy, Table 5 reports fund performance separately for expansion and recession periods defined using the NBER recession indicator. For the funds selected to be included in the top FCS portfolio by the latent skill-holdings model we find that the alpha performance is a bit higher in expansions (66 bps/month versus 54 bps/month on average). Although these funds' alpha estimate is negative (-27 bps/month) in recessions, this estimate is imprecisely estimated and far from statistically significant. Still, this evidence is indicative that funds with stock picking talent mostly benefit from such skills during expansion periods, consistent with the findings in Kacperczyk et al. (2014). We fail to find similar state dependence in the performance of the funds ranked by the benchmark model or the simple LS model that does not use holdings data. We also do not find any state dependence in the performance of the top decile portfolio.

For the FCS portfolio of bottom-ranked funds the alpha estimate is a highly significant -27 bps/month in expansions versus a whopping -105 bps/month during recession months. Moreover, the latter estimate is statistically significant, suggesting that the latent skill-holdings model that uses holdings information is particularly successful at identifying inferior funds during recessionary periods when used as an input to the FCS approach.

5. Selection of Top and Bottom Funds

The previous section shows that the FCS approach can be used to successfully identify funds with superior risk-adjusted performance. Moreover, the approach works particularly well for the latent skill-holdings model that combines holdings and returns data to admit a sharper inference on fund alpha estimates. The FCS approach is fundamentally different from existing methods such as decile sorting which keep the proportion of "top" funds fixed through time. In contrast, the FCS approach endogenously determines how many funds to include. This section provides details of both the number of funds selected by the FCS approach and their identify.

5.1. Identifying the set of top funds

The FCS approach endogenously determines how many funds get included in the set of superior funds and thus provides a new way to address how widespread superior performance is among mutual funds. We can gain important insights by studying how the set of superior funds evolves over time. Specifically, the turnover in the set of funds identified to be superior is related to the persistence in individual funds' performance. For example, if a small set of funds perform significantly better than all other funds for a sustained period of time, then the FCS should contain very few funds during such a period. Conversely, periods where few funds have outstanding performance will result in a wider set of funds being included in the FCS because no individual fund stands out.

The number of funds identified as being superior depends on the ability of the underlying performance model to accurately capture individual funds' risk-adjusted returns. For example, during periods where individual funds' alphas are estimated with large sampling errors, we would not expect t-tests such as (8) to be able to eliminate many funds from the FCS which, thus, might include too many funds. In contrast, during times where individual funds' track records are more informative and alpha estimates are sharper, we would expect the FCS approach to be more discriminating, resulting in a narrower set of funds being included.

The top panel of Figure 4 shows the evolution in the number of funds that are included in the FCS. Because this set depends on the size of the test, λ, we present results for three different values, $\lambda = \{0.1, 0.5, 0.9\}$, but focus on $\lambda = 0.90$ in most of our discussion. The number of funds included in the top FCS fluctuates considerably over time. Around 5-10 funds get selected during the two-year period from 1992-1994, followed by a considerably narrower FCS up to 2000. At this point, the FCS undergoes considerable change with a broad array of funds entering and leaving the FCS up to 2007. During the last 18 months of the sample the top FCS grows in size and includes around 40 individual funds.

Interestingly, the sensitivity of the number of funds selected with regards to the value of λ fluctuates a great deal over time. For example, in 2009 and 2010 the most stringent FCS with $\lambda = 0.90$ only selects a single fund, whereas the FCS that uses $\lambda = 0.10$ includes up to 60 individual funds. At other times, notably during early parts of the sample, the number of funds selected depends less on λ. On average the FCS with $\lambda = 0.90$ includes 8 funds while the FCS based on $\lambda = 0.50$ and $\lambda = 0.10$ on average includes 16 and 22 funds, respectively. Of course, there is nothing wrong with finding such sensitivity of the results to λ. This parameter should be chosen to reflect the trade-off between an investor's utility from correctly including a genuinely superior fund versus her disutility from wrongly including an inferior fund.

Because the number of funds in our data increases over time, it is useful to also consider the percentage of funds identified by the FCS approach as being top performers. The bottom panel in Figure 4 provides a time-series plot of this proportion for the three different values of λ considered above. The percentage of top performers varies greatly over time. It fluctuates between 0.5% and 4% in the early part of the sample (assuming $\lambda = 0.90$) and is quite low up to 2000. After this period, 2% or more of the funds regularly get selected and the proportion of included funds peaks at around 4% in 2011-2012. For the smaller values of λ the percentage of funds is of course a bit higher but it rarely exceeds 6%. On average the tight FCS includes 1.3% of the funds while the medium and wide FCS portfolios include 3.0 and 4.2% of the funds, respectively.

Figure 5 presents more details on the funds that are identified as superior by the FCS approach. Specifically, for each fund that gets selected by the FCS approach at least once during the sample, Figure 5 shows when the fund is chosen. The labeling on the y-axis is arbitrary but maps one-to-one to the fund ID. We see that a single fund–Fidelity Select Technology–gets selected for most of the five-year period from 1995-2000. Between 2008 and 2011 a single fund–T Rowe Price Media and Telecommunications–is again selected.

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

5.2. Identifying the set of worst funds

Figures 6 and 7 present plots similar to those presented in figures 4 and 5, but now applied to the set of funds identified to have inferior performance. Figure 6 shows that the set of funds with inferior performance fluctuates considerably through time, peaking at close to 5% but often containing

1% or less of the funds. Once again, the panels in Figure 6 show considerable differences in how many funds get included among the set of inferior funds across different values of λ. For example, up to 10% of the funds get included when we set $\lambda = 0.10$. On average, the bottom FCS with $\lambda = 0.90$ includes 11 funds while the FCS with $\lambda = 0.50$ and $\lambda = 0.10$ on average include 25 and 38 funds, respectively. These figures correspond to 1.7%, 4.5% and 6.7% of the total number of funds, respectively. Thus, the set of inferior funds is broader and includes more funds than the equivalent FCS for the superior funds.

Figure 7 shows the inclusion plot for individual funds with inferior performance. While on average more funds get selected for the bottom FCS than for the top FCS, we continue to see periods during which only a single fund is selected by the tight FCS approach applied to the bottom funds. For example, between 1993 and 1995 only the Centurion Growth Fund gets included in the tight FCS while between 2001 and 2004, the tight FCS of inferior funds only includes the Phoenix Oakhurst Strategy F fund.

5.3. Turnover and durations of portfolios

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

$$Turnover = \frac{1}{2}\sum_{i=1}^{N_t} |\Delta w_{it}|, \tag{28}$$

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 portfolio composed of the tight FCS of best funds is 0.19, while the corresponding figure for the FCS portfolio based on the worst funds is 0.26. For comparison the monthly turnover of the top and bottom decile portfolios are 0.12 and 0.11, respectively. Thus, the turnover is somewhat higher for the FCS portfolios, which is what we would expect given that they typically contain fewer funds than the broader-diversified decile portfolios.

To help explain what generates the higher turnover in the tight FCS portfolios, Figure 8 provides details on the average duration of funds in the FCS (left column) and decile portfolios (right column) with top performers in the upper panels and bottom performers in the lower panels. Focusing on top-ranked funds the figure shows that the main reason for the high turnover in the top FCS portfolios is that approximately half of all funds remain in the FCS portfolio for only one or two months, while another 10% of the funds get selected for three consecutive months. In contrast, only about 35% of funds remain in the top decile portfolio for three months or less. There are also fewer funds with very long durations in the top FCS portfolio compared to the top decile portfolio. Similar conclusions hold for the portfolios of bottom-ranked funds shown in the lower panels of Figure 8.

5.4. Determinants of the size of the set of superior and inferior funds

Our empirical analysis up to this point shows that the width of the set of funds identified to have superior performance varies a great deal over time. There are several reasons why we find such time variation, including (i) competition among funds (Hoberg et al. (2015); (ii) time-variation and state-dependence in managers' ability to outperform (Ferson and Schadt (1996); (iii) decreasing returns to scale in fund performance (Glode et al. (2011)); and (iv) random sampling variation associated with estimation error in the alpha estimates.

To see whether the proportion of funds identified to have either superior or inferior performance depends on the state of the economy, we perform a simple analysis that regresses the percentage of funds in either the top or the bottom FCS on a range of state variables commonly used to capture the state of the economy. Specifically, we use the 1-month T-bill rate, the term spread, the default spread (i.e., the difference between the yield on BAA- and AAA rated bond portfolios) and the dividend yield as our state variables. These variables feature prominently in the literature on return predictability for the broad stock market (Welch and Goyal (2008)) and have also been used to model state-dependence in mutual fund performance (Ferson and Schadt (1996), Avramov and Wermers (2006), and Banegas et al. (2013)).

Table 6 reports the outcome of this analysis. We find that many of the state variables are significantly correlated with the width of the top FCS portfolio. Specifically, the proportion of funds included in the top FCS tends to be higher in environments with a high short interest rate, a steep yield curve (high term spread), and a small default yield. In contrast, the dividend yield is not significantly correlated with the proportion of funds deemed to have superior performance.[24]

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

We conclude from this evidence that economic states with higher expected future growth (higher term spread) and lower risk premia (smaller default yield) are associated with a broader set of funds being identified as superior.[25] Such states are more likely to occur in economic expansions and so our finding is consistent with that of Kacperczyk et al. (2014) that some funds can pick stocks during economic upturns. Conversely, inferior performance tends to become more widespread during recessions.

An alternative way to address how difficult it is for mutual funds to generate abnormal performance and how this varies through time is by studying how many funds have significantly positive or negative alpha estimates. The top panel in Figure 9 plots the proportion of funds with a statistically significant alpha estimate, identified using the methodology of Romano and Wolf (2005) which is ideally suited to address this question. The number starts at 16% in 1990 (approximately 25 funds), drops to less than 4% in 1999 (approximately 15-20 funds), and ranges between 5% and 10% for the remainder of the sample, corresponding to between 40 and 75 funds. This shows that there is significant time variation in the scope for individual funds to outperform. The downward trend from 1990 through 1999 and, again, from 2006 until the end of the sample (2012) is consistent with competition intensifying through time (Pastor et al. (2014)).

The bottom panel of Figure 9 plots the proportion of funds with significantly negative alpha estimates. On average about 20-25% of funds have significantly negative alpha estimates, however

[24]A similar set of results apply if we focus on the number of funds (rather than the proportion of funds), although now the T-bill rate is insignificant while conversely the default yield variable is negatively correlated with the number of funds in the top FCS.

[25]We also regressed the percentage of funds in the top and bottom FCS portfolios on the NBER recession indicator. We found borderline significant evidence that the percentage of inferior funds is higher during such periods. In contrast, there was no evidence that the fraction of superior funds depends on the underlying economic state.

the scope for underperformance has varied substantially through time. Less than 15% of funds had significantly negative alphas in 1994 and from 2010-2011, whereas more than 25% of funds had significantly negative alpha estimates during much of the decade from 1999-2009.

5.5. Discriminating between superior and inferior funds

The FCS methodology uses a sequence of pairwise comparisons of individual funds' performance to determine if there exists a fund whose performance, measured relative to that of another fund, is clearly inferior. As we have seen, the ability to eliminate inferior funds–as measured by the breadth of the set of superior funds–varies over time. Another way to convey insights into how the FCS algorithm works over time is by studying the average value of the test statistic in the elimination steps leading up to the final fund confidence set. Figure 10 plots this value over time. The average value of the t-test used to eliminate inferior funds (top panel) hovers between -2 and -3, but drops significantly in 2007 and 2008 to values between -5 and -6.

The average value of the test statistic used to eliminate superior funds from the set of inferior funds (bottom panel) again hovers between -2 and -3 although it varies less over time compared to the average value of the test statistic used to determine the set of superior funds.

6. Style and Industry Exposures

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

6.1. *Style Exposures*

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

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

Interestingly, the portfolio of superior funds consists of funds whose average size ($3.3bn) is considerably higher than that of both the average fund ($1.7bn) and the inferior funds ($0.5bn). Turning to the style factors, the portfolio of superior funds holds firms with higher size, lower book-to-market ratios and higher momentum than the average fund while the portfolio of inferior funds holds firms with lower size, higher book-to-market ratios and higher momentum than the average fund. Both types of funds also have a smaller return gap. The inflows are much larger for the superior funds 2.7% per month as compared to only -0.85% and 1.4% for inferior and average funds, respectively. Superior funds also have a lower expense ratio and a lower turnover than inferior funds. This is consistent with the findings of Carhart (1997) that higher expense ratios and higher turnover tend to reduce net return performance.

Figure 11 plots style exposures for funds in the top (left column) and bottom FCS along with the style exposures for the average fund. Comparing the plots for the top FCS portfolio to that of the average fund, the former is seen to be more volatile. This is unsurprising because the FCS portfolio consists of far fewer funds than the overall universe and so benefits less from a "portfolio diversification" effect. However, it may also be indicative of style rotations among the top funds.

Compared to the average fund, the top FCS funds over-weight small stocks in 1996, 2001-2003 and, again, from 2009-2012. Conversely, the top FCS portfolio predominantly overweights large cap stocks during 2005-2008 (with exception of a brief spell in 2007). The top FCS portfolio tilts towards growth stocks from 1994 to 2003 and, again, from 2009 to 2011 but this portfolio overweights value stocks from 2003-2009 and during the last two years of the sample. As shown in the bottom panel of Figure 11, the top FCS stocks have a notably higher exposure to the momentum factor than the average fund only in brief spells during 1998, 2002 and 2012, suggesting that this portfolio is not predominantly capturing rewards for momentum risk.

A very different picture emerges for the bottom FCS funds whose factor loadings are shown in the right panels of Figure 11. These funds underweight small cap stocks between 1993 and 1999 and again from 2009 to 2012, but mostly overweight small stocks during the interim period from 1997-2008. The bottom-ranked FCS funds overweighted value stocks throughout most of the sample while their exposure to the momentum risk factor did not deviate much from that of the average fund.

Figure 12 shows histograms depicting the performance of the top FCS funds (top row) for different values of the size, book-to-market and momentum risk factors sorted into quintiles. For example, the left-most blocks show the average risk-adjusted performance of the FCS funds during months with the lowest quintile of realized values of the different risk factors, while the right-most blocks show the average performance during months in which the realized value of the factors were in the top quintile.

The top left panel reveals a systematic and monotonically increasing relation between the performance of the top-rated funds and the size factor: the higher the size factor, the better the funds in the top FCS approach performed during a particular month. The relationship between the realized value of the book-to-market factor and the performance of the top FCS funds is inverse: the top-rated funds perform far better (with an alpha close to 100 bps/month) during months where this factor was at its lowest, i.e., during months where growth stocks outperformed value stocks. Finally, as shown in the right panel, there is a less systematic relation between the value of the momentum factor and the performance of the top funds, although the top FCS funds tend to perform a little better on average during months with a higher realized value of the momentum factor.

The bottom row of panels in Figure 12 presents a similar set of plots for the worst FCS funds. There is some evidence that the bottom FCS funds performed worse in periods where small stocks underperformed large stocks (left panel), value stocks outperformed growth stocks (middle panel) and during times with a high realized value of the momentum factor. The variation in the bottom FCS funds' performance across different realizations of the risk factors tends to be weaker than that identified for the top FCS funds, however.

6.2. Industry Concentrations

Table 8 reports time series averages of industry concentrations for the top FCS funds, the bottom FCS funds, and average across all funds. The industry concentration are presented for all periods, expansions only, and for recessions only. We find that the funds in the top FCS have been more concentrated in business services, computers, and electronic equipment, than the funds in the bottom FCS and the average funds.

To get a sense of the dynamic industry tilting strategies implied by the portfolios of superior and inferior funds, Figure 13 plots one-year rolling averages of industry concentrations for the top FCS funds and the average fund. Our analysis is based on an industry classification that uses 48 industries and we show results for the eight industries with the largest (absolute) difference in industry concentration (relative to the average). The top funds greatly overweight computer and electronic equipment stocks after 2005 and also overweight business services and machinery during shorter spells. Conversely, the top funds underweight retail stocks and, in particular, banking stocks throughout most of the sample.

Very different industry concentrations emerge for the inferior FCS funds. As shown in Figure 14, these funds overweight apparel, computer, machinery and electronic equipment stocks in the early nineties, while conversely banking, business services and pharmaceutical products are underweighted during this period. Towards the end of the sample banking and communication shares get overweighted by the inferior FCS funds.

7. Conclusion

This paper presents a new approach to selecting funds with superior performance. By conducting a large set of pair-wise performance comparisons across a large set of mutual funds, the approach iteratively eliminates funds with inferior performance. In line with the finding in recent studies that only a relatively narrow–and declining–set of funds is capable of outperforming on a risk-adjusted basis, our results suggest that it is important to choose a stringent procedure that is capable of eliminating funds with inferior performance.

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

Second, and perhaps somewhat surprisingly, despite the considerable sampling error surrounding estimates of individual funds' alphas, our results show that funds' track records can be sufficiently informative to make it possible to discriminate superior from inferior funds. Consistent with the importance of having a good performance model, the sharpest empirical results–with alpha estimates exceeding 50 bps/month for a portfolio of superior funds–are obtained when we combine returns and holdings data to obtain a conditional (time-varying) alpha estimate.

Third, the proportion of funds–as well as the identify of the individual funds–deemed to be superior varies considerably over time and generally is far smaller than the 5-10% often assumed in studies that use portfolio decile sorts to gauge top funds' performance. Moreover, the fraction of funds deemed to be superior is correlated with a range of variables commonly used to measure the state of the economy.

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

Appendix: Forecasts of fund performance

This appendix explains how we estimate the unknown parameters of the latent state models and use these models to generate predictions of the conditional alpha. For each fund, i, we observe a sample of excess returns, R_{it}. We then cast the return model into state space form as follows:

$$R_{it} = G_{it}(F_{it-1}) + \varepsilon_{it},$$
$$F_{it} = \nu_i F_{it-1} + \eta_{it}.$$

We focus on models where R_{it} is a linear function of the signal. Define $\tilde{F}_{it}$ and $\tilde{P}_{it}$ 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 $\tilde{F}_{it-1}$,

$$G_{it}(F_{it-1}) \approx G_{it}(\tilde{F}_{it-1}) + G_{F,it}(F_{it-1} - \tilde{F}_{it-1}),$$

where

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

Given starting values for $\tilde{F}_{i0}$ and $\tilde{P}_{i0}$, the following recursions constitute the extended Kalman filter:

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

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

$$\hat{F}_{i,t} = \hat{F}_{i,t}(\hat{\theta}_{i\tau})$$

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

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

The forecast of alpha at time $\tau + 1$, given information at time τ, is therefore a function of the forecast of the signal and the parameter estimates available at time τ.

References

Avramov, D., Wermers, R., 2006. Investing in mutual funds when returns are predictable. Journal of Financial Economics 81, 339–377.

Banegas, A., Gillen, B., Timmermann, A., Wermers, R., 2013. The cross-section of conditional mutual fund performance in European stock markets. Journal of Financial Economics 108, 269–726.

Barras, L., O. Scaillet, and R. Wermers, 2010, False Discoveries in Mutual Fund Performance: Measuring Luck in Estimated Alphas. Journal of Finance 65, 179-216.

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

Blake, D., A.G. Rossi, A. Timmermann, I. Tonks, and R. Wermers, 2013, Decentralized Investment Management: Evidence from the Pension Fund Industry. Journal of Finance 68, 1133-1178.

Bollen, N., Busse, J., 2004. Short-term persistence in mutual fund performance. Review of Financial Studies 18, 569–597.

Cai, Z., 2007. Trending time-varying coefficient time series models with serially correlated errors. Journal of Econometrics 136, 163–188.

Carhart, M., 1997. On persistence in mutual fund performance. Journal of Finance 52, 569–597.

Chen, B., Hong, Y., 2012. Testing for smooth structural changes in time series models via nonparametric regression. Econometrica 80, 1157–1183.

Christopherson, J., Ferson, W., Glassman, D., 1998. Conditioning manager alphas on economic information: Another look at the persistence of performance. Review of Financial Studies 11, 111–142.

Daniel, K., Grinblatt, M., Titman, S., Wermers, R., 1997. Measuring mutual fund performance with characteristic-based benchmarks. Journal of Finance 52, 1035–1058.

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

Doornik, J., 2006. Object-Oriented Matrix Programming Using Ox, 3rd Edition. London: Timberlake Consultants Press and Oxford: www.doornik.com.

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.

Kacperczyk, M., van Nieuwerburgh, S., Veldkamp, L., 2014. Time-varying fund manager skill. Journal of Finance 69, 1455-1484.

Kosowski, R., 2011. Do mutual funds perform when it matters most to investors. Quarterly Journal of Finance 1, 607–664.

Liu, J., Timmermann, A., 2013. Optimal convergence trade strategies. Review of Financial Studies 26 (4), 1048–1086.

Mamaysky, H., Spiegel, M., Zhang, H., 2007. Improved forecasting of mutual fund alphas and betas. Review of Finance 11, 359–400.

Mamaysky, H., Spiegel, M., Zhang, H., 2008. Estimating the dynamics of mutual fund alphas and betas. Review of Financial Studies 21 (1), 233–264.

Nyblom, J., 1989. Testing for the constancy of parameters over time. Journal of the American Statistical Association 84, 223–230.

Pastor, L., R.F. Stambaugh, and L.A. Taylor, 2015, Scale and Skill in Active Management. Journal of Financial Economics 116, 23-45.

Patton, A., Timmermann, A., 2010. Monotonicity in asset returns: New tests with applications to the term structure, the CAPM and portfolio sorts. Journal of Financial Economics 98, 605–625.

Romano, J., Wolf, M., 2005. Stepwise multiple testing as formalized data snooping. Econometrica 73 (4), 1237–1282.

Welch, I., and A. Goyal, 2008. A comprehensive look at the empirical performance of equity premium prediction. The Review of Financial Studies 21 (4), 1455–1508.

Wermers, R., 2000, Mutual Fund Performance: An Empirical Decomposition into Stock-Picking Talent, Style, Transactions Costs, and Expenses. Journal of Finance 55, 1655-1695.

White, H., 2000. A reality check for data snooping. Econometrica 68 (5), 1097–1126.

Table 1: Cross-section of alpha estimates

	Benchmark	Latent State	Latent State Holdings
Mean	-0.057	-0.057	-0.066
Min	-1.114	-1.168	-1.610
Max	0.942	0.813	1.713
5%	-0.361	-0.370	-0.376
10%	-0.272	-0.279	-0.273
25%	-0.166	-0.162	-0.157
50%	-0.058	-0.051	-0.062
75%	0.053	0.056	0.034
90%	0.169	0.167	0.136
95%	0.255	0.244	0.211

This table shows the cross-sectional distribution of four-factor alpha estimates obtained from monthly returns data on U.S. equity mutual funds over the period 1980:06-2012:12. 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 three columns report results for different performance models. The benchmark model assumes that alpha is constant and estimated by OLS using a 60-month rolling window. The latent skill (LS) model allows for a time-varying alpha and uses returns data to compute the expected value of each fund's alpha. The latent skill-holdings (LSH) model also allows for a time-varying alpha but combines returns and holdings data to estimate each fund's alpha. The extended Kalman filter is used to extract the latent signal underlying the models with time-varying alphas. The reported estimates of the cross-sectional distribution is based on the individual funds' average alpha estimates.

Table 2: Risk-adjusted performance for alpha-ranked decile portfolios

Portfolio	Benchmark		Latent State		Latent State Holdings	
P1A	-0.22	(-2.66)	-0.25	(-3.65)	-0.18	(-2.51)
P1B	-0.16	(-2.53)	-0.17	(-2.66)	-0.14	(-2.24)
P1	-0.19	(-2.74)	-0.21	(-3.32)	-0.16	(-2.51)
P2	-0.14	(-2.46)	-0.15	(-2.57)	-0.14	(-2.51)
P3	-0.08	(-1.60)	-0.10	(-2.06)	-0.10	(-2.15)
P4	-0.09	(-1.76)	-0.08	(-1.75)	-0.09	(-1.86)
P5	-0.07	(-1.44)	-0.08	(-1.74)	-0.09	(-2.25)
P6	-0.07	(-1.45)	-0.08	(-1.76)	-0.06	(-1.43)
P7	-0.03	(-0.65)	-0.05	(-1.15)	-0.05	(-1.23)
P8	-0.05	(-1.08)	-0.03	(-0.61)	-0.02	(-0.45)
P9	-0.02	(-0.42)	0.04	(0.75)	0.00	(0.04)
P10	0.06	(0.83)	0.05	(0.70)	0.02	(0.31)
P10A	0.00	(0.07)	0.05	(0.75)	-0.00	(-0.04)
P10B	0.11	(1.26)	0.05	(0.57)	0.05	(0.59)
P10-P1	0.25	(3.46)	0.26	(4.94)	0.18	(4.00)
MR	0.07		0.00		0.00	

This table reports four-factor alphas for a set of alpha-ranked decile portfolios. Each month we rank the set of mutual funds in existence according to their expected alphas and allocate them to equal-weighted decile portfolios listed in increasing order from funds with the lowest alphas ($P1$) to funds with the highest alphas ($P10$). The alpha ranking is repeated each month during the sample 1985:07-2012:12 and so produces a time series of portfolio returns from which the reported alpha estimates are computed. The columns report results for three different performance models. The benchmark model assumes that alpha is constant and estimated by OLS using a 60-month rolling window. The latent skill (LS) model allows for a time-varying alpha and uses returns data to compute the expected value of next month's alpha. The latent skill-holdings (LSH) model also allows for a time-varying alpha but combines returns and holdings data to estimate each fund's alpha. $P1A$ and $P1B$ are constructed by splitting the funds in the bottom decile portfolio into two new portfolios of equal size. $P10A$ and $P10B$ are constructed in the same way. $P10 - P1$ is a portfolio constructed from the difference between the top and bottom decile portfolios. The table reports alpha estimates in percentage terms followed by t-statistics in brackets. If a particular performance model provides an accurate ranking of the funds we would expect the alpha estimates to increase monotonically from $P1$ to $P10$. We test for this monotonically increasing pattern against a flat or a decreasing relation using the monotonic relation (MR) test proposed by Patton and Timmermann (2010). Low p-values are evidence against a flat or decreating patter and so are suggestive of a precise ranking of alphas.

Table 3: Risk-adjusted performance for FCS portfolios of superior funds

Panel A: Top 15 percent of funds

	Benchmark		Latent State		Latent State Holdings	
	Equal	Value	Equal	Value	Equal	Value
FCS-tight	0.30 (1.24)	0.36 (1.41)	0.25 (1.94)	0.25 (1.81)	0.54 (2.13)	0.52 (2.08)
FCS-medium	0.27 (1.59)	0.23 (1.33)	0.15 (1.53)	0.14 (1.24)	0.18 (1.35)	0.18 (1.24)
FCS-wide	0.20 (1.51)	0.19 (1.38)	0.10 (1.05)	0.02 (0.16)	0.06 (0.51)	-0.02 (-0.14)
P10	0.06 (0.83)	0.07 (0.89)	0.05 (0.70)	0.04 (0.61)	0.02 (0.31)	0.07 (1.06)

Panel B: Top 5 percent of funds

	Benchmark		Latent State		Latent State Holdings	
	Equal	Value	Equal	Value	Equal	Value
FCS-tight	0.36 (1.34)	0.39 (1.41)	0.10 (0.57)	0.10 (0.50)	0.48 (1.77)	0.47 (1.74)
FCS-medium	0.45 (2.28)	0.41 (2.02)	0.14 (1.06)	0.18 (0.99)	0.36 (1.92)	0.43 (2.04)
FCS-wide	0.36 (2.03)	0.28 (1.56)	0.08 (0.63)	0.06 (0.40)	0.24 (1.61)	0.21 (1.40)
P10	0.06 (0.83)	0.07 (0.89)	0.05 (0.70)	0.04 (0.61)	0.02 (0.31)	0.07 (1.06)

Panel C: Top 25 percent of funds

	Benchmark		Latent State		Latent State Holdings	
	Equal	Value	Equal	Value	Equal	Value
FCS-tight	0.30 (1.33)	0.35 (1.50)	0.17 (1.60)	0.16 (1.44)	0.38 (1.94)	0.38 (1.91)
FCS-medium	0.19 (1.24)	0.20 (1.30)	0.12 (1.48)	0.11 (1.20)	0.16 (1.18)	0.17 (1.20)
FCS-wide	0.14 (1.20)	0.14 (1.16)	0.08 (1.07)	0.04 (0.51)	0.05 (0.48)	-0.00 (-0.04)
P10	0.06 (0.83)	0.07 (0.89)	0.05 (0.70)	0.04 (0.61)	0.02 (0.31)	0.07 (1.06)

Panel D: Romano and Wolf

	Benchmark		Latent State		Latent State Holdings	
	Equal	Value	Equal	Value	Equal	Value
FCS-tight	0.20 (0.95)	0.28 (1.30)	0.08 (1.05)	0.11 (1.19)	0.48 (2.02)	0.53 (2.19)
FCS-medium	0.14 (1.04)	0.18 (1.30)	0.10 (1.51)	0.10 (1.36)	0.14 (1.01)	0.13 (0.85)
FCS-wide	0.08 (0.61)	0.07 (0.59)	0.06 (0.94)	0.09 (1.29)	0.13 (1.02)	0.11 (0.85)
P10	0.06 (0.83)	0.07 (0.89)	0.05 (0.70)	0.04 (0.61)	0.02 (0.31)	0.07 (1.06)

This table reports four-factor alphas for a set of portfolios formed by equal- or value-weighting stocks identified as being superior by the fund confidence set (FCS) approach. Each month we apply the FCS approach to all funds in existence so as to identify the set of superior funds. We then form a portfolio of these funds and record its return during the following month. This procedure is repeated each month during the sample 1985:07-2012:12 and so produces a time series of portfolio returns from which the reported four-factor alpha estimates are computed. In each panel the top three rows report results for three 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 and third rows give rise to medium ($\lambda = 0.50$) and wide ($\lambda = 0.10$) fund confidence sets. For comparison we also show results for the alpha-ranked top decile portfolio ($P10$). The FCS approach requires us to choose an initial set of candidate funds from which the superior funds are identified. Panel A uses the funds in the top 15 percent of the distribution of alpha estimates, while Panels B and C use the top 5 and top 25 percent of funds ranked by their expected alphas. Panel D uses the set of funds with significantly positive alpha, identified using the approach of Romano and Wolf (2005). The columns report results for three different performance models. The benchmark model assumes that alpha is constant and estimated by OLS using a 60-month rolling window. The latent skill (LS) model allows for a time-varying alpha and uses returns data to compute the expected value of next month's alpha. The latent skill-holdings (LSH) model also allows for a time-varying alpha but combines returns and holdings data to estimate each fund's alpha. Alpha estimates are reported in percent per month with t-statistics reported in brackets.

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

	Benchmark		Latent State		Latent State Holdings	
	Equal	Value	Equal	Value	Equal	Value
Panel A: Bottom 15 percent of funds						
FCS-tight	-0.43 (-3.04)	-0.41 (-2.78)	-0.36 (-3.23)	-0.27 (-2.49)	-0.32 (-2.48)	-0.26 (-1.74)
FCS-medium	-0.31 (-2.58)	-0.19 (-1.55)	-0.27 (-3.45)	0.01 (0.11)	-0.22 (-2.99)	-0.19 (-1.75)
FCS-wide	-0.21 (-3.09)	-0.14 (-1.78)	-0.25 (-3.32)	-0.01 (-0.12)	-0.19 (-2.99)	-0.15 (-1.54)
P1	-0.19 (-2.73)	-0.18 (-2.09)	-0.21 (-3.32)	-0.12 (-1.66)	-0.16 (-2.51)	-0.16 (-1.87)
Panel B: Bottom 5 percent of funds						
	Benchmark		Latent State		Latent State Holdings	
	Equal	Value	Equal	Value	Equal	Value
FCS-tight	-0.57 (-3.85)	-0.57 (-3.92)	-0.56 (-3.41)	-0.44 (-2.50)	-0.45 (-2.99)	-0.49 (-2.75)
FCS-medium	-0.49 (-3.82)	-0.45 (-3.47)	-0.41 (-4.14)	-0.14 (-1.23)	-0.37 (-3.97)	-0.30 (-2.28)
FCS-wide	-0.40 (-4.04)	-0.39 (-4.37)	-0.36 (-3.78)	-0.00 (-0.04)	-0.29 (-3.92)	-0.29 (-2.56)
P1	-0.19 (-2.73)	-0.18 (-2.09)	-0.21 (-3.32)	-0.12 (-1.66)	-0.16 (-2.51)	-0.16 (-1.87)
Panel C: Bottom 25 percent of funds						
	Benchmark		Latent State		Latent State Holdings	
	Equal	Value	Equal	Value	Equal	Value
FCS-tight	-0.40 (-2.82)	-0.35 (-2.31)	-0.21 (-2.12)	-0.17 (-1.58)	-0.26 (-2.68)	-0.16 (-1.44)
FCS-medium	-0.35 (-3.39)	-0.23 (-2.14)	-0.24 (-3.37)	-0.06 (-0.75)	-0.21 (-3.07)	-0.19 (-1.99)
FCS-wide	-0.18 (-2.46)	-0.10 (-1.33)	-0.23 (-3.44)	-0.04 (-0.59)	-0.19 (-3.08)	-0.13 (-1.60)
P1	-0.19 (-2.73)	-0.18 (-2.09)	-0.21 (-3.32)	-0.12 (-1.66)	-0.16 (-2.51)	-0.16 (-1.87)
Panel D: Romano and Wolf						
	Benchmark		Latent State		Latent State Holdings	
	Equal	Value	Equal	Value	Equal	Value
FCS-tight	-0.45 (-3.36)	-0.40 (-2.81)	-0.20 (-2.77)	-0.11 (-1.26)	-0.23 (-2.31)	-0.17 (-1.55)
FCS-medium	-0.25 (-3.35)	-0.11 (-1.35)	-0.20 (-3.01)	-0.11 (-1.54)	-0.15 (-2.20)	-0.13 (-1.52)
FCS-wide	-0.14 (-2.28)	-0.02 (-0.34)	-0.17 (-3.22)	-0.05 (-0.97)	-0.20 (-3.59)	-0.16 (-2.04)
P1	-0.19 (-2.73)	-0.18 (-2.09)	-0.21 (-3.32)	-0.12 (-1.66)	-0.16 (-2.51)	-0.16 (-1.87)

This table reports four-factor alphas for a set of portfolios formed by equal- or value-weighting stocks identified as being inferior by the fund confidence set (FCS) approach. Each month we apply the FCS approach to all funds in existence so as to identify the set of inferior funds. We then form a portfolio of these funds and record its return during the following month. This procedure is repeated each month during the sample 1985:07-2012:12 and so produces a time series of portfolio returns from which the reported four-factor alpha estimates are computed. In each panel the top three rows report results for three 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 and third rows give rise to medium ($\lambda = 0.50$) and wide ($\lambda = 0.10$) fund confidence sets. For comparison we also show results for the alpha-ranked bottom decile portfolio ($P1$). The FCS approach requires us to choose an initial set of candidate funds from which the inferior funds are identified. Panel A uses the funds in the top 15 percent of the distribution of alpha estimates, while Panels B and C use the top 5 and top 25 percent of funds ranked by their expected alphas. Panel D uses the set of funds with significantly positive alpha, identified using the approach of Romano and Wolf (2005). The columns report results for three different performance models. The benchmark model assumes that alpha is constant and estimated by OLS using a 60-month rolling window. The latent skill (LS) model allows for a time-varying alpha and uses returns data to compute the expected value of next month's alpha. The latent skill-holdings (LSH) model also allows for a time-varying alpha but combines returns and holdings data to estimate each fund's alpha. Alpha estimates are reported in percent per month with t-statistics reported in brackets.

Table 5: Risk-adjusted performance for FCS portfolios in expansions and recessions

Panel A: Superior Funds

	Benchmark	Latent State	Latent State Holdings
		Overall	
FCS	0.30 (1.24)	0.25 (1.94)	0.54 (2.13)
P10	0.06 (0.83)	0.05 (0.70)	0.02 (0.31)
		Expansions	
FCS	0.26 (0.98)	0.30 (2.14)	0.66 (2.53)
P10	0.07 (0.85)	0.05 (0.77)	0.03 (0.53)
		Recessions	
FCS	0.06 (0.07)	0.26 (0.85)	-0.27 (-0.48)
P10	-0.06 (-0.37)	-0.05 (-0.29)	0.02 (0.09)

Panel B: Inferior Funds

	Benchmark	Latent State	Latent State Holdings
		Overall	
FCS	-0.43 (-3.04)	-0.36 (-3.23)	-0.32 (-2.48)
P1	-0.19 (-2.73)	-0.21 (-3.32)	-0.16 (-2.51)
		Expansions	
FCS	-0.44 (-3.23)	-0.35 (-2.75)	-0.27 (-2.24)
P1	-0.22 (-3.34)	-0.24 (-3.68)	-0.19 (-3.09)
		Recessions	
FCS	-0.39 (-0.56)	-1.20 (-3.31)	-1.05 (-2.09)
P1	-0.33 (-1.97)	-0.27 (-1.90)	-0.25 (-1.54)

This table reports four-factor alphas in expansions and recessions for a set of portfolios formed by equal-weighting stocks identified as being superior (Panel A) or inferior (Panel B) by the fund confidence set (FCS) approach. Each month we apply the FCS approach to all funds in existence so as to identify the set of superior or inferior funds. We then form a portfolio of these funds and record its return during the following month. This procedure is repeated each month during the sample 1985:07-2012:12 and so produces a time series of portfolio returns from which the reported four-factor alpha estimates are computed. The FCS portfolios are equally weighted, selected from the top 15 percent of alpha-ranked funds and set $\lambda = 0.90$), resulting in a narrower set of funds being included. For comparison we also show results for the alpha-ranked bottom decile portfolio ($P1$). The columns report results for three different performance models. The benchmark model assumes that alpha is constant and estimated by OLS using a 60-month rolling window. The latent skill (LS) model allows for a time-varying alpha and uses returns data to compute the expected value of next month's alpha. The latent skill-holdings (LSH) model also allows for a time-varying alpha but combines returns and holdings data to estimate each fund's alpha. The sample is split into recession and expansion states based on the NBER recession indicator. Alpha estimates are reported in percent per month with t-statistics reported in brackets.

Table 6: Size of fund confidence set and macroeconomic state variables

	Top FCS		Bottom FCS	
State variable	Number of funds	Percentage of funds	Number of funds	Percentage of funds
Constant	-4.14 (-0.59)	-0.00 (-0.28)	30.75 (2.23)	0.12 (2.23)
tbl	-76.76 (-1.47)	0.45 (2.65)	-172.75 (-1.42)	0.76 (2.70)
tms	175.72 (2.25)	1.23 (4.05)	-213.10 (-1.42)	0.30 (0.87)
dfy	-457.72 (-3.13)	-0.88 (-2.38)	966.30 (2.62)	1.33 (1.65)
dy	-3.68 (-2.25)	0.00 (0.67)	3.71 (1.26)	0.04 (2.53)

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

Table 7: Attributes of funds selected by the FCS

	Overall			Expansions			Recessions		
	Superior	Inferior	Average	Superior	Inferior	Average	Superior	Inferior	Average
CS-1m($\times$100)	0.36	-0.03	0.03	0.49	0.00	0.04	-0.51	-0.25	0.03
CS-3m($\times$100)	0.42	-0.15	0.04	0.55	-0.11	0.05	-0.50	-0.45	0.04
TNA	3381.16	508.36	1765.90	3563.31	557.77	1800.89	2116.82	165.45	1765.90
Size	4.20	3.88	4.14	4.16	4.00	4.14	4.52	3.06	4.14
BTM	2.84	2.97	2.89	2.87	2.98	2.89	2.64	2.86	2.89
MOM	3.17	3.19	3.14	3.19	3.18	3.13	3.02	3.29	3.14
RetGap12($\times$100)	-0.06	-0.05	-0.04	-0.13	-0.06	-0.07	0.42	0.02	-0.04
Flow($\times$100)	2.72	-0.85	1.40	3.33	-0.50	1.69	-1.50	-3.29	1.40
ExpRatio($\times$100)	1.13	1.65	1.22	1.17	1.65	1.23	0.88	1.63	1.22
Turnover	1.06	1.39	0.83	1.13	1.40	0.83	0.59	1.29	0.83

This table reports several attributes of the funds selected by the FCS procedure and assumes that $\lambda = 0.9$. The table reports attributes for the FCS portfolio of superior funds, the FCS portfolio of inferior funds and the average across funds. Attributes are calculated by first taking cross-sectional averages of attributes of the funds in the portfolio at a given point in time, then taking time series averages. The attributes include the portfolio-weighted characteristic selectivity measure, one month (CS-1m) and three months (CS-3m) in percent, total net assets of funds in millions of dollars (TNA), portfolio holdings-based style attributes in the size (Size), book-to-market (BTM), and momentum (MOM) dimensions, lagged net return, compounded over the 12 months prior to each portfolio formation date (RetGap12), monthly percentage net cash inflows (Flow), computed as TNA minus one-quarter-lagged TNA, divided by three, fund expense ratio (ExpRat), and percent monthly turnover (Turnover). These attributes are presented for all periods, for expansions only and for recessions only.

Table 8: Industry concentrations of funds selected by the FCS

	Overall			Expansions			Recessions		
	Superior	Inferior	Average	Superior	Inferior	Average	Superior	Inferior	Average
Agriculture	0.09	0.20	0.21	0.11	0.16	0.19	0.00	0.49	0.21
Food Products	0.84	1.36	1.56	0.75	1.32	1.53	1.48	1.62	1.56
Candy & Soda	0.16	0.27	0.11	0.17	0.29	0.11	0.09	0.12	0.11
Beer & Liquor	0.82	1.12	0.95	0.92	1.21	0.93	0.12	0.46	0.95
Tobacco Products	0.25	0.22	0.34	0.28	0.24	0.34	0.04	0.05	0.34
Recreation	0.24	0.33	0.34	0.27	0.34	0.35	0.04	0.28	0.34
Entertainment	1.56	0.88	0.99	1.27	0.94	1.01	3.61	0.48	0.99
Printing & Publishing	1.67	0.74	1.00	1.51	0.82	1.04	2.85	0.17	1.00
Consumer Goods	1.17	2.76	1.98	1.20	2.83	1.96	0.90	2.24	1.98
Apparel	0.61	1.88	0.85	0.69	1.97	0.84	0.04	1.24	0.85
Healthcare	0.90	1.26	1.31	1.01	1.22	1.33	0.08	1.52	1.31
Medical Equipment	1.19	3.12	2.13	1.13	2.61	2.10	1.65	6.78	2.13
Pharmaceutical Products	3.96	6.46	7.01	4.29	6.34	6.84	1.60	7.27	7.01
Chemicals	1.18	1.82	2.36	1.27	1.79	2.37	0.55	2.05	2.36
Rubber and Plastic Products	0.21	0.32	0.33	0.16	0.23	0.33	0.59	0.94	0.33
Textiles	0.09	0.23	0.25	0.10	0.25	0.26	0.00	0.12	0.25
Construction Materials	0.63	1.20	0.97	0.69	1.27	0.99	0.17	0.75	0.97
Construction	0.22	0.99	0.74	0.25	0.98	0.74	0.00	1.08	0.74
Steel Works Etc	0.58	1.67	1.15	0.61	1.78	1.15	0.31	0.91	1.15
Fabricated Products	0.03	0.10	0.10	0.03	0.09	0.11	0.00	0.17	0.10
Machinery	3.53	3.44	2.89	3.35	3.49	2.87	4.84	3.11	2.89
Electrical Equipment	0.87	1.28	1.50	0.98	1.33	1.51	0.12	0.92	1.50
Automobiles and Trucks	0.97	1.39	1.32	0.98	1.48	1.39	0.84	0.78	1.32
Aircraft	0.44	1.23	1.13	0.49	1.30	1.12	0.10	0.71	1.13
Shipbuilding and Railroad Equipment	0.03	0.10	0.17	0.03	0.10	0.17	0.01	0.16	0.17
Defense	0.12	0.26	0.28	0.13	0.28	0.27	0.01	0.14	0.28
Precious Metals	0.25	0.06	1.15	0.28	0.06	1.18	0.06	0.04	1.15
Non-Metallic and Industrial Metal Mining	0.26	0.30	0.65	0.29	0.31	0.67	0.04	0.24	0.65
Coal	0.17	0.13	0.15	0.19	0.12	0.14	0.00	0.19	0.15
Petroleum and Natural Gas	3.56	4.71	5.81	3.96	4.85	5.64	0.72	3.65	5.81
Utilities	1.04	2.68	4.12	1.14	2.81	4.03	0.30	1.79	4.12
Communication	10.93	4.49	4.79	9.02	4.80	4.78	24.61	2.27	4.79
Personal Services	0.62	0.80	0.65	0.70	0.79	0.63	0.04	0.83	0.65
Business Services	13.66	7.88	8.34	12.93	7.59	8.34	18.90	9.95	8.34
Computers	13.12	7.45	4.82	13.60	7.23	4.83	9.65	9.02	4.82
Electronic Equipment	11.11	6.61	5.35	11.60	6.37	5.43	7.59	8.31	5.35
Measuring and Control Equipment	1.45	1.30	1.30	1.62	1.16	1.29	0.22	2.33	1.30
Business Supplies	0.99	1.35	1.61	1.02	1.46	1.63	0.78	0.61	1.61
Shipping Containers	0.13	0.30	0.27	0.15	0.29	0.26	0.00	0.35	0.27
Transportation	1.79	1.65	2.20	1.68	1.61	2.20	2.60	1.94	2.20
Wholesale	1.19	2.94	1.91	1.34	2.90	1.90	0.14	3.23	1.91
Retail	5.67	6.12	6.47	5.47	6.24	6.49	7.05	5.30	6.47
Restaurants and Hotels and Motels	1.17	1.70	1.22	1.26	1.72	1.22	0.53	1.53	1.22
Banking	4.23	7.33	8.47	4.48	7.70	8.71	2.39	4.69	8.47
Insurance	4.14	4.04	5.43	4.20	4.13	5.52	3.66	3.38	5.43
Real Estate	0.16	0.09	0.25	0.18	0.09	0.25	0.00	0.12	0.25
Trading	1.17	2.60	2.37	1.30	2.39	2.36	0.21	4.12	2.37
Almost Nothing	0.82	0.82	0.72	0.87	0.73	0.68	0.47	1.52	0.72

The table reports average industry concentration of the funds selected by the FCS procedure and assumes that $\lambda = 0.9$. The table reports industry concentrations in percent for the FCS portfolio of superior funds, the FCS portfolio of inferior funds, and the average across funds. Average industry concentrations are calculated by first taking cross-sectional averages of the industry concentrations of the funds in the portfolio at a given point in time, then taking time series averages. These industry concentrations are presented for all periods, for expansions only and for recessions only.

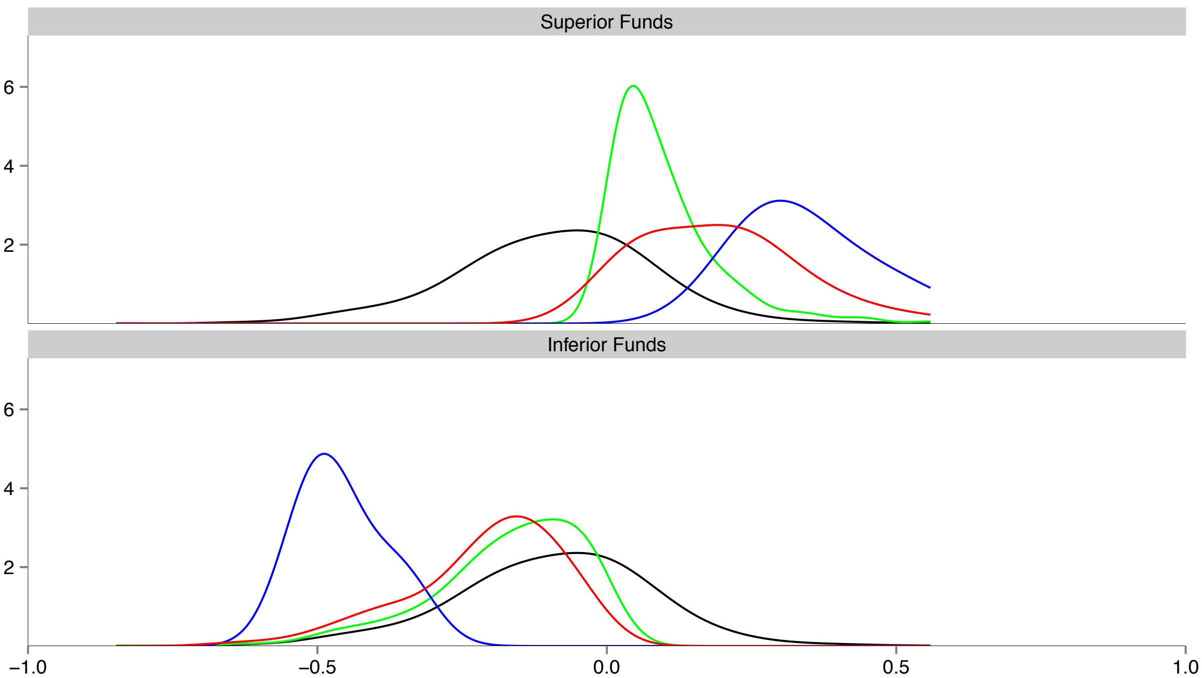

Figure 1: The figure presents distributions of predictive alphas for funds included in equal weighted portfolios consisting of different subsets of funds on a single month in our sample (July 2006) . The top panel presents distributions of predictive alphas for portfolios of superior funds. The black line represents a portfolio consisting of all funds. The green line is for at portfolio with only positive forecasts. The red line is the portfolio based on the Romano Wolf test. The blue line is the FCS where $\lambda = 0.90$ is assumed. The bottom portfolio presents distributions for forecasts included in portfolios of inferior funds. The black line represents a portfolio of all funds. The green line represents a portfolio consisting of funds with negative alpha forecasts. The red line is the portfolio based on the Romano Wolf test. Finally, the blue line represents the FCS portfolio, where $\lambda = 0.90$ is assumed.

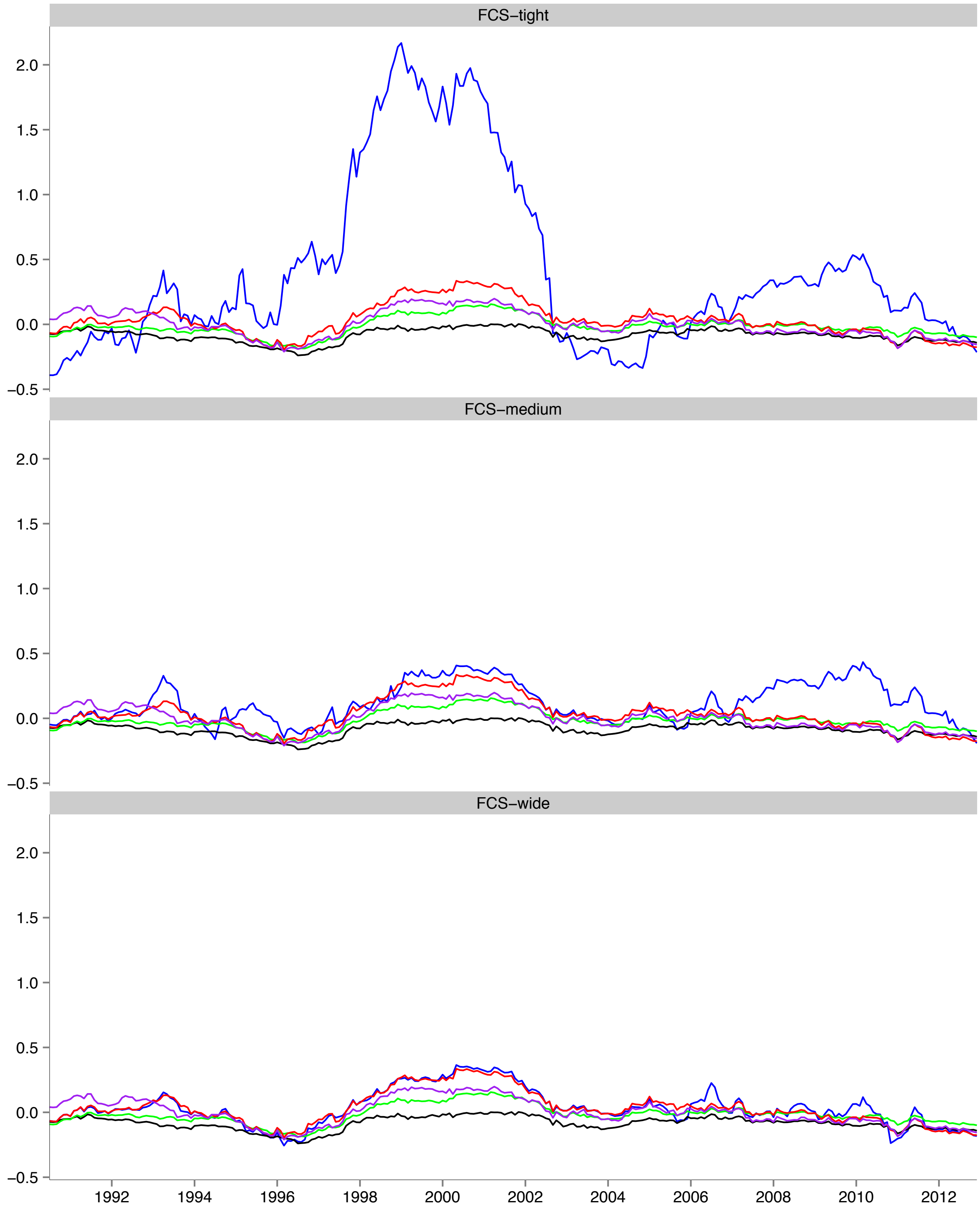

Figure 2: Time variation in alpha estimates for superior funds. This figure shows nonparametric estimates of local time-variation in four-factor alpha estimates (in percent)for three FCS portfolios consisting of funds with superior performance. The top, middle and bottom panels set with $\lambda = 0.10, 0.50, 0.90$, respectively and so correspond to tight, medium and broad fund confidence sets. Blue lines track the performance of the FCS portfolios. For comparison the black line tracks the variation in the alpha of the average fund. Green lines track the performance of a portfolio consisting of all funds with positive forecasts of alpha. Purple lines track the top decile portfolio ($P10$). Finally, red lines track the performance of a portfolio consisting of the top 15% of funds with positive forecasts of alpha. We use the latent skill-holdings model with a time-varying alpha to compute each fund's alpha estimate and thus use both returns and holdings data.

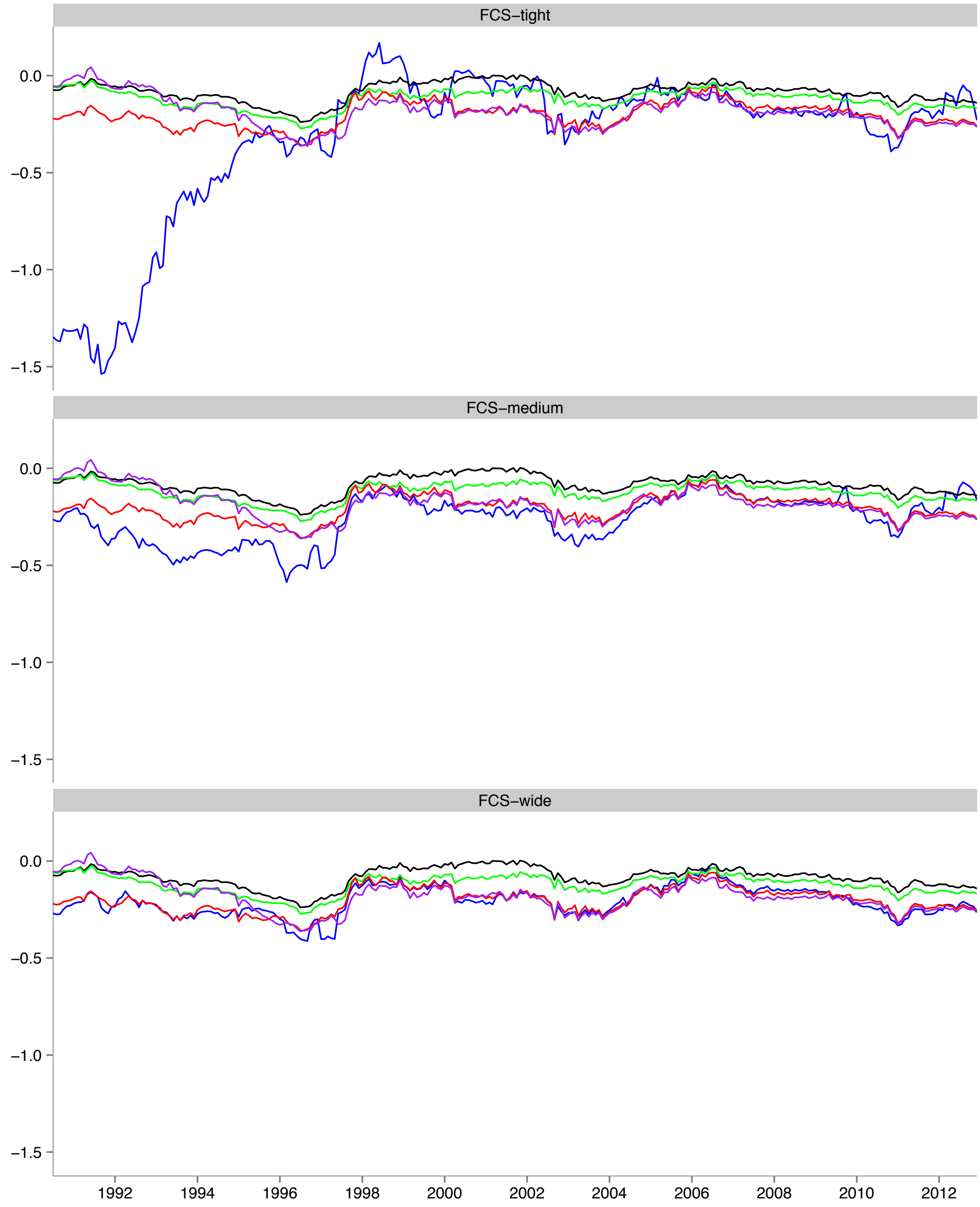

Figure 3: Time variation in alpha estimates for inferior funds. This figure shows nonparametric estimates of local time-variation in four-factor alpha estimates (in percent)for three FCS portfolios consisting of funds with inferior performance. The top, middle and bottom panels set with $\lambda = 0.10, 0.50, 0.90$, respectively and so correspond to tight, medium and broad fund confidence sets. Blue lines track the performance of the FCS portfolios. For comparison the black line tracks the variation in the alpha of the average fund. Green lines track the performance of a portfolio consisting of all funds with negative forecasts of alpha. Purple lines track the bottom decile portfolio ($P1$). Finally, red lines track the performance of a portfolio consisting of the bottom 15% of funds with negative forecasts of alpha. We use the latent skill-holdings model with a time-varying alpha to compute each fund's alpha estimate and thus use both returns and holdings data.

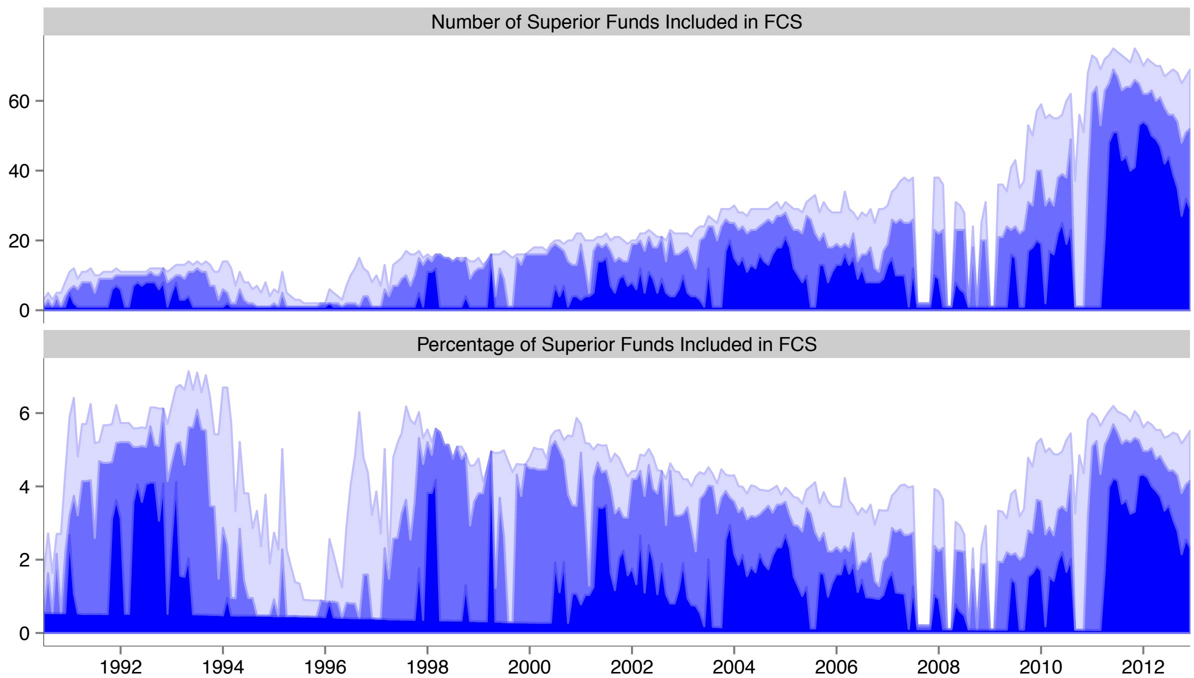

Figure 4: Number of funds in the FCS for superior funds. The top panel shows the evolution in the number of funds included in the tight, medium and wide fund confidence sets. The bottom panel shows the corresponding evolution in the percentage of funds in the FCS. These funds are selected from the top 15 percent of funds ranked by 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 $\lambda = 0.50$ (medium) and $\lambda = 0.10$ (wide set).

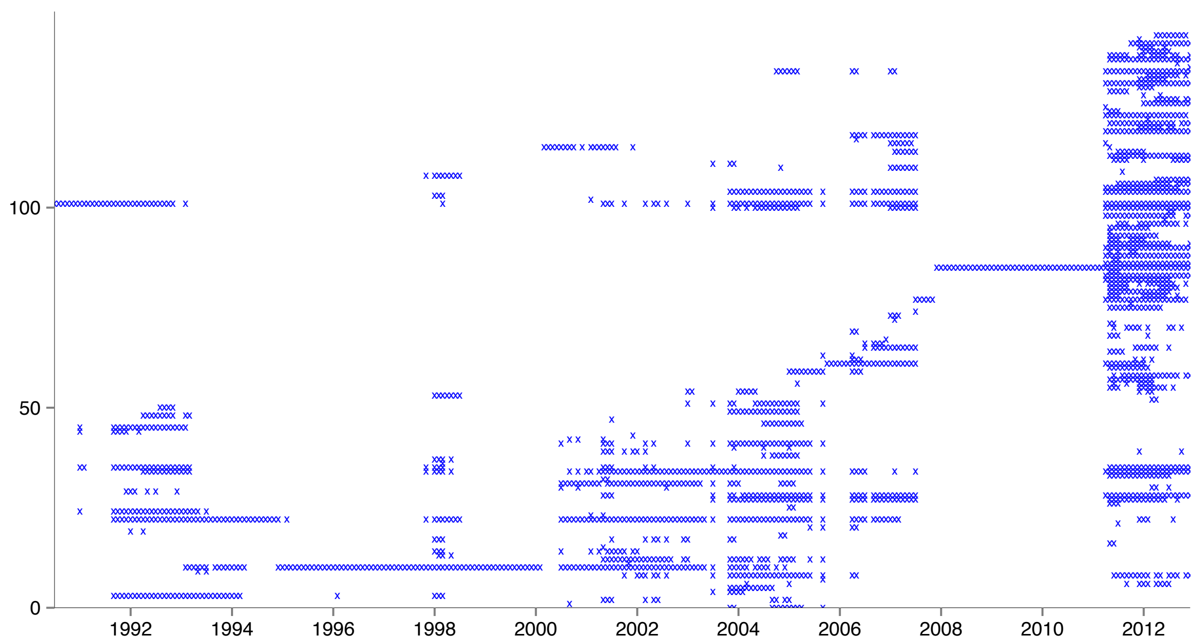

Figure 5: Inclusion of individual funds in the FCS for superior funds. The plot illustrates the composition of the top FCS portfolio. Each fund that is included in the FCS at least once during the sample gets a unique number on the y-axis and a cross shows when this fund is included in the FCS for superior funds. There is no cardinal ordering for the vertical axis. The FCS is based on the time-varying alpha model with latent skill that combines returns and holdings data, considers the top 15 percent of funds as candidate funds and assumes $\lambda = 0.90$.

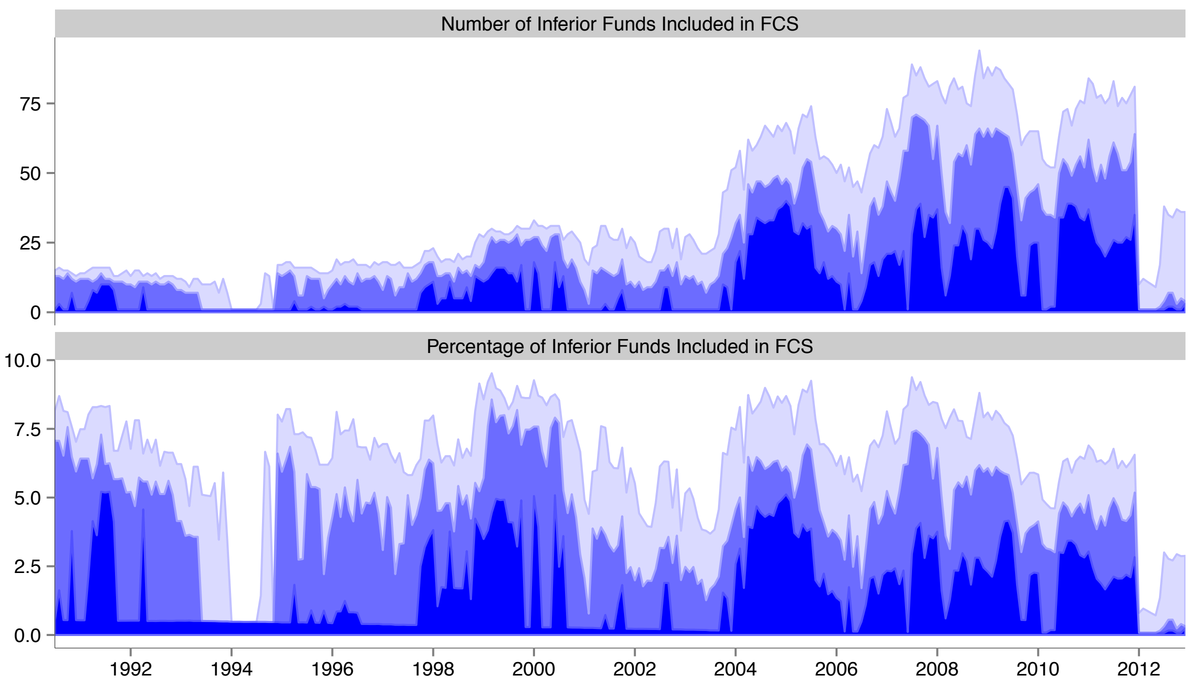

Figure 6: Number of funds in the FCS for inferior funds. The top panel shows the evolution in the number of funds included in the tight, medium and wide fund confidence sets. The bottom panel shows the corresponding evolution in the percentage of funds in the FCS. These funds are selected from the bottom 15 percent of funds ranked by 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 $\lambda = 0.50$ (medium) and $\lambda = 0.10$ (wide set).

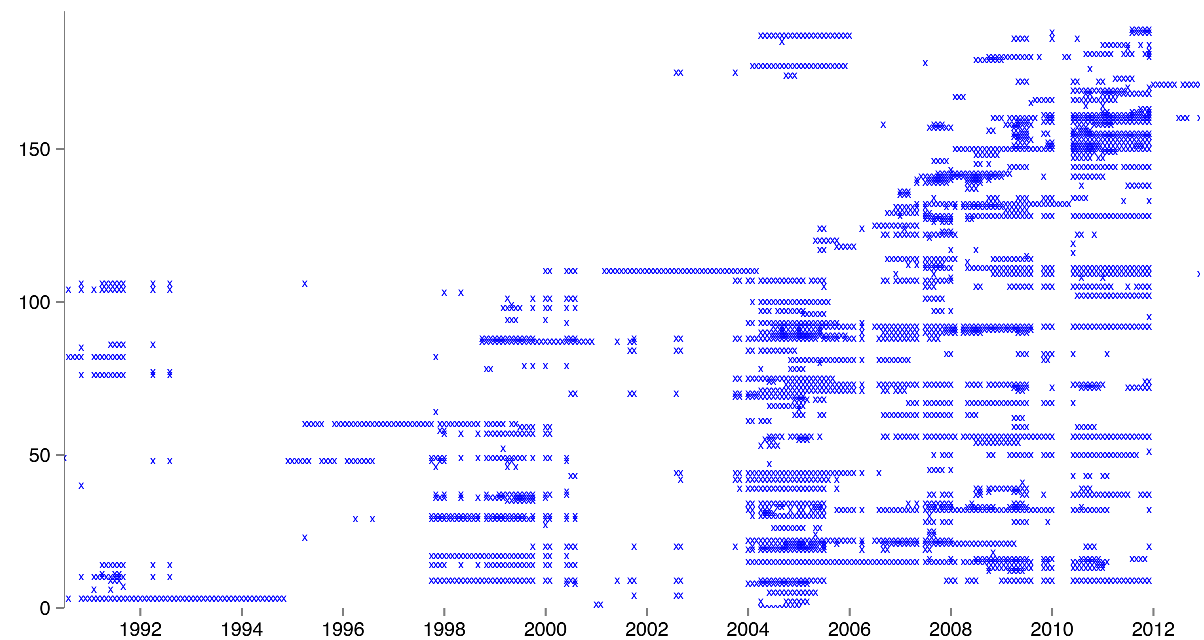

Figure 7: Inclusion of individual funds in the FCS for inferior funds. The plot illustrates the composition of the FCS portfolio for inferior funds. Each fund that is included in the FCS at least once during the sample gets a unique number on the y-axis and a cross shows when this fund is included in the FCS for inferior funds. There is no cardinal ordering for the vertical axis. The FCS is based on the time-varying alpha model with latent skill that combines returns and holdings data, considers the top 15 percent of funds as candidate funds and assumes $\lambda = 0.90$.

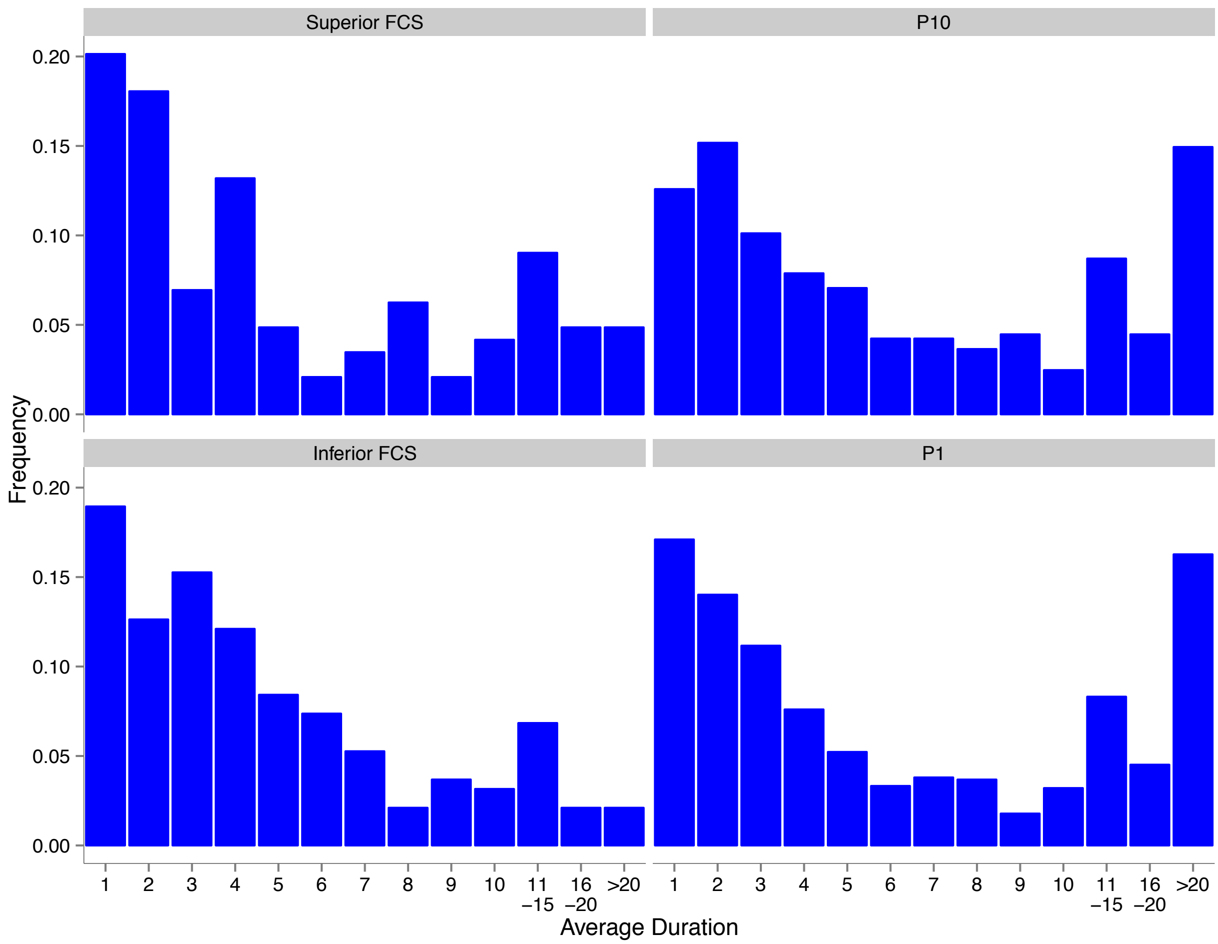

Figure 8: Average duration of funds included in different portfolios. The histograms show the duration (in months) that a fund stays in four different portfolios, namely the FCS portfolio of superior and inferior funds (top left and bottom left corner, respectively), both assumign $\lambda = 0.90$, and the top and bottom decile portfolios (top and bottom right corners, respectively) Only funds that get included in the portfolio at least once are represented in the plot.

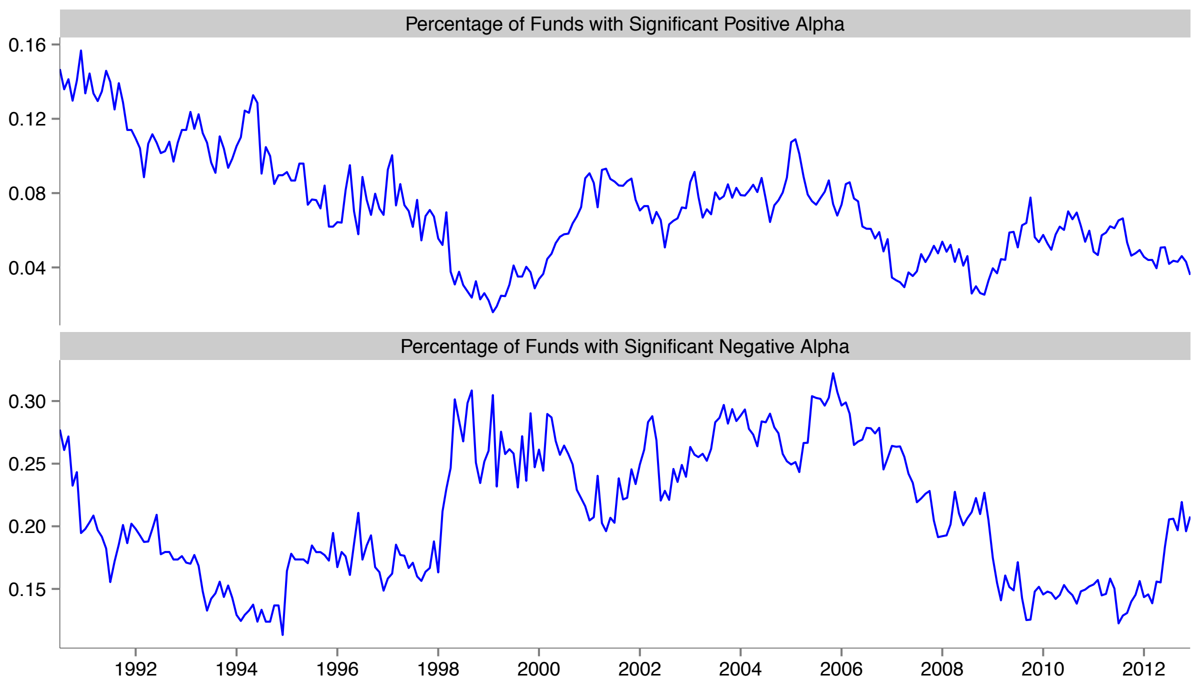

Figure 9: Evolution in the proportion of funds with significantly positive or negative alpha estimates. The top panel plots the proportion of funds with a statistically significant alpha estimate while the bottom panel plots the proportion of funds with a statistically significant negative alpha estimate. Each month we determine the set of funds with significantly positive (or negative) alphas by applying the Romano-Wolf (2005) step-wise bootstrap procedure to the alpha estimates obtained from the latent skill model that combines data on fund returns and holdings.

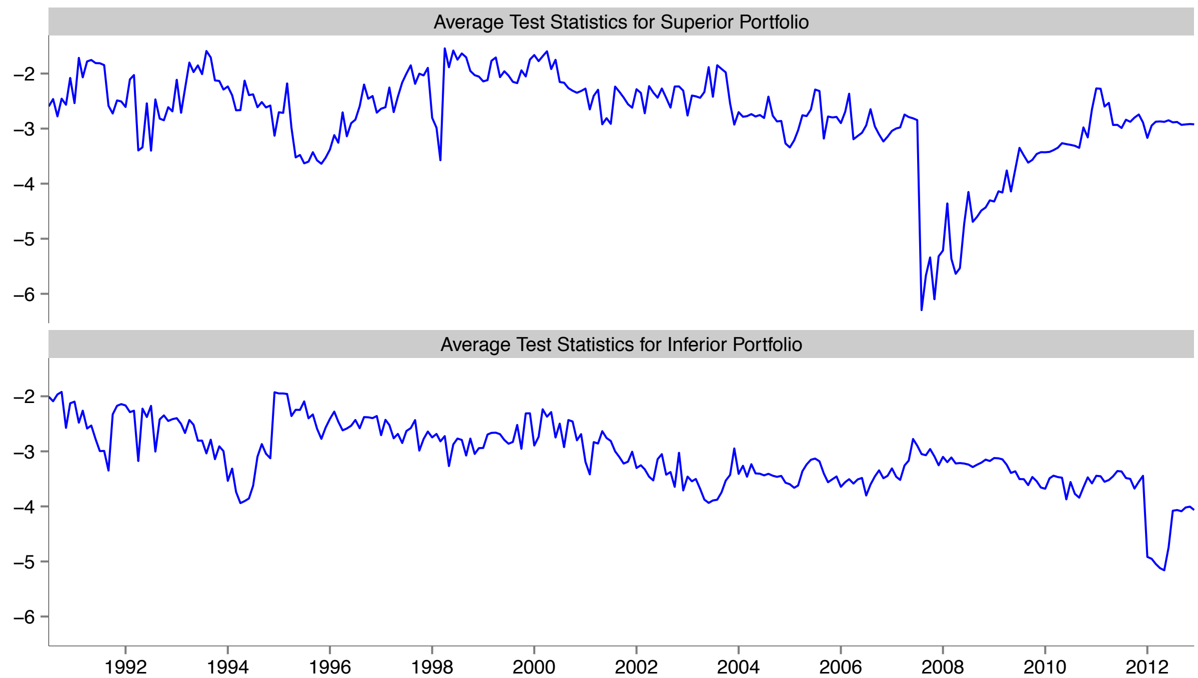

Figure 10: Average value of the test statistic used to eliminate funds from the FCS. The top panel shows the evolution in the average value of the test statistics used to exclude funds from the FCS for superior funds. The bottom panel shows the corresponding values of the test statistics used to exclude superior funds from the FCS for inferior funds. In both cases we assume $\lambda = 0.90$.

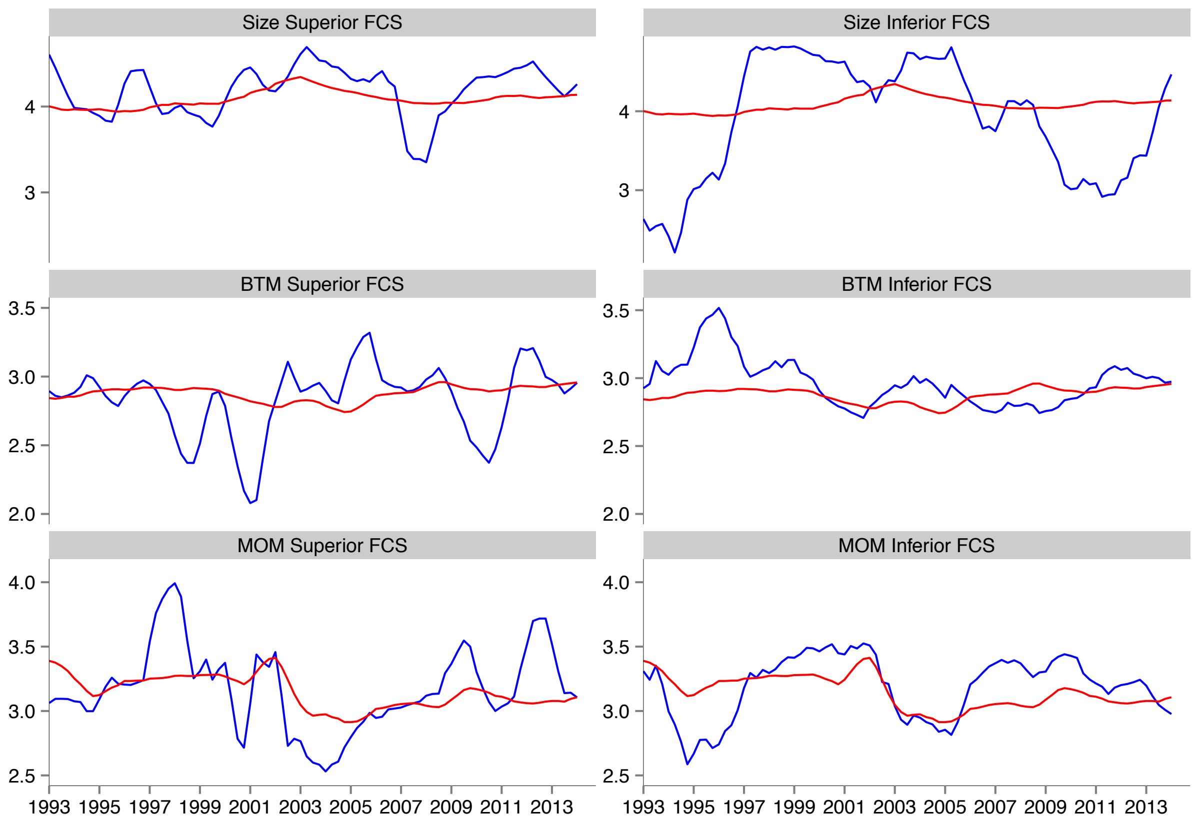

Figure 11: Loadings on risk factors. The left column shows one-year rolling averages of exposures to the size, book-to-market and momentum risk factors for the FCS portfolio consisting of superior funds. The left column shows one-year rolling averages of exposures to the same risk factors for the FCS portfolio consisting of inferior funds. The plots assume $\lambda = 0.90$ and are based on the time-varying alpha model that combines data on fund returns and holdings to estimate alphas.

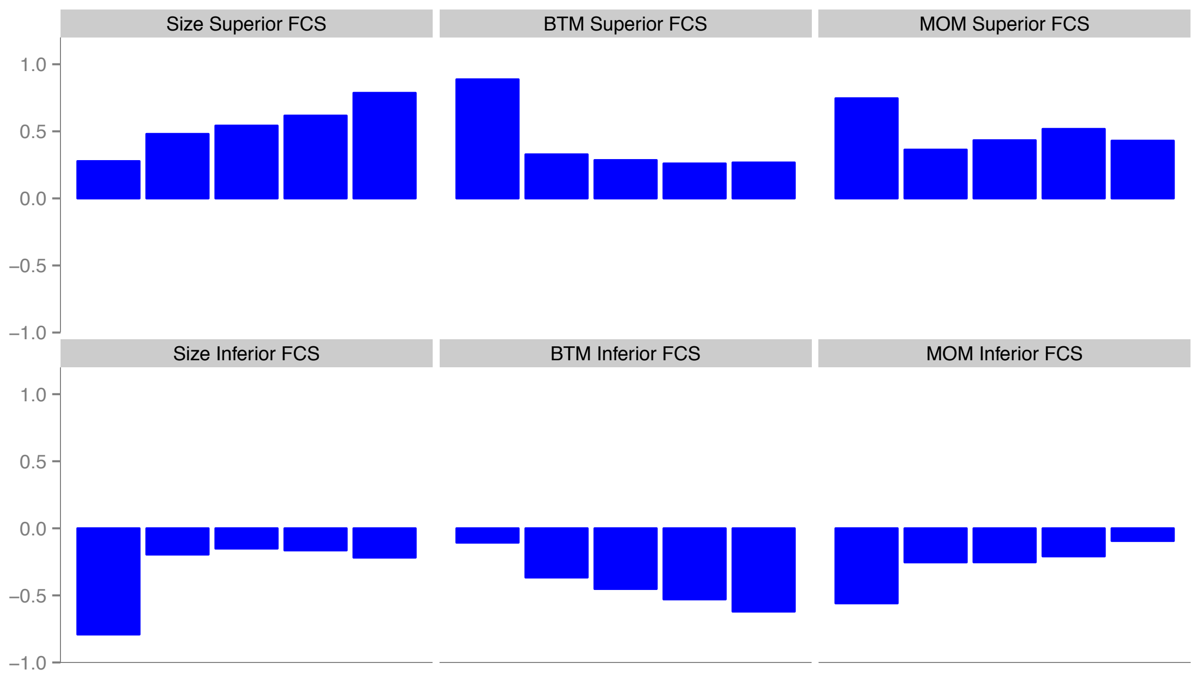

Figure 12: Performance of the superior and inferior FCS portfolios for different realizations of the risk factors. The top row shows histograms of the four-factor alphas for the FCS portfolio of superior funds for different values of the size factor (left panel), book-to-market factor (middle panel)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

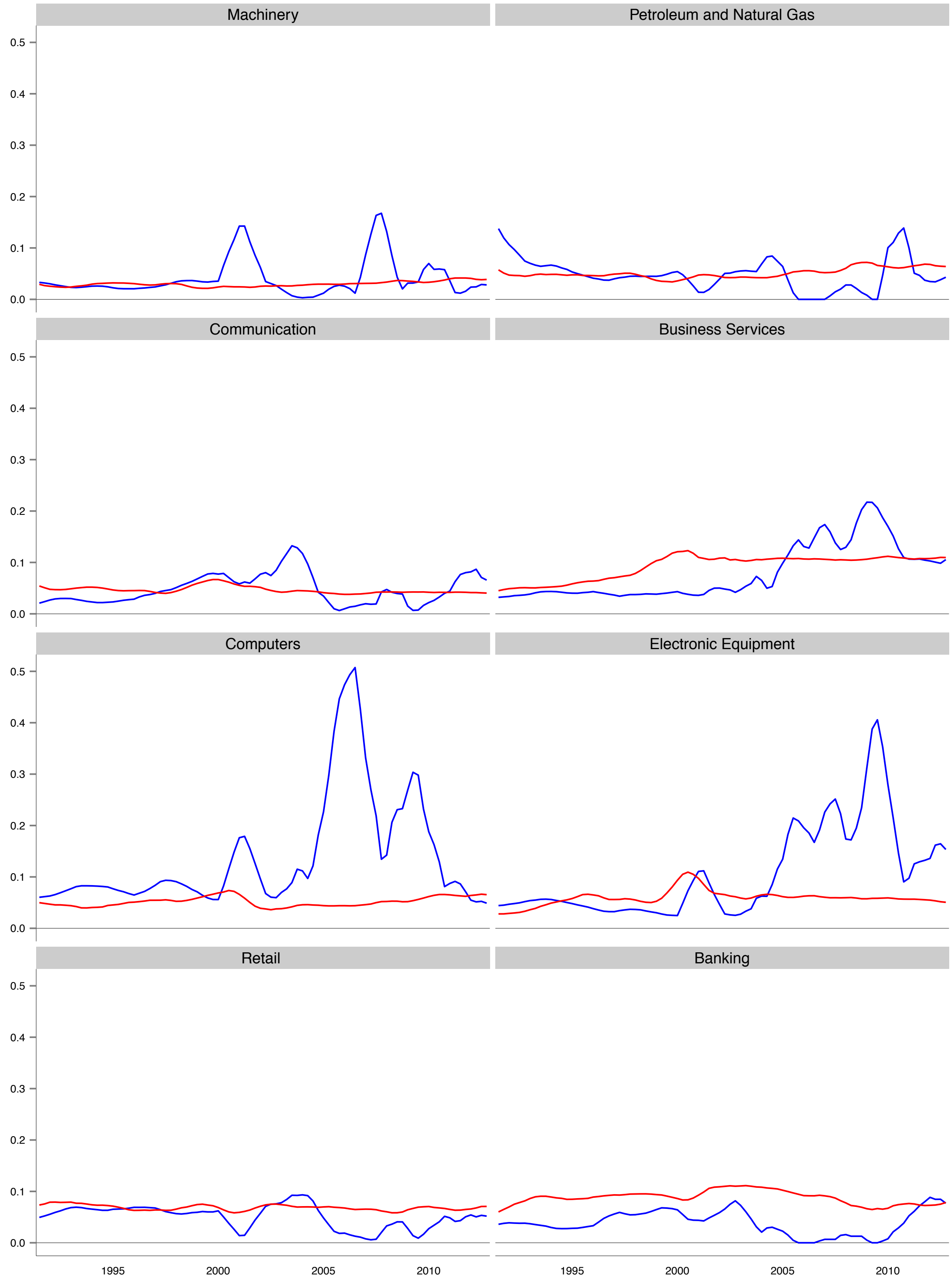

Figure 13: Industry concentrations for the top FCS portfolio. The figure presents one-year rolling averages of industry concentrations for the FCS portfolio composed of funds identified to have superior performance (blue line). We have chosen the 8 industries with the highest maximum absolute values in industry concentration. For comparison the red lines show the average concentration in the same industries computed across all funds in our sample.The plots assume $\lambda = 0.90$ and are based on the time-varying alpha model that combines data on fund returns and holdings to estimate alphas.

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