Can Mutual Fund "Stars" Really Pick Stocks? New Evidence from a Bootstrap Analysis

Robert Kosowski
Financial Markets Group
London School of Economics and Political Science
Houghton Street
London WC2A 2AE
United Kingdom
Phone: 0207-852-3574
R.L.Kosowski@lse.ac.uk

Allan Timmermann
Hal White
Department of Economics
University of California, San Diego
9500 Gilman Drive
La Jolla, CA 92093-0508
Phone: (858) 534-4860
atimmerm@ucsd.edu

Russ Wermers
Department of Finance
Robert H. Smith School of Business
University of Maryland at College Park
College Park, MD 20742-1815
Phone: (301) 405-0572
rwermers@rhsmith.umd.edu

September 2001

We thank participants at the 2001 CEPR/JFI "Institutional Investors and Financial Markets: New Frontiers" symposium at INSEAD (especially Peter Bossaerts, the discussant), the 2001 Western Finance Association annual meetings in Tucson, Arizona (especially David Hsieh, the discussant), the Empirical Finance Conference at the London School of Economics (especially Greg Connor, the discussant), and workshop participants at the University of California, San Diego, Humboldt-Universitaet zu Berlin, Imperial College (London), and the University of Pennsylvania. This paper will be presented at the 2002 American Finance Association annual meetings in Atlanta, Georgia.

Can Mutual Fund "Stars" Really Pick Stocks?
New Evidence from a Bootstrap Analysis

Abstract

We apply an innovative bootstrap statistical technique to examine the performance of the U.S. equity mutual fund industry during the 1962 to 1994 period. Using this new method, we bootstrap the distribution of the performance measure (the "alpha") across mutual funds to determine whether funds with the best alphas are simply lucky, or whether managers of these funds possess genuine stockpicking skills–this bootstrap technique is necessary because of the complicated form of the distribution of alphas across funds and the non-normal nature of individual funds' alphas. Our results indicate that, controlling for luck, fund managers that pick stocks well enough to more than cover their costs do exist. That is, the distribution of alphas computed from bootstrapped fund returns (and assuming that no stockpicking talent exists) has a much smaller right tail than the distribution of alphas computed from actual fund returns. Unfortunately for investors, our bootstrap results also show strong evidence of funds with significant inferior performance. Further, our evidence suggests that stockpicking skills are most clearly evident among growth-fund managers. In general, our study supports the value of active mutual fund management, although it also highlights the drawbacks of funds actively managed by those who cannot pick stocks.

Introduction

Was Peter Lynch, former manager of the Fidelity Magellan fund, a "star" stockpicker, or was he simply endowed with stellar luck? The popular press seems to assume that his stock picks beat their benchmarks due to his unusual acumen in identifying underpriced stocks. In addition, Marcus (1990) asserts that the prolonged superior performance of the Fidelity Magellan fund is difficult to explain as a purely random outcome, where one assumes that the Magellan managers have no true stockpicking skills.

Some recent studies are supportive of the existence of *subgroups* of fund managers with superior stockpicking talents, although other studies conclude that the *average* fund finds it difficult to beat its benchmarks.[1] Chen, Jegadeesh, and Wermers (2000) examine the stockholdings and the active trades of mutual funds, and find that growth-oriented funds have unique skills in identifying underpriced large-capitalization growth stocks. Wermers (2000) finds that high-turnover mutual funds pick stocks that substantially beat the Standard and Poor's 500 index over the 1975 to 1994 period. Studies of UK mutual funds mirror the U.S. results: the average UK fund underperforms market indexes, while some fund managers outperform.[2]

The apparent superior performance of a small group of "star" fund managers, such as Mr. Lynch of Magellan, raises the question of whether this is credible evidence of genuine stockpicking skills, or whether it simply reflects the luck of individual fund managers. Hundreds of new funds are launched every year; by September 2000, there were over 4,200 U.S. equity mutual funds holding assets valued at almost $4.4 trillion. Among this huge universe of funds, it is natural to expect that some funds will outperform market indexes by a large amount, simply by chance. However, past studies of mutual fund performance do not explicitly recognize and model the role of luck in performance outcomes.[3] The literature on performance persistence, to some extent, is motivated by the need to measure performance during a period separate from the ranking period in an attempt to control for luck. However, persistence tests are susceptible to model misspecification that is

[1] For evidence of the underperformance of the average mutual fund, on a style-adjusted basis, see, for example, Carhart (1997).

[2] For example, Leger (1997) reports underperformance of the average fund and weak performance persistence in a study of 71 UK investment trusts. Using a sample of 2,375 unit trusts in existence during 1972-1995, Blake and Timmermann (1998) find underperformance (on a risk-adjusted basis) by the average UK fund, persistence in performance, and existence of a survivorship bias of 0.8 percent per year.

[3] Many papers on mutual fund performance examine the difference in performance between the best funds and an average fund, or between the best and worst groups of funds. Other papers adjust for return premia accruing to the characteristics of stockholdings. Although these methods attempt to correct for variation in common returns, they generally do not correct for idiosyncratic variation among funds.

present both in the ranking and the measurement periods. For example, a mutual fund manager holding a portfolio of value stocks would consistently beat a benchmark that does not control for the value premium.[4] Also, finding short-term persistence in subsets of funds (e.g., the portfolio of last-year's best growth funds) does not reveal whether individual funds can outperform the market over longer investment horizons, since the identity of the top performers changes over time.

This paper provides the first comprehensive examination of mutual fund performance ("alpha") that explicitly controls for luck without potential bias from misspecification. This is not because we claim our models are correctly specified–they may not be. Rather, our approach, which uses a bootstrap statistical approach, is robust to possible misspecification.[5] In bootstrapping performance estimates, we explicitly model and control for the idiosyncratic variation in mutual fund returns. However, our approach is not limited to computing whether the "best" or "worst" fund's performance was abnormal after controlling for idiosyncratic risk. As we model the full, cross-sectional distribution of performance estimates, we can also test how widespread genuine abnormal performance is by considering the statistical significance of various percentiles of this distribution.

To illustrate the central idea of our tests, suppose that we are told that a particular fund has an alpha of six percent per year over a five-year period. This would be, prima facie, an extremely impressive performance record. However, if we know that this fund is, in fact, the best performer among a group of 1,000 funds, then the fund's alpha would not appear to be nearly as impressive. Effectively, when the top and bottom funds are selected based on an *ex-post* ranking applied to the full set of funds in existence, it is far more difficult to assess the significance of alpha outliers. In particular, any such analysis needs to account for the non-normality of alphas, which, in turn, can be induced through several sources–non-normal benchmark or individual security returns, dynamic factor loading strategies, and time-series and cross-sectional (across funds) correlation in the idiosyncratic return component may all result in non-normal distributions of estimated alphas. In addition, managers with timing abilities may exhibit returns having co-skewness with the market portfolio.[6]

[4]For example, Grinblatt and Titman (1993) find evidence of persistence in mutual fund returns, before expenses and trading costs are deducted. Carhart (1997) and Wermers (1997) find that momentum in stock returns (and momentum investing by funds) explains the majority of this persistence, thus eliminating the momentum misspecification. However, Edelen (1999) show that another problem arises from the persistence of cash inflows from consumers to mutual funds. These persistent flows tend to consistently reduce the returns of some popular funds, and might be considered another source of model misspecification.

[5]Our approach is based on the bootstrap introduced by Efron (1979). For a detailed discussion of the properties of the bootstrap, see, for example, Efron and Tibishari (1993) or Hall (1992).

[6]Also, consumer cashflows induce a "negative market timing" effect on funds, as discussed by Edelen (1999).

Our objective in this paper is simple. In the face of mutual fund alphas that deviate, in the cross-section, significantly from being distributed normally, we address the following question. By random chance, how many fund managers will appear to be investing "stars" among a large sample of funds, simply due to luck, and how does this compare to the number we actually observe? To address this issue, we apply our bootstrap technique to the monthly net returns of the universe of U.S. equity funds during the 1962 to 1994 period, giving us results for one of the largest cross-sectional and time-series samples ever analyzed. We apply our bootstrap analysis to three classes of performance measurement models: (1) unconditional models of performance (e.g., Jensen (1968) and Carhart (1997)), (2) modified models of performance that include conditioning variables to account for changing factor loadings and factor return premia (e.g., Ferson and Schadt (1996)), and (3) modified models that include conditioning variables to account both for the dynamic nature of factor loadings and return premia as well as for the dynamic nature of alphas (Christopherson, Ferson, and Glassman (1998)).

Across all three classes of measures, our bootstrap tests indicate that, controlling for sampling variability (luck), superior funds that beat their benchmarks by an economically and statistically significant amount do exist. Notably, our tests also show strong evidence of inferior funds, controlling for sampling variability. We do not find it surprising that large numbers of "inferior" managers exist in our sample, since performance measurement is a difficult task requiring, for precision, a long fund lifespan. This evidence of inferior fund management is consistent with consumers who have difficulty in identifying the few (if any) fund managers that can "beat the market" (especially in terms of judging the skills of managers of relatively new funds). Much more surprising is our finding that a relatively large number of managers having superior stockpicking talents exist, with skills that are more than sufficient to cover costs (even after adjusting for the return premia to style investing). This evidence resuscitates active equity fund management as a worthwhile endeavor, at least for those consumers who somehow found a skilled manager.

Specifically, across our entire menu of performance evaluation models, we find that the extreme alphas of the top (and bottom) ten percent of funds are very unlikely to be a result of sampling variability (luck). This evidence can be illustrated by returning to our prior question of how many funds we would expect, by pure chance, to achieve a certain alpha. Of the approximately 1,000 mutual funds in our sample at the end of 1994, our bootstrap results indicate that, by luck alone, nine should achieve an alpha of at least six percent per year over (at least) a five-year period.

In reality, 60 funds exceeded this alpha. As our analysis shows, this is sufficient, statistically, to provide overwhelming evidence that some fund managers have superior talents in picking stocks. Overall, our study provides compelling evidence that, adjusting for all expenses and costs (except for load fees and taxes), the superior alphas of mutual fund stars survive. The key to our study is the bootstrap analysis, which allows us to analyze the complicated, non-normal cross-sectional distribution of mutual fund alphas in search of stars.

Further investigation indicates that most mutual fund stars are managers of funds having a growth-oriented investing strategy. This result is consistent with prior evidence at the pre-expense level that indicates that managers of growth-oriented funds can pick stocks that beat their benchmarks (at least before trading costs are considered), but value-oriented funds cannot (see, for example, Chen, Jegadeesh, and Wermers (2000)).

Finally, our bootstrap results may provide guidance for consumers wishing to use performance records to identify superior funds. For example, our bootstrap indicates that, among the subgroup of fund managers having an alpha exceeding two percent per year over a five-year (or longer) period, we expect that half had stockpicking talents during the historical period, while the other half were simply lucky. However, a caveat to this message is that it remains to future research to determine whether a fund manager with superior past talents is any more likely than a randomly chosen manager to have useful stockpicking talents in the future.

Our paper proceeds as follows. Section I describes the mutual fund database used in our study, while Section II presents the performance measures used in our bootstrapping procedure. The details of the bootstrapping procedure are described in Section III. Section IV provides empirical results, the robustness of which are further explored in Section V. Section VI buttresses our methods by presenting a Monte Carlo study of the statistical properties of our bootstrap approach under different scenarios. We conclude the paper in Section VII.

I Data

We examine monthly net returns data from the Center for Research in Security Prices (CRSP) mutual fund files, which is first used in Carhart (1997). The CRSP database contains monthly data on net returns for all mutual funds existing at any time between January 1, 1962 and December 31, 1999, with no minimum survival requirement for funds to be included in the database. Further details on this mutual fund database are available from CRSP.

Although investment objective information is available from the CRSP database, we use investment objective information from a different source, the CDA-Spectrum mutual fund files from Thomson Financial, Inc., of Rockville, Maryland. We use CDA investment objectives because CRSP investment objective data is unreliable before 1992.[7] This CDA database, and the technique for matching it with the CRSP database, are described in Wermers (1999, 2000). Since both the CRSP and CDA databases contain essentially all mutual funds existing during our sample period (with the exception of some very small funds), our merged database is essentially free of survival bias.[8]

Since CDA investment objectives are available from 1975 to 1994, our core CRSP mutual fund net returns database covers the 1975 to 1994 period (inclusive). We supplement this sample with the 1962 to 1974 (inclusive) net returns of all funds existing on January 1, 1975, which is the first date that CDA investment objectives become available. These pre-1975 returns are added to include as many funds as possible in our regression-based tests, which require a minimum return history for a fund to be included in the tests.[9]

In our study, we focus on funds that predominantly hold diversified portfolios of U.S. equities.[10] Our final database contains monthly net returns data on 1,734 diversified U.S. equity funds that exist for at least a portion of the period from January 31, 1962 to December 31, 1994. We study the performance of the full sample of funds, as well as funds in each investment objective category. These investment objective categories include aggressive growth funds (262), growth funds (945), growth and income funds (345), and balanced or income funds (182).[11]

We require a fund to have at least 60 valid monthly net return observations to be included in

---

[7]Investment objective information on funds in the CRSP database is often missing for at least some years (and sometimes all years) of the fund's existence before 1992. In addition, CRSP reports investment objective information, when available, from four different sources. As these sources classify funds in different ways, it is often difficult to determine the precise investment objective of a fund. The CDA-Spectrum files report investment objectives in a more consistent manner across funds and over time.

[8]A small number of very small funds could not be matched between the CRSP and CDA files–that is, they were usually present in the CRSP database, but not in the CDA database. Wermers (2000) discusses this limitation of the matching procedure; however, we note that these funds are generally very small funds with a short life during our sample period. Since we require a minimum return history for a fund to be included in our regression tests, the majority of these unmatched funds would be excluded from our tests in any case.

[9]Although backfilling fund data before 1975 has the potential to induce a slight survival bias in our tests during that period, we ran all of our major tests, in an earlier version of this paper, without including this pre-1975 data. All major results are similar to the results shown in this version of our paper.

[10]We include a mutual fund in our sample only if the fund held at least 50 percent of its assets in U.S. equities during the majority of its existence.

[11]Income funds and balanced funds are combined in our study, as the number in each category is relatively small (and because funds in these two categories make similar investments). Descriptions of the types of investments made by funds in all categories are available in Grinblatt, Titman, and Wermers (1995).

our baseline bootstrap tests.[12] This minimum data requirement, while potentially introducing a survival bias, is necessary to allow more precise regression parameter estimates for our more complex models of performance. However, as a robustness check, we apply the bootstrap (using some of our simpler models) with a minimum history requirement of 18, 30, and 90 months, respectively. The results from these robustness tests, which will be presented in detail in a later section, generally show that survival bias has a minimal impact on our findings.

Some caveats are in order for our study. As with most studies of mutual fund performance, our study uses the mutual fund as the unit of measurement, and not the mutual fund manager. Precise data on mutual fund managers over this time period are not currently available. It is likely that the manager of a fund is as much responsible for the success (or failure) of the fund as other influences on the portfolio choices of the fund (such as the research team). However, if stockpicking ability resides at the manager level, and not the fund level, we expect that our use of the fund as the unit of observation would only add noise (and not bias) to our tests of performance, making it more difficult to reject the null of no genuine performance ability, even if managers do have such abilities.

It is also true that the same manager may manage more than one fund at the same time, or that funds may "herd" in their investment choices by mimicking the portfolio holdings of competitor funds. Both of these concerns may introduce correlation between the performance of different funds–however, we address this issue by adopting, in an extension of our tests, a bootstrap that accounts for dependencies in idiosyncratic risk across funds.

II Performance Measures

Several different measures of the performance (alpha) of managed equity portfolios have been promoted in the recent academic literature. In order to demonstrate the robustness of our bootstrap technique, we will apply it to several of these models, and relegate the interpretation of which model is most appropriate to the reader. In this section, we briefly describe each of the performance measures used in our paper. These models can be grouped into three classes: unconditional models, conditional-β models, and conditional-α-and-β models.

[12] These monthly returns need not be contiguous, but any gap in returns results in the next non-missing return observation being discarded, since this return is cumulated (by CRSP) since the last non-missing return observation (and cannot be used in our regressions).

A. Unconditional Measures of Alpha

Our first class of models employs unconditional performance measures. The first three measures of performance in this class are the Jensen (1968), the Fama and French (1993), and the Carhart (1997) alphas. The Carhart (1997) four-factor regression model is

$$r_{i,t} = \alpha_i + b_i \cdot RMRF_t + s_i \cdot SMB_t + h_i \cdot HML_t + p_i \cdot PR1YR_t + \varepsilon_{i,t} \ , \qquad (1)$$

where $r_{i,t}$ is the month t excess return on the managed portfolio (net return minus the T-bill return), $RMRF_t$ is the month t excess return on a value-weighted aggregate market proxy portfolio; and SMB_t, HML_t, and $PR1YR_t$ equal the month t returns on value-weighted, zero-investment factor-mimicking portfolios for size, book-to-market equity, and one-year momentum in stock returns, respectively. The Fama and French alpha is computed using the Carhart model of Equation (1), excluding the momentum factor ($PR1YR_t$), while the Jensen alpha is computed using the market excess return as the only benchmark:

$$r_{i,t} = \alpha_i + b_i \cdot RMRF_t + \varepsilon_{i,t} \qquad (2)$$

Since market-timing abilities can bias inferences of selectivity abilities (see, for example, Grinblatt and Titman (1989)), we add two more performance models that control for the presence of timing ability. These are the Treynor and Mazuy (1966) model,

$$r_{i,t} = \alpha_i + b_i \cdot RMRF_t + \gamma_i \cdot [RMRF_t]^2 + \varepsilon_{i,t} \ , \qquad (3)$$

and the Merton and Henriksson (1981) model,

$$r_{i,t} = \alpha_i + b_i \cdot RMRF_t + \gamma_i \cdot [RMRF_t]^+ + \varepsilon_{i,t}, \qquad (4)$$

where α_i is the measure of selectivity for the managed portfolio, controlling for any market-timing abilities (under certain assumptions), and $[RMRF_t]^+ = \max(0, RMRF_t)$. We include these regressions in order to demonstrate the robustness of our bootstrap tests to models that explicitly allow for the presence of timing abilities.

B. Conditional Measures of Alpha

Our second class of models, to control for time-varying factor loadings (of mutual fund portfolios) and factor returns, uses the technique of Ferson and Schadt (1996) to modify the Jensen model of Equation (2) as follows:

$$r_{i,t} = \alpha_i + b_i \cdot RMRF_t + \sum_{j=1}^{K} B_{i,j}[z_{j,t-1} \cdot RMRF_t] + \varepsilon_{i,t} \ , \tag{5}$$

where $z_{j,t-1} = Z_{j,t-1} - E(Z_j)$, the time $t-1$ deviation of public information variable j from its unconditional mean, and $B_{i,j}$ is the fund's "beta response" to the value of $z_{j,t-1}$. Therefore, the Ferson and Schadt measure computes the alpha of a managed portfolio, controlling for any investment strategies that use economic information that is publicly available to dynamically modify the portfolio's beta due to the predictability of factor returns.

Ferson and Schadt find that three public information variables are useful instruments for economic information that can help to predict market returns: (1) the lagged level of the one-month Treasury bill yield (TBILL), (2) the lagged dividend yield of the CRSP value-weighted New York Stock Exchange (NYSE) and American Stock Exchange (AMEX) stock index (DY), and (3) a lagged measure of the slope of the term structure (TERM).[13] We construct TBILL using the 30-day annualized yield on Treasury bills; DY as the price level at the end of the previous month on the CRSP value weighted NYSE/AMEX index, divided into the previous 12 months of dividend payments for the index; and TERM as the constant-maturity 10-year Treasury bond yield less the three-month Treasury bill yield. Further details on the construction of these instruments are available in Ferson and Schadt (1996).

In an analogous fashion, we modify Equation (1) to add the Ferson and Schadt predictive variables; specifically, the regressors include each of the three lagged information variables multiplied by each of the four Carhart regressors, plus the unconditional Carhart regressors (for a total of 16 regressors plus the intercept). Similar procedures are used to create a conditional version of the Fama-French model, as well as the Treynor-Mazuy and Merton-Henriksson models of Equations (3 and 4). Details on the construction of the conditional versions of these models are available in

[13]These variables have been shown to be useful for predicting security returns and risks over time by other studies as well (see, for example, Pesaran and Timmermann (1995)). Two other instruments, a lagged quality spread in the corporate bond market (QUAL), and a dummy variable for the month of January (JAN), were found to be insignificant in predicting returns by Ferson and Schadt.

Ferson and Schadt (1996).

In most cases, therefore, $K = 3$ in Equation (5). However, for some of our more complex models, we perform tests with only a single conditioning variable, the lagged level of the one-month Treasury bill yield (TBILL), to increase the degrees-of-freedom in our tests (in this case, $K = 1$).

Our third class of models incorporates the Christopherson, Ferson and Glassman (1998) conditional framework that allows both the alpha and the factor loadings of a fund to vary through time. For example, the Jensen model of Equation (2) is modified as follows:

$$r_{i,t} = \alpha_i + \sum_{j=1}^{K} A_{i,j} \cdot z_{j,t-1} + b_i \cdot RMRF_t + \sum_{j=1}^{K} B_{i,j}[z_{j,t-1} \cdot RMRF_t] + \varepsilon_{i,t} \ ,$$

This model computes the alpha of a managed portfolio, controlling for any investment strategies that use economic information that is publicly available to either change the portfolio's beta, or to add stocks with abnormally high expected returns, conditional on the information. Analogous conditional alpha and beta versions of the Carhart, Fama-French, Treynor-Mazuy, and Merton-Henriksson models are also employed in our study. Again, in most instances, we use three information variables ($K = 3$), although we also employ versions of the models with only one information variable ($K = 1$).

III Bootstrap Evaluation of Fund Alphas

A. The Bootstrap Approach

The bootstrap is a nonparametric approach to statistical inference. In evaluating mutual fund performance, there are several advantages in using the bootstrap. Specifically, traditional parametric methods use *a priori* assumptions about the shape of the distribution from which individual fund alphas are drawn. For example, Alexander et al. (1998) report the statistical significance of estimates of Jensen's alpha ($\hat{\alpha}$) under the assumption that they are drawn from a normal distribution.

However, as we will demonstrate in a later section, the empirical distribution of residuals from Jensen (and other) regressions is highly non-normal for most mutual funds in our sample. Thus, the distribution of $\hat{\alpha}$ may be poorly approximated by normality. While the central limit theorem justifies regarding the normal distribution as a first-order approximation to the true distribution of $\hat{\alpha}$, the bootstrap can improve on this approximation (see, for example, Bickel and Freedman (1984) and Hall (1986)).

In addition, refinements of the bootstrap (which we will implement) provide a general approach for dealing with unknown time-series dependencies that are due, for example, to heteroskedasticity or serial correlation in the residuals from performance regressions. These bootstrap refinements also provide a way of handling unknown cross-sectional correlations (across funds) in regression residuals, thus avoiding the estimation of a very large covariance matrix for these residuals. See the Appendix for further details on the bootstrap approach.

The hypothesis that the manager of the very best fund cannot pick stocks well enough to cover costs (cannot produce a positive alpha) can be stated as follows:

$$\begin{aligned} H_0 &: \max_{i=1,\dots,L} \alpha_i \leq 0 \text{ , and} \\ H_A &: \max_{i=1,\dots,L} \alpha_i > 0. \end{aligned}$$

However, we are interested not only in analyzing the maximum alpha, but also in determining whether managers of other top funds can pick stocks. For example, do the top five funds–or, alternatively, the top five percent of funds–generate higher returns than would be expected by random chance? To address this question, we evaluate the alpha of several of the top-ranked funds.

To illustrate, suppose that a group of funds have been ranked by their alphas, and let i^* be the rank of a given fund. When testing whether this fund manager can pick stocks, the null and alternative hypotheses are

$$\begin{aligned} H_0 &: \alpha_{i^*} \leq 0 \text{ , and} \\ H_A &: \alpha_{i^*} > 0. \end{aligned}$$

We also test whether managers of the worst-ranked funds pick stocks well enough to cover their costs. Here, the hypothesis test takes the following form for a fund with rank i^*(ranked from the minimum mutual fund alpha in our dataset):

$$\begin{aligned} H_0 &: \alpha_{i^*} \geq 0 \text{ , and} \\ H_A &: \alpha_{i^*} < 0. \end{aligned}$$

We evaluate, separately, the distribution of the $\hat{\alpha}$ values and the distribution of the estimated t-statistic of $\hat{\alpha}$, $\hat{t}_{\hat{\alpha}}$.[14] Although $\hat{\alpha}$ measures the economic size of abnormal performance, it has a relatively high coverage error in construction of confidence intervals.[15] Alternatively, $\hat{t}_{\hat{\alpha}}$ is a pivotal statistic, and, thus, generates lower coverage errors.[16] Also, $\hat{t}_{\hat{\alpha}}$ has another attractive statistical property. Specifically, funds with a shorter history of monthly net returns will have an alpha estimated with less precision, and will tend to generate alphas that are outliers. The t-statistic provides a correction for these spurious outliers by normalizing the estimated alpha by the estimated precision of the alpha estimate–it is related to the well-known "information ratio" method of performance measurement of Treynor and Black (1973).

Using this performance measure, the null and alternative hypotheses are (for the highest ranked fund):

$$\begin{aligned} H_0 &: \max_{i=1,\ldots,L} t_i \leq 0, \text{ and} \\ H_A &: \max_{i=1,\ldots,L} t_i > 0. \end{aligned}$$

For the lowest-ranked fund, the null and alternative hypotheses are given by reversing the inequalities above.

B. Implementation

In this section, we illustrate our bootstrapping procedure for generating fund alphas with the Carhart (1997) four-factor model of Equation (1). The application of this bootstrap procedure to other models used in our paper is very similar, with the only modification of the following steps being the substitution of the appropriate model of performance.

To prepare for our bootstrap procedure, we use the Carhart model to compute the OLS-estimated alphas, factor loadings, and residuals using the time-series of monthly net returns for

[14] We estimate $\hat{t}_{\hat{\alpha}}$ using a heteroskedasticity and autocorrelation adjusted estimate of the standard error.

[15] The coverage probability is the probability that the confidence interval includes the true parameter, and the coverage error is the difference between the true and nominal coverage.

[16] A pivotal statistic is one that is not a function of nuisance parameters, such as $Var(\varepsilon_{it})$. For further details, see the Appendix.

fund i :

$$r_{it} = \hat{\alpha}_i + \hat{\beta}_i RMRF_t + \hat{s}_i SMB_t + \hat{h}_i HML_t + \hat{p}_i PR1YR_t + \hat{\varepsilon}_{i,t} \ ,$$

where r_{it} is the net return of fund i during month t, in excess of the T-bill rate; $RMRF_t$ is the month t return, net of T-bills, on a value-weighted aggregate market proxy portfolio; and SMB_t, HML_t, and $PR1YR_t$ are the month t returns on value-weighted, zero-investment factor-mimicking portfolios for the market capitalization of equity, the ratio of the book value to the market value of equity, and the one-year past return of stocks, respectively. For fund i, the coefficient estimates, $\{\hat{\alpha}_i$, $\hat{\beta}_i$, $\hat{s}_i$, $\hat{h}_i$, and $\hat{p}_i\}$, are saved, as well as the time-series of estimated residuals, $\{\hat{\varepsilon}_{i,t}\ ,\ t = 1, T_i\}$.

Next, for our bootstrap procedure, we use two different approaches. The first approach resamples the saved residuals only, while the second resamples both the factor returns and the residuals. We first describe residual-only resampling, followed by a discussion of residual and factor resampling.

B.1 Residual-Only Resampling

In a portfolio context, residual-only resampling is used to help control for non-normal security returns, which can also remain at the portfolio level.[17] In addition, dynamic factor loading strategies, as well as time-series and cross-sectional (across funds) correlation in the idiosyncratic return component may all result in non-normal distributions of estimated alphas. An example is given by Grinblatt, Titman, and Wermers (1995), who show that the majority of funds use a momentum strategy in picking stocks.

For residual (only) resampling, we draw a sample with replacement from the fund i residuals that are saved from the first step, creating a time-series of resampled residuals, $\{\hat{\varepsilon}^b_{i,t}\ ,\ t = s^b_1, s^b_2, ..., s^b_{T_i}\}$, where b=1 (for bootstrap resample number one), and, as indicated, where a sample is drawn having the same number of residuals (e.g., the same number of time periods, T_i) as the original sample. This resampling procedure is repeated for the remaining bootstrap iterations, $b = 2, ..., B$.

Next, for each bootstrap iteration, b, a time-series of (bootstrapped) monthly net returns is constructed for this fund, imposing the null hypothesis of zero true performance ($\alpha_i = 0$):

[17]Co-skewness of individual security returns may not diversify away in large portfolios, especially if an omitted factor such as an industry factor is responsible. Also, many mutual fund managers hold relatively undiversified portfolios of stocks, taking large bets on a subgroup of stocks in their portfolios, which can exacerbate this effect (e.g., the Janus family of funds). Finally, funds occasionally hold derivatives in their portfolios.

$$\{r_{i,t}^b = \hat{\beta}_i RMRF_t + \hat{s}_i SMB_t + \hat{h}_i HML_t + \hat{p}_i PR1YR_t + \hat{\varepsilon}_{i,t}^b\ ,\ \ t = s_1^b, s_2^b, ..., s_{T_i}^b\}, \tag{6}$$

where $s_1^b, s_2^b, ..., s_{T_i}^b$ is the time reordering imposed by resampling the residuals in bootstrap iteration b. As indicated by Equation (6), this sequence of artificial returns has a true alpha (intercept) of zero, since the residuals are drawn from a sample that is mean zero by construction. However, when we next regress the returns for a given bootstrap sample, b, on the Carhart factors, a positive estimated alpha may result, since that bootstrap may have drawn an abnormally high number of positive residuals, or, conversely, a negative alpha may result if an abnormally high number of negative residuals are drawn. Of course, the estimate of alpha will also depend on the correlation of the resampled residuals with the factor-mimicking returns that are matched with them.

After estimating the Carhart alpha (and t-statistic of the alpha) for all bootstrap iterations for fund i, we repeat this procedure for all other funds in our sample. The end result of this procedure is an empirical distribution for the alpha of each mutual fund, based solely on resampling the residuals of the original performance regression.

B.2 Residual and Factor Resampling

Residual and factor resampling, in a portfolio context, controls for any of the above non-normalities in residuals that may have a market component. For example, investing on momentum has an industry component (Moskowitz and Grinblatt (1998)), which may, in turn, have a market component. In addition, managers with timing abilities may exhibit returns having co-skewness with the market portfolio.[18]

For residual and factor resampling, we augment the residual resampling procedure with factor returns that are resampled independently of the residuals. When resampling these factor returns, the same draw is used across all funds, giving the following data for bootstrap iteration b for fund i:

$$\{RMRF_t^b, SMB_t^b, HML_t^b,\ \text{and}\ PR1YR_t^b,\ t = u_1^b, u_2^b, ..., u_{T_i}^b\}\ \text{and}\ \{\hat{\varepsilon}_{i,t}^b,\ t = s_1^b, s_2^b, ..., s_{T_i}^b\}$$

Resampling factor returns as well as residuals allows for sampling variation in the coefficient esti-

[18] Also, consumer cashflows induce a "negative market timing" effect on funds, as discussed by Edelen (1999).

mates, $\{\hat{\beta}_i,\ \hat{s}_i,\ \hat{h}_i$, and $\hat{p}_i\}$, that results from using a particular draw of factor realizations, as well as residuals, over the sample period of 1962 to 1994.

Next, for each bootstrap iteration, b, a time-series of (bootstrapped) monthly net returns is constructed for fund i, again imposing the null hypothesis of zero true performance ($\alpha_i = 0$):

$$
\begin{aligned}
\{r_{i,t}^b &= \hat{\beta}_i RMRF_{t_F} + \hat{s}_i SMB_{t_F} + \hat{h}_i HML_{t_F} + \hat{p}_i PR1YR_{t_F} + \hat{\epsilon}_{i,t_\epsilon}^b \ ,\ t_F = u_1^b, u_2^b, ..., u_{T_i}^b \\
\text{and } t_\epsilon &= s_1^b, s_2^b, ..., s_{T_i}^b\},
\end{aligned}
$$

where $u_1^b, u_2^b, ..., u_{T_i}^b$ and $s_1^b, s_2^b, ..., s_{T_i}^b$ are the (matched) time reorderings imposed by resampling the factor returns and residuals, respectively, in bootstrap iteration b.

Repeating these steps across funds, $i = 1, ..., N$, and bootstrap iterations, $b = 1, ..., B$, we then build the cross-sectional distribution of the alpha estimates, $\hat{\alpha}_i^b$, or their t-statistics, $\hat{t}_{\hat{\alpha}_i}^b$, resulting purely from sampling variation, as we impose the null of no abnormal performance. For example, in the case of bootstrapping the distribution of the maximum $\hat{\alpha}$ performance measure, this information can be represented as follows:

		Fund number				
		1	2		N	
Alpha estimates based on unmodified returns	0	$\hat{\alpha}_1$	$\hat{\alpha}_2$	...	$\hat{\alpha}_N$	$\Longrightarrow \max\{\hat{\alpha}\}$
	1	$\hat{\alpha}_1^1$	$\hat{\alpha}_2^1$	...	$\hat{\alpha}_N^1$	$\Longrightarrow \max\{\hat{\alpha}^1\}$
	2	$\hat{\alpha}_1^2$	$\hat{\alpha}_2^2$	...	$\hat{\alpha}_N^2$	$\Longrightarrow \max\{\hat{\alpha}^2\}$
Alpha estimates based on bootstrapped returns	3	$\hat{\alpha}_1^3$	$\hat{\alpha}_2^3$	...	$\hat{\alpha}_N^3$	$\Longrightarrow \max\{\hat{\alpha}^3\}$
(“B” bootstrapped sets of α estimates)	...					
	B	$\hat{\alpha}_1^B$	$\hat{\alpha}_2^B$	...	$\hat{\alpha}_N^B$	$\Longrightarrow \max\{\hat{\alpha}^B\}$

If we find that very few of the bootstrap iterations generated a maximum α-estimate as high as that observed in the raw (unmodified) data, this suggests that sampling variation (luck) is not the source of performance, but that genuine stockpicking skills actually exist. In all of our bootstrap tests described above, we run 1,000 bootstrap iterations ($B = 1,000$).

C. Accounting for Cross-Sectional Dependencies in Idiosyncratic Returns

The above procedures assume that the factors absorb all common variation in mutual fund returns, i.e., the model is well-specified. By construction they thus do not capture any possible cross-sectional (across funds) correlation in residuals. However, funds often hold the same stocks, in part because they tend to herd in their investments to some degree (see Wermers (1999) for evidence of herding behavior among mutual funds). To refine our bootstrap procedure to capture cross-sectional correlation in residuals, we implement a bootstrap method that draws residuals, across funds, for identical time periods. That is, rather than drawing sequences of time periods, t_i, that are unique to each fund, i, we draw T time periods from the set $\{t = 1, ..., T\}$, then resample residuals for this reindexed time sequence across **all** funds, thus preserving any cross-sectional correlation in the residuals.

In Section IV, we will show empirical findings that suggest that lack of cross-sectional dependence is not a serious concern.[19] However, we also study, in Section V, a cross-sectional bootstrap procedure that accounts for possible dependencies across individual fund residuals.

IV Empirical Results

A. Cross-Sectional (Nonbootstrapped) Distribution of Alphas

Our sample in this paper includes all funds that existed anytime during the 1975 to 1994 period (augmented with 1962 to 1974 returns data), and that had an investment objective consistent with holding predominantly U.S. equities. Before beginning our bootstrap tests of the significance of alpha outliers, we present statistics that describe the cross-sectional distribution of fund alphas for our sample of funds, for each of the performance models described in Section II. These distributions allow us to address, across all performance measurement models:

- the level of alpha for a "typical" fund in our sample (average or median),
- the cross-sectional dispersion in alphas (across funds),
- the normality of the estimated alphas, and

[19] A bootstrap procedure that wrongly assumes independence across idiosyncratic components of fund returns could in principle exaggerate the dispersion of the alpha estimates and lead to exaggerated $p-$values for the top and bottom funds.

- the model that appears to fit our mutual fund returns with the greatest precision (the least tracking error).

To address the first two issues, Panel A of Table I presents alphas for the average and median funds, as well as the cross-sectional standard deviation of alphas, for all five unconditional models of performance: the Jensen measure (model 1), the Treynor-Mazuy model (TM; model 2), the Henriksson-Merton model (HM; model 3), the Fama-French three-factor model (FF; model 4) and the four-factor Carhart model (model 5). Panel A also shows the cross-sectional average t-statistic associated with our fund alphas for each performance model. To be included in the results shown in any panel of this table, a mutual fund must have 60 or more non-missing monthly net returns available in the CRSP mutual fund database. These monthly returns need not be contiguous, but any gap in returns results in the next non-missing return observation being discarded, since this return is cumulated (by CRSP) since the last non-missing return observation (and cannot be used in our regressions).

Our results are surprisingly similar for all models of performance–the average and median alpha estimates are quite small, and the average estimated t-statistic for these alphas is insignificant under all models. For example, the Jensen model (model 1) exhibits an average alpha of only 0.3 basis points per month (3.6 basis points, annualized), a median alpha of 0.6 basis points per month (7.2 basis points, annualized), and an average t-statistic that is close to zero. Although the average and median alphas for the HM model, which includes a control for timing ability, are somewhat larger in magnitude, the average t-statistic remains insignificant.

Together, our average and median alphas indicate that there is little evidence that the typical mutual fund, having at least a five-year return history, significantly beats its benchmarks after expenses and trading costs. Our results are slightly different from some prior studies (e.g., Carhart (1997) or Wermers (2000)) that have shown negative (but, small in magnitude) and significant average Carhart alphas. The difference in results is due to the slightly different sample of fund returns that our study starts with, as well as the fact that we require a minimum of 60 monthly net returns for inclusion of a fund in our tests to facilitate the precise estimate of regression parameters.[20] While this 60-month return requirement imposes a slight survival bias in our test,

[20]Specifically, we analyze mutual fund returns for the period 1962 to 1994, while Wermers (2000) analyzes returns from 1975 to 1994 and Carhart analyzes returns from 1962 to 1993. In addition, these latter papers measure the alpha of value-weighted or equal-weighted portfolios of funds, which allows them to include funds having a very short life. Our study is focused on individual funds rather than portfolios of funds, which requires us to exclude funds

we will show in a later section of our paper that it does not significantly affect our inferences about the tails of the alpha distributions.

Interestingly, Panel A shows that the cross-sectional (across funds) standard deviations of these alphas are quite large for all five models, when compared to the average alphas. These standard deviation estimates range from 0.26 (for the Jensen model) to 0.44 (for the HM model) percent per month, which indicates that the tails of our alpha distributions contain substantial numbers of funds having fairly dramatic levels of performance (either positive or negative).

Panel B of Table I presents statistics for alpha distributions computed using conditional models of performance. These conditional models control for predictability in factor returns, as well as for time-varying factor loadings used by mutual funds to take advantage of the predictability. As discussed in Section II, we follow Ferson and Schadt (1996) in using three instruments to control for predictable variation in factor loadings and in factor returns: (1) the lagged level of the one-month yield on Treasury bills (TBILL), (2) the lagged dividend yield of the CRSP value-weighted New York Stock Exchange (NYSE) and American Stock Exchange (AMEX) index (DY), and (3) a lagged measure of the slope of the term structure (TERM).

These "conditional-β" factor models include conditional versions of all five models shown in Panel A. For the Fama-French and Carhart models, we present two conditional versions: one with only one conditioning variable (TBILL), and the second with all three conditioning variables. The first version is shown since it reduces the number of regressors included in the model–which increases the degrees-of-freedom, and, in turn, the precision of alpha estimates.

All of the conditional-β models of Panel B exhibit very small and insignificant average and median estimates of alphas. Although there are some differences in alphas between the unconditional models of Panel A and the conditional-β models of Panel B, the alpha estimates are remarkably similar.

Panel C presents regression results for conditional models of performance that allow for predictability in both factor returns and in returns unrelated to the factors. These "conditional-α-and-β" models include conditional versions of the Fama-French and Carhart models. Again, the alphas of these conditional models are similar to those of their unconditional counterparts, and indicate that the alpha of the average mutual fund is small in magnitude and statistically insignificant.

---

having a very short life. However, in unreported tests, we find that an equal-weighted portfolio of all funds in our sample (including short-lived funds) exhibited a Carhart alpha of about -0.6 percent per year during our sample period, which is statistically significant at the 10 percent confidence level.

Our results so far indicate that the average mutual fund manager, at best, picks stocks only well enough to cover trading costs and other expenses. However, we have made no statement about the ability of subgroups of fund managers to pick stocks, although the dispersion of our alpha estimates indicates that, perhaps, both superior and inferior stockpickers exist in our sample. Before we address this issue with our bootstrap procedure, we explore whether our alpha estimates appear to be normally distributed–almost all previous papers make this normality assumption when analyzing alphas.

B. The Normality of Alphas

To motivate the need for a bootstrap analysis, we analyze the distribution of alphas generated by the various models. To test for the normality of alpha estimates, each panel of Table I reports the Jarque-Bera test for normality (see the "Rejection of Normality" statistic in each panel). This test is based on the skewness and kurtosis of the distribution of the estimated regression residuals resulting from the application of each performance measurement model.

Our results show that normality is rejected for 54-57 percent of funds, depending on the unconditional model that is used. Moreover, we find, in unreported results, that the rejections tend to be very large for many of the funds. Similarly, Panels B and C indicate that the percentage of funds for which normality of the residuals is rejected remains near 50 percent when the conditional-β and conditional-α-and-β models are used.

This strong finding of non-normal alpha estimates challenges the validity of earlier research that relies on the normality assumption for inference tests. This challenge to the standard t- and F-tests of the significance of fund alphas strongly indicates the need to bootstrap the level of significance of fund alphas in order to precisely determine whether significant outperformers (or underperformers) really exist in our sample.

C. Bootstrap Analysis of the Significance of Alpha Outliers

In this section, we analyze the tails of the cross-sectional distribution of alphas for our sample of funds using our bootstrapping procedure. First, however, we illustrate the range of alphas by comparing alphas for mutual fund "winners" and "losers."

C.1 The Spread in Alphas Between "Winner" and "Loser" Funds

Our results in Section A indicate that a wide dispersion in alphas exists for our sample of funds. However, exactly how big is the alpha spread between the best and worst performers? Panel A of Figure I shows this spread, for all models of performance. Each column also shows the performance of the fund with the marginal 1% and 5% alpha in the right (left) as a percentage of the performance of the top (bottom) fund.

Across most models, the spread is similar in magnitude–the exceptions are the unconditional and conditional-β versions of the HM model (models 3 and 8) and the conditional-α-and-β version of the Jensen and Carhart models with three conditioning variables (models 13 and 17). The larger HM model spread indicates that controlling for timing-related biases may be important for funds with extreme alpha values, while the larger conditional Jensen and Carhart spread indicates that this model is overfitted (since it has 19 regressors plus an intercept). This is also confirmed by poor model selection criterion of Models 13 and 17. For most models, however, the best-worst spread in alphas is about three percent per month (36 percent, annualized).

Panel B of Figure I shows the best-worst spread in our second measure, the t-statistic, across different performance models. Each column also shows the marginal 1% and 5% $t-$statistic in the right (left) tail of the distribution as a percentage of the highest (lowest) $t-$statistic. Again, most models exhibit similar best-worst spreads in t-statistics of alphas.

C.2 Choice of "Representative Models" of Performance

Since our bootstrap tests that will follow involve very extensive computations, we first choose a representative model from each class of performance measures. Each of these three representative models will then be used to conduct a complete bootstrap analysis of our sample of funds to address whether significant positive alphas really exist. In unreported tests, we used several other models to generate our baseline bootstrap tests, and found that results were similar to those of our representative models of performance.

To select a representative model from a given class of models, we determine which model provides the best "fit" for the average mutual fund's returns using a Schwartz Information Criterion (SIC) test. The SIC test statistic trades off the fit of a model against its complexity, and is a commonly used criteria for model selection. We compute the SIC statistic for each mutual fund in our sample (for each model), then present the average SIC statistic in Table I for each model.

Panel A indicates that, among unconditional models, the Fama-French three-factor model (model 4) provides the best fit for the average mutual fund–this model exhibits an average SIC value of 0.459. In fact, this model exhibits the best fit among all three classes of models, so our most extensive bootstrap tests will be conducted with the unconditional FF model.

Among conditional-β models, the FF model with one conditioning variable (model 9) generates the lowest SIC value, 0.477. Finally, among the conditional α-and-β models, the FF model with one conditioning variable (model 14) generated the lowest average SIC value, 0.490.[21] Thus, in our bootstrap tests, we use these three versions of the FF model as our representative models.[22]

C.3 Baseline Bootstrap Tests

Table II presents results for our baseline bootstrap tests. In each panel of the table, we show several points in the tails of the alpha distribution, using the unconditional (Panel A), conditional-β (Panel B), and conditional-α-and-β (Panel C) versions of the FF model. Specifically, each panel of Table II presents the estimated alpha at several points in the left and right tails of the alpha distribution. These sample points include the top and bottom five funds, as well as funds falling at several different percentile points in the left and right tails of the alpha distribution.

For example, the fifth-ranked fund in our sample (which can be a different fund under each of the three models), exhibits FF alphas of 0.95, 0.97, and 1.35 percent per month, respectively, for the unconditional, conditional-β, and conditional-α-and-β versions of the model (see the "5.max" alpha of each panel). Although, at first blush, these impressive alphas indicate that the fifth-best fund manager, under each model specification, has solid stockpicking talents, these results could be purely due to luck. As discussed in Section III.A, we compute a second measure of fund performance–the estimated t-statistic for the estimated alpha. As discussed in that section, the t-statistic has some advantageous statistical properties when constructing bootstrapped distributions. In addition, the alpha t-statistic may be a better indicator of performance than alpha, since it scales alpha by its standard error (which tends to be larger for shorter-lived funds).[23] Thus, the distribution of bootstrapped t-statistics for extreme values of the (unmodified return) t-statistic is likely to have

[21] In unreported tests, we examined the median instead of the mean value of the SIC statistic. The use of the median resulted in the same choice of the "best" models.

[22] The choice of the optimal number of conditioning variables is based on the model that maximized the Schwartz Information Criterion. We confirm Ferson and Schadt (1996)'s finding that the instruments QUAL and JAN do not add explanatory power.

[23] Indeed, the t-statistic of the alpha is closely related to the information ratio of Treynor and Black (1973).

better properties (fewer problems with high variance) than the distribution of bootstrapped alpha estimates in that region.[24]

To determine the role of luck in these outcomes, we present the bootstrapped p-values for the t-statistic of the alpha–as discussed in Section III.B, these p-values are bootstrapped with the constraint that the true alpha of the best fund is zero.[25] These p-values, all of which are less than 0.009, strongly indicate that the best fund managers (under each model) has stockpicking skills sufficient to provide fund shareholders with reliably positive alphas.

Looking across the entire right tail of the alpha distribution (the right side of each panel of Table I), we find that our evidence strongly indicates (at the 99 percent confidence level) that the fund managers in this tail of the alpha distribution had true stockpicking talents during our sample period. The results for the left tail of the alpha distribution (the left side of each panel of Table I) closely mirror the right-tail results–with the exception of the bottom four funds, the bootstrapped p-values indicate that fund managers in this tail of the alpha distribution (with 95 percent confidence under each performance model) cannot pick stocks well enough to cover their costs. To summarize our results of this section, only the performance of the bottom few funds can be attributed to (bad) luck. Further away from the extreme tails, regardless of the model, funds exhibit levels of performance that cannot be explained by random chance alone.[26]

Figure II provides a comparison of the distribution of fund alphas generated by unmodified fund returns with the distribution generated by the bootstrap. This figure, which also uses residual resampling with the unconditional FF model, shows the density of both distributions over their entire range.[27] The reader is reminded that, since the bootstrap density is generated under the constraint that the true alpha is zero, all dispersion in estimated alphas is due to random resampling variation.

The two densities are quite different–the distribution of alphas based on unmodified returns has

[24]Also, our reported t-statistics use Newey-West (1987) heteroskedasticity and autocorrelation consistent standard errors.

[25]To clarify the interpretation of the p-value that is bootstrapped, the p-value of 0.001 indicates that, by pure chance, the probability of the fifth-best fund (from an ex-post alpha sort) generating an alpha of 0.97 percent per month or higher (under the conditional-β model) is only 0.001.

[26]In unreported tests, we repeat the baseline bootstrap analysis of Table II for the unconditional four-factor Carhart model (Model 5) and the unconditional HM model (Model 3). The results lead to the same conclusions as the analysis based on the three versions of the FF model used in Table II.

[27]The kernel density estimator replaces the "boxes" in a histogram by "bumps" that are smooth and the kernel function is a weighting function that determines the shape of the bumps. This plot was generated using a Gaussian kernel function. The optimal bandwidth controls the smoothness of the density estimate and is calculated according to equation 3.31 of Silverman (1986).

more probability mass in the left and, especially, in the right tail, and far less mass in the center of the distribution than the bootstrapped alpha distribution. Again, this illustrates that our sample of funds exhibit alphas with significantly fatter tails than can be explained by random sampling error, which illustrates that many superior, and inferior funds exist.

One should note, at this point, that our sample of funds has a large range of lifespans. It is clear that a fund with a short lifespan, thus, having relatively few monthly returns, will exhibit an alpha estimate (from any performance model) that is less precise than a fund with a longer history of returns (although we require funds to have at least 60 monthly return observations to reduce the influence of short-lived funds). Funds with shorter lifespans, thus, introduce noise in the bootstrap distribution of alphas, as well, and will tend to disproportionately populate the extreme tails of the non-bootstrapped and bootstrapped alpha distributions.

From a different perspective, we can use the bootstrapped distribution of the estimated alpha to calculate how many funds (out of the total set of funds with a track record of at least five years) we would expect, by chance alone, to achieve a given level of performance. This number can then be compared to the number of funds that actually achieved this level of performance in our sample. Panel A of Figure III plots the number of funds from the original and from the bootstrapped distribution that surpass each given level of performance, while Panel B plots the numbers that perform below each given level.

For example, Panel A indicates that 26 funds exhibited an alpha estimate below -0.5 percent per month, compared to an expected number of only eight funds by random chance. Further, Panel B indicates that we would expect only nine funds to achieve an estimated alpha of more than 0.5 percent per month. In reality 60 funds achieved this alpha.

Overall, the results in this section provide strong evidence that a subgroup of our sample of fund managers possess superior stockpicking skills. We also confirm the existence of fund managers who cannot pick stocks–or, more precisely, who cannot pick stocks well enough to cover their costs. Another interpretation of our bootstrap results may provide guidance for consumers wishing to use performance records to identify superior funds. For example, our bootstrap indicates that, among the subgroup of fund managers having an alpha exceeding two percent per year over a five-year (or longer) period, we expect that half have stockpicking talents, while the other half are simply lucky. In addition, among the subgroup of fund managers having an alpha below minus two percent per year over a five-year (or longer) period, we expect that half are unable to pick stocks well enough to cover costs, while the other half are simply unlucky. However, a caveat to this message is that a manager with past talents does not necessary have future abilities.

Of course, it is not very surprising that "inferior" managers can exist in financial markets, since measuring performance is a difficult task that requires a long time-series of returns (i.e., a long fund lifespan) for precision. Finding evidence of inferior managers is consistent with a world where no managers, or very few managers, can pick stocks, but where consumers have difficulty in precisely determining which managers (if any) may have these skills. More surprising is our finding that a relatively large number of superior stockpickers exist, with skills that are sufficient to more than cover costs (even after adjusting for the return premia accruing to style investing).

D. Bootstrap Analysis for Investment Objective Subgroups

Prior research indicates that managers of growth-oriented funds have better stockpicking talents than managers of income-oriented funds. For example, Chen, Jegadeesh, and Wermers (2000) find that the average growth-oriented fund manager buys stocks with abnormal returns that are two percent per year higher than stocks the manager sells. By contrast, the average income-oriented fund manager does not exhibit any stockpicking talents. Accordingly, we next divide our sample of funds by their self-declared investment objectives to see whether our inferences about the tail of the alpha distribution are affected by the investment style of a fund.

Tables III, IV and V report bootstrap results for each investment objective subgroup, for the unconditional, conditional-β, and conditional-α-and-β versions of the FF model, respectively. For example, Panel A of Table III shows the results of our bootstrap tests, conducted only on those funds having an investment objective of "aggressive-growth." For this subgroup of funds, we can reject, using the t-statistic of alpha as our performance measure, which (as we argued in an earlier section) is a better-behaved statistic, that the observed right tail is purely due to luck. Specifically, our bootstrapped p-values indicate that the various points that we present in the right tail of the alpha distribution are all significant at the 97 percent confidence level. In contrast, we cannot reject that the inferior performance of funds in the left tail is due to bad luck.

The results for the conditional models applied to aggressive-growth funds (Tables IV and V) are consistent with the unconditional model results (Table II). Thus, the evidence is very strong that the high alphas observed among this category of funds are due to the presence of managers with stockpicking talents, but we cannot confirm the presence of inferior aggressive-growth managers.

Results for growth-fund managers are similar for the right-tail of the alpha distribution, but differ somewhat for the left tail. Using the t-statistic of the alpha as a measure of performance, the results in bootstrapped p-values (across all three performance models) that generally reject that poor performance is due to luck.

For growth and income managers, the results are more ambiguous. Specifically, high t-statistics of alphas, are not overwhelmingly attributed to funds with superior managers, according to the bootstrapped p-values. However, the evidence for inferior funds is much stronger for this category, as the majority of bootstrapped p-values allow rejection that luck plays a large role.

Finally, the combined category of balanced funds and income funds shows quite different results than the others. Neither the right-tail nor the left-tail results for this category of funds indicates that we can reject that luck plays the major role in performance outcomes.

We also plot, in Figure IV, the kernel density estimate for the bootstrapped and actual performance distributions for funds grouped by the four investment objective categories. One can clearly see that aggressive growth funds, shown in Panel A, have alphas that are distributed much further to the right than their bootstrapped counterparts. Although more subdued, the same observation can be made for growth funds in Panel B. However, Panel C and, especially, Panel D show that bootstrapped alphas seem to fit their actual alpha counterparts fairly well.

On the whole, our bootstrap results are consistent with past studies: growth-oriented funds exhibit much stronger abilities in picking stocks that beat their benchmarks than income- (value-) oriented funds. These results remain present after trading costs and expenses, once we adjust for the non-normality of growth-fund alphas. The bootstrapped results for income-oriented funds show much weaker evidence of the existence of mutual fund "stars."

V Sensitivity Analysis

The method that we have adopted to analyze the tails of the distribution of alphas is new in the context of evaluating mutual fund performance. Therefore, we conduct a sensitivity analysis of our results, both in terms of the bootstrap resampling procedures, and in terms of the use of ranking information on individual funds versus portfolios of funds. Another issue we investigate is the effect on the results of the minimum number of observations required to include a fund in the bootstrap tests.

We will show, in this section, that our conclusions are robust to the choice of bootstrap procedure. In addition, while our bootstrap results for the extreme sections of the alpha tails is sensitive to the minimum number of return observations (the lifespan of the fund), we find that the p-values of the top and bottom 1-10 percent of individual funds or portfolios of funds are largely insensitive to this choice. These results indicate that our main findings in this paper are robust.

A. Time Series Dependence

The appropriate bootstrap procedure should mimic the sampling mechanism that generated the original data. Since this is unknown, we explore the sensitivity of our results with respect to a host of different assumptions. Our benchmark results assume that, conditional on the factor realizations, the residuals are independently and identically distributed. While this may seem a strong assumption, it does allow for conditional dependence in returns through the time-series behavior of the factors. In addition, this simple bootstrap has some robustness properties against violation of the independence and identical distribution assumption such as heteroskedasticity (Hall (1992)).

To explicitly allow for dependence across returns in different time periods conditional on the factor realizations, we adopt the stationary bootstrap suggested by Politis and Romano (1994). For general dependent data, the simple bootstrap fails to capture the dependence structure of the data and require nontrivial modifications. Politis and Romano (1994) propose a bootstrap method that resamples data blocks of random length to form a pseudo-time series. This procedure uses a sequence of independently and identically distributed variables drawn from a geometric distribution that determines the length of the blocks and a similar sequence of variables with uniform distribution to arrange blocks to yield a stationary pseudo-time series.

The average block length for the stationary bootstrap depends on a parameter, q, and is given by $1/q$. When $q = 1$, we have independent resampling, while when $q = 1/2$, the average block length is 2. To explore the sensitivity of our results with respect to this choice, we vary this parameter from 1 to 0.1 corresponding to an average block length of 10 observations. Table VI shows that there is very little sensitivity to this parameter for the unconditional factor model. Slightly higher sensitivity is found for the conditional beta model (Table VII), particularly when we randomize over the factor draws (factor-residual resampling) in addition to randomizing over the individual funds' residuals. In no case, using the $t-$statistic is the statistical significance for the best fund overturned. Similarly the worst fund is statistically insignificant in all scenarios.

B. Portfolios of Funds

Another experiment is to consider the bootstrapped distribution of portfolios of alpha estimates rather than looking at the sampling distribution of the marginal alpha estimates. In practice most investment strategies would hold not just a single fund of a particular rank and would instead be diversified across several (top) funds. We account for this by computing the statistical significance across equal-weighted portfolios of funds. For the worst performers, these portfolios are formed by starting with the worst fund, including the second worst fund and so on, moving progressively further towards the centre of the performance distribution. Likewise, for the best performers we start with the top fund, then include the second best fund in a two-fund portfolio and so on.

Naturally the setup from Section IV has to be changed slightly when we base the test statistic on portfolios comprising the best or worst funds as opposed to individually ranked funds. The alpha performance measure is now based on $\alpha^p_{0i^*} = (1/i^*) \sum_{i=N-i^*+1}^{N} \alpha_{0i}$, where N is again the total number of funds so that the Nth ranked fund is the top performer, fund number $N-1$ is the second best performer and so on. The null hypothesis that the equal-weighted portfolio comprising the i^*th to the Nth fund does not outperform against the alternative that it genuinely outperforms can be stated as follows:

$$\begin{aligned} H_0 &: \alpha^p_{0i^*} \leq 0 \\ H_A &: \alpha^p_{0i^*} > 0, \end{aligned}$$

Results from this exercise based on 1%, 2%, 3%, 5% and 10% quantile portfolios are presented in Tables VIII and IX. Compared to the earlier analysis based on individual funds, the main effect of using equal-weighted portfolios is to smooth the tail probabilities. There is again some evidence that the performance of portfolios comprising the worst funds with aggressive growth or balanced and income investment objectives reflects random chance. For the best-performing funds there is only evidence that the performance of the funds with a balanced and income investment objective may be due to chance.

C. Length of Data Records

Short-lived funds tend to generate higher dispersion and therefore more extreme alpha estimates than long-lived funds. This leads to non-trivial heteroskedasticity across the distribution of alpha estimates. In an attempt to correct for this effect, the baseline bootstrap results imposed a minimum of 60 observations to exclude funds that are very short-lived. Because of concerns about possible survivorship bias, discussed extensively by Brown, Goetzmann, Ibbotson, and Ross (1992), it is important to investigate the effect of this restriction on our results. Thus, a final point we investigate is the sensitivity of the baseline results with respect to the choice of the minimum number of observations available for the funds. We consider subsamples of funds that have survived for longer than a certain minimum length of time, say $T > T^* = 18, 30, 60$, and 90 months.[28]

The results, reported in Table X for the unconditional model shows that the extreme left tail of the performance distribution are poorly estimated when funds with only 18 or 30 observations are included. As expected, the largest effect is found for the alpha performance measure which, in contrast with the t-statistic, does not scale by the standard error of the estimated alpha. As argued earlier, this causes the problem associated with cross-sectional heterogeneity to get exacerbated in the presence of short data records.

D. Cross-Sectional Bootstrap

To take into account the potential cross-sectional dependence between fund returns we apply a cross-sectional bootstrap. This procedure differs from the algorithm described in section 1 only in that, for a given bootstrap, we now employ the same bootstrap index across all funds. Since some funds may as a result be allocated bootstrap index entries from periods when they did not have a return, we drop a fund if it did not have at least 60 observations after applying the bootstrap index. As Table XI shows our results from the cross-sectional bootstrap support our previous conclusions.

Across all investment objectives the p-values show that the best funds' abnormal performance from the 90% quantile to the 99% quantile cannot be explained by sampling variation. Interestingly, in the left tail of the performance distribution the findings are a bit more sensitive to the assumed bootstrap although on balance the results continue to suggest that the worst funds' performance is genuine and not simply due to bad luck. Panel B and C of Table XI show that we obtain the same conclusions when using the conditional β and conditional α models.

Our conclusions concerning individual investment objectives above are also robust to the use of the cross-sectional bootstrap as Table XII shows. It suggests that the best aggressive growth and

[28]The most complex conditional α and β factor model includes four risk factors and five conditioning variables, which leads to 29 regressors and an intercept. As we impose the condition that there should be at least 30 degrees of freedom, this requires that a fund has 60 observations.

growth funds have genuine stock picking skills whereas there is very little evidence that balanced and income funds possess superior stock picking skills. In the left tail of the performance distribution, the results indicate that the performance of the worst aggressive growth and balanced and income funds can be explained by sampling variation.[29]

VI Monte Carlo Analysis of the Bootstrap Procedures

Our data set of mutual fund performance records consists of an unbalanced panel with residuals that are drawn from a non-normal distribution with fat tails. In order to examine how this affects the properties of our bootstrap methodology such as its size and power under reasonable levels of true abnormal performance, we carry out Monte Carlo simulations using artificially generated data for fund returns. We generate artificial returns data according to the simple single-index factor model:

$$\begin{aligned} r_{it} &= \alpha + \beta X_t + \varepsilon_{it},\ i = 1, ..., N,\ t = 1, ..., T_i, \\ X_t &\sim N(0, \sigma_x^2),\ \varepsilon_{it} \sim N(0, \sigma_\varepsilon^2). \end{aligned} \tag{7}$$

We set $N = 1734$ and $T = 396$. These values correspond to the number of funds in our sample and the length of our time series. The Monte Carlo simulations assume that the factors are normally distributed with mean zero and variance that match that found in the actual data. Residuals are also assumed to be uncorrelated across firms. Based on the sample estimates from the single-factor model in Table I, we set $\beta = 0.903$, while the standard deviation of the return on the factors is fixed at $\sigma_x = 0.04458$ and $\sigma_\varepsilon = 0.0110$.[30]

To study the properties of the bootstrap procedure, we impose the following levels of true abnormal performance (in percent per month) under the alternative: (a) $\alpha = -0.2$, (b) $\alpha = -0.02$, (c) $\alpha = 0$, (d) $\alpha = 0.02$, and (e) $\alpha = 0.2$. For these levels of abnormal performance, we expect the performance of the top funds to be statistically significant in scenarios (d) and (e) and the performance of the bottom funds to be statistically significant in scenarios (a) and (b). Scenario

[29] In unreported tests we applied the cross-sectional bootstrap to individual investment objectives based on the conditional β and conditional α and β model. We found that the results supported our conclusions from the unconditional model.

[30] We only require a single bootstrap sample for each Monte Carlo draw. This lowers the computational burden very substantially and is a consequence of Corollary 2.4 in White (2000). We therefore report the average values for 1000 Monte Carlo replications, each of which uses one bootstrap resample.

(c) imposes the null hypothesis of zero true performance. We would therefore expect both the top and the bottom portfolios to generate statistically insignificant alpha performance.

Table XIII shows that this is exactly what we find. Moreover the table also shows that both the $\alpha-$estimates and the t-statistics identify large positive or negative performance when this is genuinely present. When the genuine abnormal performance is quite large (plus or minus 0.2 percent per month), all quantiles of the cross-sectional performance distribution identify the abnormal performance. On the other hand, when abnormal performance is modest (plus or minus 0.02 percent per month), the far tails of the performance distribution do not identify the abnormal performance, whereas the quantiles a bit further inside the tails do so. Under the assumption of no genuine skills, the p-values of the quantiles are distributed around 50% showing that the bootstrap is properly centered and that its distribution is symmetric as it should be under our setup.[31]

VII Conclusion

We test whether the superior performance of "star" mutual fund managers is due to luck or genuine stockpicking skills. In particular, we examine the statistical significance of the performance of the "best" and "worst" funds, based on their alphas, by means of a flexible bootstrap procedure applied to a large number of unconditional and conditional factor models. Our findings indicate that the performance of the best and worst managers is not due to luck, that is, cannot be explained by sampling variability. We also uncover large differences between the performance of funds with different investment objectives. Whereas there is strong evidence of genuine superior performance for a group of growth-oriented funds, the evidence is much weaker for income-oriented funds.

We expect that the methods adopted in this study will also prove useful in addressing other questions in finance where interest lies in analyzing the best and worst performers drawn from a large population that has been ex-post sorted. The advantages of our approach include not needing to specify the exact shape of the distribution from which alphas are drawn, as well as controlling for potential 'data snooping' biases arising from an ex-post sort on alphas. Our approach also offers potentially higher statistical power . It is well-known that it is difficult to establish the statistical significance of individual funds' alpha estimates which are typically estimated with considerable

[31]In unreported results we found that the conclusions are unaltered if we assume that the residuals ε_{it} follow a $t-$distribution with 6 degrees of freedom and the simulations are based on an unbalanced panel in which the funds have the same life spans as in the original data.

noise. However, by considering the entire distribution of alpha estimates across funds, and thus incorporating information about the number of funds at various levels of performance, our tests exploit additional information that is helpful in identifying abnormal performance.

Finally, our evidence, which points to the existence of fund managers with superior stock-picking skills, may provide significant hope for investors who chase mutual fund stars. However, finding a manager with a large alpha is no guarantee of future abilities–we leave this issue to future research.

Appendix

Description of the Bootstrap Approach

Let F_i be the population cumulative distribution function of the returns data, r_{it}, while the empirical distribution function, $\hat{F}_i$, is given by

$$\hat{F}_i(r) \equiv T_i^{-1} \sum_{t=1}^{T_i} I\{r_{it} \leq r\}, \tag{8}$$

where T_i is the number of return observations available for fund i and $I\{.\}$ is an indicator function which takes a value of one when its argument holds, otherwise is zero. When the number of observations, T_i, goes to infinity, $\hat{F}_i$ converges to F_i, under general conditions. Estimated performance statistics such as $\hat{\theta}_i = \theta_i(\hat{F}_i)$ will also converge, generally speaking, to the population values $\theta_i(F_i)$. Moreover, provided that θ_i is a sufficiently well-behaved functional of its argument, the central limit theorem ensures that $\sqrt{T_i}(\theta_i(\hat{F}_i) - \theta_i(F_i))$ converges in distribution to a normal random variable.

Further refinements in the limiting distribution are possible, specifically those arising from Edgeworth expansions. The bootstrap provides a way to approximate the distribution of $\sqrt{T_i}(\theta_i(\hat{F}_i) - \theta_i(F_i))$ using the distribution of $\sqrt{T_i}(\theta_i(F_i^*) - \theta_i(\hat{F}_i))$, where

$$F_i^*(r) \equiv T_i^{-1} \sum_{t=1}^{T_i} I\{r_{it}^* \leq r\} \tag{9}$$

and $\{r_{it}^*\}$ is a resample of $\{r_{it}\}$, i.e. $r_{it}^* = r_{i\tilde{t}}$, where $\tilde{t}$ is a random variable drawn from $\{1, ..., T_i\}$. Because realizations of F_i^* can be generated at will, we can build up as precise an estimate of the distribution of $\sqrt{T_i}(\theta_i(\hat{F}_i) - \theta_i(F_i))$ as we desire, using realizations of $\sqrt{T_i}(\theta_i(F_i^*) - \theta_i(\hat{F}_i))$. When θ_i is a pivotal statistic, that is the asymptotic distribution of $\sqrt{T_i}(\theta_i(\hat{F}_i) - \theta_i(F_i))$ is independent of nuisance parameters (such as the regression error variance), the bootstrap method can provide a close approximation to the refined Edgeworth approximation to the asymptotic distribution of interest. Otherwise, the bootstrap gives a useful approximation to the normal component of the limiting distribution of interest. In particular, standardized parameter estimators, such as t-statistics are pivotal, whereas nonstandardized estimators, such as $\hat{\alpha}_i$ are non-pivotal. See Hall (1992) for further discussion.

Here, the effect of "luck" due to random sampling variation in the distribution of the performance estimates, $\hat{\alpha}_i$, of a given fund will be reflected in the properties of the distribution of $\hat{\alpha}_i - \alpha_{i0}$, where α_{i0} is the hypothesized value of α_i under the null. We denote this distribution by $G_i(u)$, i.e. $G_i(u) = \Pr(\hat{\alpha}_i - \alpha_{i0} \leq u)$. Since α_i is not known, for each fund we approximate it by the cumulative distribution function of $\hat{\alpha}_i^b - \hat{\alpha}_i$, the bootstrapped performance distribution that is given by

$$\hat{G}_i^*(u) = \frac{1}{B} \sum_{b=1}^{B} I\{\hat{\alpha}_i^b - \hat{\alpha}_i \leq u\}. \tag{10}$$

Here $\hat{\alpha}_i^b$ is the estimate of α_i from the bth realization of the bootstrap method just described.

Second, and perhaps most importantly for our analysis, standard statistical methods cannot be relied on when assessing the performance of the 'best' or 'worst' funds drawn from a larger universe. The identity of the funds with extreme performance is not known in advance, as this information is based on ranking the full set of funds *ex-post*. If the individual funds' performance estimates were joint normally distributed, then the best fund's performance would be the maximum value drawn from a multivariate normal distribution whose dimension depends on the number of funds in existence. This distribution also depends on the entire covariance matrix characterizing the joint distribution of the individual funds' returns and is thus generally not possible to tabulate. Because so many funds are being considered, the sampling distribution from which the best and worst performance estimates are drawn has much fatter tails than the sampling distribution of any one individual fund's performance estimate. If we wish to draw inferences about the significance of the best and worst funds' performance, we thus have to adopt an approach that effectively evaluates the entire distribution of fund performance.[32] For example, suppose we are interested in a particular quantile of the distribution of performance across N funds. This is a function of the individual funds' performance distributions and can thus be written as $q(\hat{F}_1, ..., \hat{F}_N)$.

The bootstrap allows us to simulate the sampling distribution of the performance measures that determine whether the best and worst funds produce genuinely abnormal performance. The question of the existence of abnormal performance is concerned with the tails of the sampling distribution of performance across funds. For each point in the distribution of performance across

[32] Ferson and Schadt (1996) calculate Bonferroni bounds for the maximum and minimum alpha statistics. Although robust to arbitrary assumptions concerning correlations across funds' alpha statistics, this approach can give rise to very conservative inference (since it must hold for all possible correlations) and can thus fail to identify genuine abnormal performance.

funds, we use the bootstrap to compute the percentage of funds that achieve a certain level of performance due to luck:

$$\hat{G}_B(u) = \frac{1}{B \cdot N} \sum_{b=1}^{B} \sum_{i=1}^{N} I\{\hat{\alpha}_i^b - \hat{\alpha}_i \leq u\}. \tag{11}$$

Under the null hypothesis, this distribution approximates that of

$$G_0(u) = N^{-1} \sum_{i=1}^{N} I\{\hat{\alpha}_i - \alpha_{i0} \leq u\}, \tag{12}$$

which is the true distribution of the distance of $\hat{\alpha}$ from its value under the null hypothesis, across our sample of funds.

References

Admati, A., and S. Ross, 1985, Measuring Investment Performance in a Rational Expectations Equilibrium Model, Journal of Business 58, 1-26.

Admati, A.R., S. Bhattacharya, S. Ross, and P. Pfleiderer, 1986, On Timing and Selectivity, Journal of Finance, 41, 715-730.

Alexander, G.J., J.V. Bailey and W.C. Sharpe, 1998, Investments, Prentice Hall.

Bickel, P.J., and Freedman, D.A., 1984, Some Asymptotics on the Bootstrap, Annals of Statistics 9, 1196-1271.

Blake, D. and A. Timmermann, 1998, Mutual Fund Performance: Evidence from the UK, European Review of Finance, 57-77.

Black, A., Fraser, P. and Power, D. 1992, UK unit trust performance: a passive time-varying approach, Journal of Banking and Finance 16, 1015-1033

Breen, W., L.R. Glosten, and R. Jagannathan, 1989, Economic significance of predictable variations in stock index returns, Journal of Finance 44, 1177-1190.

Brown, S.J., W. Goetzmann, R.G. Ibbotson, and S.A. Ross, 1992, Survivorship bias in performance studies, Review of Financial Studies 5, 553-580.

Carhart, M., 1997, On Persistence in Mutual Fund Performance, Journal of Finance, Vol LII, No 1, 57-82.

Chen, H.L., N.Jegadesh and R. Wermers, 2000, An Examination of the Stockholdings and Trades of Fund Managers, Journal of Financial and Quantitative Analysis, September 2000.

Chevalier, J. and G. Ellison,1997, Are Some Mutual Fund Managers Better Than Others? Cross-Sectional Patterns in Behavior and Performance, Journal of Finance, 54, 3.

Christopherson, J.A., W.E. Ferson and D.A. Glassman, 1998, Conditioning Manager Alphas on Economic Information: Another Look at the Persistence of Performance, Review of Financial Studies, Vol.11, No. 1, pp.111-142.

Cochrane, J., 1996, A Cross-Sectional Test of a Production-based Asset Pricing Model, Journal of Political Economy 104, 572-621.

Edelen, Roger M., 1999, Investor Flows and the Assessed Performance of Open-end Mutual Funds, Journal of Financial Economics 53, 439–466.

Efron, B., 1979, Bootstrap methods: Another look at the Jackknife. Annals of Statistics 7, 1-26.

Efron, B., 1987, Better bootstrap confidence intervals, *Journal of the American Statistical Association* 82, 171-200.

Efron, B. and R. J. Tibishari, 1993, *An Introduction to the Bootstrap*, Chapman and Hall.

Ferson, W.E., 1985, Changes in Expected risk Premiums and Security Risk Measures, *Proceedings of the European Finance Association*, August.

Ferson, W.E. and C.R. Harvey, 1993, The risk and predictability of international equity returns, *Review of Financial Studies* 6, 527-566.

Ferson, W.E. and R.W. Schadt, 1996, Measuring Fund Strategy and Performance in Changing Economic Conditions, *Journal of Finance* 51, 425-461.

Ghysels, E., 1998, On Stable Factor Structures in the Pricing of Risk: Do Time-varying Betas Help or Hurt?, *Journal of Finance* 53, 549-574.

Grinblatt, Mark and Sheridan Titman, 1989, "Portfolio Performance Evaluation: Old Issues And New Insights," *Review of Financial Studies*, 2, 393-422.

Grinblatt, M., S. Titman and R. Wermers, 1995, Momentum Investment Strategies, Portfolio Performance, and Herding: A Study of Mutual Fund Behavior, *American Economic Review 85,* 1088-1105 .

Gruber, M. J., 1995, Another puzzle: The growth of actively managed mutual funds, *Journal of Finance*, 51, 783-810.

Hall, P., 1986, On the bootstrap and confidence intervals, *Annals of Statistics* 14, 1431-1452.

Hall, P., 1992, *The Bootstrap and Edgeworth Expansion*, Springer Verlag.

Henriksson, R. 1984, Market timing and mutual fund performance, *Journal of Business* 57, 73-96.

Jagannathan, R., and Z. Wang, 1996, The Conditional CAPM and the Cross-section of Expected Returns, *Journal of Finance*, 51, 3-54.

Jensen, M., 1968, The performance of mutual funds in the period 1945-1964, *Journal of Finance* 23, 389-416.

Leger, L. 1997, UK investment trusts: performance, timing and selectivity, *Applied Economics Letters* 4, 207-210.

Marcus, Alan J., 1990, "The Magellan Fund And Market Efficiency," *Journal of Portfolio Management*, 17, 85-88.

Merton, R. and R. Henriksson, R, 1981, On market timing and investment performance II:

Statistical procedures for evaluating forecasting skills, Journal of Business 54, 513–33.

Newey, W. and K. West, 1987, A Simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, Econometrica 55, 703-708.

Pesaran, M., and A. Timmermann, 1995, Predictability of Stock returns: Robustness and economic significance, Journal of Finance 50, 1201-1228.

Politis, D. N. and Romano, J. P., 1994, The stationary bootstrap, Journal of the American Statistical Association, 89, 1303-1313.

Rosenberg, B., and V. Marathe, 1979, Tests of Capital Asset Pricing Hypotheses, Research in Finance, 1, 115-224.

Shanken, J., 1990, Intertemporal Asset Pricing: An Empirical Investigation, Journal of Econometrics, 35, 99-120.

Sharpe, W., 1998, Major Investment Styles, Journal of Portfolio Management, 68-74.

Silverman, B. W., 1986, Density Estimation for Statistics and Data Analysis, Chapman and Hall: London.

Singh, K. , 1986, Discussion of Wu, Annals of Statistics, 14, 1328-1330.

Treynor, Jack L. and Fischer Black, 1973, "How To Use Security Analysis To Improve Portfolio Selection," Journal of Business, 46, 66-86.

Treynor, J., and K. Mazuy, 1966, Can mutual funds outguess the market?, Harvard Business Review 44, 131-36.

Wermers, R., 1997, Momentum Investment Strategies of Mutual Funds, Performance Persistence, and Survivorship Bias, Working Paper.

Wermers, R., 1999, Mutual Fund Herding and the Impact on Stock Prices, Journal of Finance.

Wermers, R., 2000, Mutual Fund Performance: An Empirical Decomposition into Stock-Picking Talent, Style, Transaction Costs, and Expenses, Journal of Finance 55, 1655 - 1703.

White, H., 2000, A Reality Check for Data Snooping, Econometrica 68, 1097-1126

Table I

Cross-Sectional (Nonbootstrapped) Distribution of Alphas for Different Models

This table presents statistics on the distribution of performance measures (alphas) from several different models. These performance measurement models include the single factor Jensen, Treynor-Mazuy and Henriksson-Merton as well as the the multi-factor Fama-French and Carhart models. Reported in each panel are the cross-sectional: average alpha, average t-statistic, median alpha and standard deviation of alpha (for alpha; we report average t-statistics based on heteroskedasticity and autocorrelation consistent standard errors). Also the Schwartz Information Criterion is presented, which is a model selection criterion that trades off goodness of fit against degrees of freedom. Panel A reports these statistics for unconditional versions of the models, panel B for conditional-beta versions, and panel C for conditional alpha and beta versions. The conditioning variables for panels B and C include the lagged level of the one-month Treasury bill yield (z_{1t}), the lagged dividend yield of the value-weighted New York Stock Exchange and American Stock Exchange (AMEX) stock index (z2t) and a lagged measure of the slope of the term structure (z_{3t}).

Panel A. Unconditional Factor Models

Model number	1	2	3	4	5
Description	Jensen	TM	HM	FF	Carhart
Average α (in percent per month)	0.003	0.039	0.085	0.043	-0.008
Average t_α	-0.004	0.107	0.219	0.193	-0.147
Median α (in percent per month)	0.006	0.025	0.052	0.010	-0.032
Standard Deviation of α	0.255	0.300	0.438	0.288	0.283
Model Selection Criteria and Diagnostics					
Schwartz Information Criterion	0.547	0.551	0.555	0.459	0.460
Rejection of Normality (% of funds)	57.86	55.45	55.24	56.39	54.30

Panel B. Conditional β Models

Model number	6	7	8	9	10	11	12
Description	Jensen	TM	HM	FF I	FF II	Carhart I	Carhart II
Average α (in percent per month)	0.012	0.038	0.064	0.043	0.038	-0.010	-0.021
Average t_α	0.048	0.081	0.105	0.241	0.167	-0.118	-0.205
Median α (in percent per month)	0.017	0.021	0.037	0.014	0.013	-0.025	-0.033
Standard Deviation of α	0.259	0.297	0.407	0.284	0.287	0.282	0.308
Conditioning Variables	z_{1t}, z_{2t}, z_{3t},	z_{1t}, z_{2t}, z_{3t},	z_{1t}, z_{2t}, z_{3t},	z_{1t}	z_{1t}, z_{2t}, z_{3t},	z_{1t}	z_{1t}, z_{2t}, z_{3t},
Model Selection Criteria and Diagnostics							
Schwartz Information Criterion	0.574	0.580	0.627	0.477	0.534	0.490	0.570
Rejection of Normality (% of funds)	54.82	53.98	53.67	53.35	48.58	51.26	47.27

Panel C. Conditional α and β Models

Model number	13	14	15	16	17
Description	Jensen	FF I	FF II	Carhart I	Carhart II
Average α (in percent per month)	0.078	0.048	0.030	0.002	0.001
Average t_α	0.136	0.240	0.057	-0.058	-0.146
Median α (in percent per month)	0.028	0.020	0.010	-0.012	-0.020
Standard Deviation of α	0.433	0.320	0.311	0.415	0.492
Conditioning Variables	z_{1t}, z_{2t}, z_{3t},	z_{1t}	z_{1t}, z_{2t}, z_{3t},	z_{1t}	z_{1t}, z_{2t}, z_{3t},
Model Selection Criteria and Diagnostics					
Schwartz Information Criterion	0.609	0.490	0.575	0.503	0.613
Rejection of Normality (% of funds)	53.93	53.35	48.58	50.47	47.17

Table II

Statistical Significance of the Best and Worst Funds' Performance For All Investment Objectives

The first row reports the OLS estimate of alpha in percent per month for each fund. The second and third row report the t-statistic of alpha based on heteroskedasticity and autocorrelation consistent standard errors as well as the bootstrapped p-value of the t-statistic.The first column on the left (right) reports results for funds with the lowest (highest) alpha and t-statistic followed by results for the funds with the second lowest (highest) alpha and t-statistic and marginal funds at different percentiles in the left (right) tail of the distribution. The p-value is based on the distribution of the best (worst) funds in 1000 bootstrap resamples. The unconditional model in Panel A is based on the three-factor Fama-French model. Panels B and C report results based on the conditional beta Fama-French model and conditional alpha and beta Fama French model, respectively.

Panel A: Unconditional Model

	bottom	2.min	3.min	4.min	5.min	min1%	min3%	min5%	min10%	min40%	max40%	max10%	max5%	max3%	max1%	5.max	4.max	3.max	2.max	top
alpha (%)	-2.361	-1.086	-0.889	-0.868	-0.862	-0.707	-0.480	-0.339	-0.252	-0.031	0.069	0.411	0.544	0.650	0.848	0.954	0.973	1.027	1.115	1.240
t-alpha	-4.499	-4.472	-4.328	-3.833	-3.803	-3.248	-2.381	-2.124	-1.621	-0.252	0.535	2.245	2.824	3.174	3.725	4.007	4.082	4.243	4.308	5.344
p-value (bootstrapped)	0.157	0.018	0.001	0.002	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	0.634	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	0.004	0.004

Panel B: Conditional Beta Model

	bottom	2.min	3.min	4.min	5.min	min1%	min3%	min5%	min10%	min40%	max40%	max10%	max5%	max3%	max1%	5.max	4.max	3.max	2.max	top
alpha (%)	-1.884	-1.190	-1.152	-1.082	-0.949	-0.757	-0.489	-0.377	-0.236	-0.023	0.069	0.400	0.521	0.632	0.853	0.972	0.986	1.028	1.037	1.176
t-alpha	-4.496	-3.987	-3.836	-3.684	-3.527	-3.177	-2.483	-2.122	-1.629	-0.201	0.523	2.362	2.788	3.131	3.903	4.245	4.312	4.358	4.636	5.830
p-value (bootstrapped)	0.127	0.057	0.013	0.006	0.002	< 0.001	< 0.001	< 0.001	< 0.001	0.925	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	0.001

Panel C: Conditional Alpha And Beta Model

	bottom	2.min	3.min	4.min	5.min	min1%	min3%	min5%	min10%	min40%	max40%	max10%	max5%	max3%	max1%	5.max	4.max	3.max	2.max	top
alpha (%)	-1.694	-1.234	-1.230	-1.204	-1.190	-0.784	-0.546	-0.415	-0.278	-0.022	0.079	0.438	0.574	0.692	0.973	1.345	1.376	1.423	1.500	1.649
t-alpha	-5.843	-4.008	-3.826	-3.642	-3.565	-3.115	-2.498	-2.217	-1.645	-0.171	0.539	2.338	2.862	3.238	3.810	4.297	4.329	4.416	5.261	5.894
p-value (bootstrapped)	0.019	0.144	0.062	0.049	0.013	0.002	< 0.001	< 0.001	< 0.001	0.989	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	0.002	0.001	0.009

TABLE III

Statistical Significance of the Best and Worst Funds' Performance by Investment Objective (Unconditional Factor Model)

The first row reports the OLS estimate of alpha in percent per month for each fund. The second and third row report the t-statistic of alpha based on heteroskedasticity and autocorrelation consistent standard errors as well as the bootstrapped p-value of the t-statistic.The first column on the left (right) reports results for funds with the lowest (highest) alpha and t-statistic followed by results for the funds with the second lowest (highest) alpha and t-statistic and marginal funds at different percentiles in the left (right) tail of the distribution. The p-value is based on the distribution of the best (worst) funds in 1000 bootstrap resamples. The results in this table are based on the unconditional Fama-French model. Panels A, B, C and D report results for different investment objectives, which were determined by the first CDA investment objective reported for a fund ignoring any subsequent changes in the objective.

Panel A: Aggressive Growth

	bottom	2.min	3.min	min1%	min3%	min5%	min10%	min40%	max40%	max10%	max5%	max3%	max1%	3.max	2.max	top
alpha (%)	-1.086	-0.889	-0.868	-0.868	-0.666	-0.529	-0.287	0.111	0.270	0.664	0.775	0.848	0.954	0.954	1.027	1.115
t-alpha	-3.259	-3.103	-2.636	-2.636	-2.081	-1.599	-1.105	0.642	1.322	2.975	3.310	3.499	3.826	3.826	3.833	3.947
p-value (bootstrapped)	0.409	0.139	0.255	0.255	0.3	0.791	0.961	1.000	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.003	0.038

Panel B: Growth

	bottom	2.min	3.min	min1%	min3%	min5%	min10%	min40%	max40%	max10%	max5%	max3%	max1%	3.max	2.max	top
alpha (%)	-2.361	-0.796	-0.707	-0.603	-0.477	-0.348	-0.240	-0.032	0.057	0.387	0.478	0.527	0.732	0.873	0.973	1.240
t-alpha	-4.328	-3.803	-3.542	-3.247	-2.354	-2.131	-1.644	-0.252	0.492	2.245	2.761	3.100	4.007	4.243	4.308	5.344
p-value (bootstrapped)	0.056	0.018	0.006	0.001	0.009	<0.001	0.001	0.590	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.004

Panel C: Growth and Income

	bottom	2.min	3.min	min1%	min3%	min5%	min10%	min40%	max40%	max10%	max5%	max3%	max1%	3.max	2.max	top
alpha (%)	-0.749	-0.706	-0.629	-0.629	-0.385	-0.310	-0.265	-0.057	0.000	0.169	0.307	0.374	0.419	0.419	0.624	0.670
t-alpha	-4.499	-4.472	-3.833	-3.833	-2.458	-2.274	-1.817	-0.605	-0.003	1.416	2.192	2.557	2.830	2.830	2.892	2.905
p-value (bootstrapped)	0.027	<0.001	0.001	0.001	0.022	0.001	<0.001	<0.001	1	0.178	<0.001	0.003	0.07	0.07	0.164	0.52

Panel D: Balanced and Income

	bottom	2.min	3.min	min1%	min3%	min5%	min10%	min40%	max40%	max10%	max5%	max3%	max1%	3.max	2.max	top
alpha (%)	-0.547	-0.522	-0.439	-0.522	-0.376	-0.312	-0.184	-0.055	0.008	0.185	0.249	0.267	0.354	0.347	0.354	0.443
t-alpha	-3.114	-2.847	-2.442	-2.847	-2.330	-2.238	-1.712	-0.538	0.066	1.401	2.072	2.290	3.168	2.884	3.168	3.440
p-value (bootstrapped)	0.201	0.069	0.119	0.069	0.065	0.003	0.021	0.013	0.929	0.37	0.044	0.077	0.009	0.002	0.009	0.075

TABLE IV

Statistical Significance of the Best and Worst Funds' Performance by Investment Objective (Conditional Beta Factor Model)

The first row reports the OLS estimate of alpha in percent per month for each fund. The second and third row report the t-statistic of alpha based on heteroskedasticity and autocorrelation consistent standard errors as well as the bootstrapped p-value of the t-statistic. The first column on the left (right) reports results for funds with the lowest (highest) alpha and t-statistic followed by results for the funds with the second lowest (highest) alpha and t-statistic and marginal funds at different percentiles in the left (right) tail of the distribution. The p-value is based on the distribution of the best (worst) funds in 1000 bootstrap resamples. The results in this table are based on the conditional beta three-factor Fama-French model. Panels A, B, C and D report results for different investment objectives, which were determined by the first CDA investment objective reported for a fund ignoring any subsequent changes in the objective.

Panel A: Aggressive Growth

	bottom	2.min	3.min	min1%	min3%	min5%	min10%	min40%	max40%	max10%	max5%	max3%	max1%	3.max	2.max	top
alpha (%)	-1.190	-1.152	-0.949	-0.949	-0.711	-0.598	-0.238	0.137	0.277	0.631	0.731	0.853	0.945	0.945	0.986	1.037
t-alpha	-3.224	-3.002	-2.873	-2.873	-2.464	-1.911	-1.074	0.710	1.454	2.819	3.225	3.405	3.903	3.903	4.002	4.093
p-value (bootstrapped)	0.358	0.178	0.081	0.081	0.014	0.136	0.975	1.000	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.003	0.03

Panel B: Growth

	bottom	2.min	3.min	min1%	min3%	min5%	min10%	min40%	max40%	max10%	max5%	max3%	max1%	3.max	2.max	top
alpha (%)	-1.884	-1.082	-0.901	-0.609	-0.491	-0.391	-0.249	-0.030	0.068	0.336	0.471	0.567	0.725	0.972	1.028	1.176
t-alpha	-3.987	-3.387	-3.245	-2.859	-2.454	-2.132	-1.622	-0.241	0.435	2.387	2.788	3.062	4.245	4.358	4.636	5.830
p-value (bootstrapped)	0.119	0.121	0.057	0.07	0.001	<0.001	<0.001	0.646	0.004	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.001

Panel C: Growth and Income

	bottom	2.min	3.min	min1%	min3%	min5%	min10%	min40%	max40%	max10%	max5%	max3%	max1%	3.max	2.max	top
alpha (%)	-0.757	-0.702	-0.590	-0.590	-0.396	-0.297	-0.231	-0.044	0.007	0.170	0.312	0.354	0.467	0.467	0.560	0.724
t-alpha	-4.496	-3.836	-3.684	-3.684	-2.842	-2.253	-1.823	-0.438	0.050	1.605	1.944	2.513	3.022	3.022	3.120	3.159
p-value (bootstrapped)	0.043	0.008	0.002	0.002	0.001	0.003	<0.001	0.021	0.99	0.008	0.051	0.002	0.019	0.019	0.059	0.305

Panel D: Balanced and Income

	bottom	2.min	3.min	min1%	min3%	min5%	min10%	min40%	max40%	max10%	max5%	max3%	max1%	3.max	2.max	top
alpha (%)	-0.415	-0.414	-0.380	-0.414	-0.282	-0.257	-0.213	-0.049	0.005	0.189	0.227	0.253	0.370	0.332	0.370	0.418
t-alpha	-3.177	-2.630	-2.177	-2.630	-2.127	-1.806	-1.614	-0.450	0.048	1.457	1.808	2.491	3.361	2.665	3.361	4.032
p-value (bootstrapped)	0.171	0.169	0.349	0.169	0.21	0.289	0.056	0.067	0.957	0.243	0.284	0.013	0.006	0.019	0.006	0.014

TABLE V

Statistical Significance of the Best and Worst Funds' Performance by Investment Objective (Conditional Alpha and Beta Factor Model)

The first row reports the OLS estimate of alpha in percent per month for each fund. The second and third row report the t-statistic of alpha based on heteroskedasticity and autocorrelation consistent standard errors as well as the bootstrapped p-value of the t-statistic. The first column on the left (right) reports results for funds with the lowest (highest) alpha and t-statistic followed by results for the funds with the second lowest (highest) alpha and t-statistic and marginal funds at different percentiles in the left (right) tail of the distribution. The p-value is based on the distribution of the best (worst) funds in 1000 bootstrap resamples. The results in this table are based on the conditional alpha and beta Fama-French model. Panels A, B, C and D report results for different investment objectives, which were determined by the first CDA investment objective reported for a fund ignoring any subsequent changes in the objective.

Panel A: Aggressive Growth

	bottom	2.min	3.min	min1%	min3%	min5%	min10%	min40%	max40%	max10%	max5%	max3%	max1%	3.max	2.max	top
alpha (%)	-1.694	-1.234	-1.230	-1.230	-0.737	-0.610	-0.344	0.138	0.279	0.672	0.758	0.893	1.119	1.119	1.274	1.423
t-alpha	-5.843	-3.023	-2.968	-2.968	-2.327	-1.808	-1.202	0.540	1.358	2.851	3.426	3.695	3.810	3.810	4.129	4.186
p-value (bootstrapped)	0.001	0.231	0.083	0.083	0.075	0.403	0.893	1.000	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.001	0.043

Panel B: Growth

	bottom	2.min	3.min	min1%	min3%	min5%	min10%	min40%	max40%	max10%	max5%	max3%	max1%	3.max	2.max	top
alpha (%)	-1.204	-1.190	-0.900	-0.740	-0.574	-0.423	-0.318	-0.024	0.071	0.410	0.541	0.643	0.973	1.376	1.500	1.649
t-alpha	-3.826	-3.558	-3.232	-3.115	-2.488	-2.287	-1.585	-0.181	0.477	2.383	2.862	3.061	4.297	4.416	5.261	5.894
p-value (bootstrapped)	0.354	0.165	0.16	0.025	0.014	0.001	0.034	0.911	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.005

Panel C: Growth and Income

	bottom	2.min	3.min	min1%	min3%	min5%	min10%	min40%	max40%	max10%	max5%	max3%	max1%	3.max	2.max	top
alpha (%)	-0.784	-0.665	-0.502	-0.502	-0.421	-0.314	-0.245	-0.053	0.018	0.219	0.276	0.421	0.498	0.498	0.518	0.631
t-alpha	-4.008	-3.642	-3.565	-3.565	-2.624	-2.145	-1.810	-0.472	0.168	1.616	2.261	2.479	2.981	2.981	3.246	3.483
p-value (bootstrapped)	0.166	0.048	0.007	0.007	0.007	0.017	0.002	0.013	0.844	0.025	0.004	0.026	0.062	0.062	0.086	0.252

Panel D: Balanced and Income

	bottom	2.min	3.min	min1%	min3%	min5%	min10%	min40%	max40%	max10%	max5%	max3%	max1%	3.max	2.max	top
alpha (%)	-0.555	-0.484	-0.413	-0.484	-0.371	-0.352	-0.230	-0.043	0.005	0.243	0.298	0.324	0.420	0.347	0.420	0.528
t-alpha	-2.877	-2.855	-2.597	-2.855	-2.476	-2.374	-1.783	-0.544	0.052	1.633	2.462	2.816	3.407	3.100	3.407	4.035
p-value (bootstrapped)	0.445	0.13	0.102	0.13	0.058	0.007	0.026	0.016	0.944	0.081	0.003	0.001	0.009	0.003	0.009	0.034

TABLE VI
Bootstrap Procedure Sensitivity Analysis - Unconditional Model

This table reports a sensitivity analysis of the statistical significance of the worst and best fund's performance to various bootstrap procedures.The first row reports the OLS estimate of alpha in percent per month for each fund. The second and third row report the t-statistic of alpha based on heteroskedasticity and autocorrelation consistent standard errors as well as the bootstrapped p-value of the t-statistic. The benchmark bootstrap in column 1 uses the uniform bootstrap to resample residuals. Columns 2 to 4 report results for the Politis and Romano (1994) stationary bootstrap using various smoothing parameters q. Column 5 reports results for factor-residual resampling instead of residual resampling. The p-value is based on the distribution of the best (worst) funds in 1000 bootstrap resamples.The results in this table are based on the unconditional Fama-French model.

Panel A. Sensitivity Analysis of the Worst Fund's Performance

	Benchmark Scenario	Stationary Bootstrap q=0.1	Stationary Bootstrap q=0.5	Stationary Bootstrap q=1	Factor-Residual Resampling
alpha (%)	-2.361	-2.361	-2.361	-2.361	-2.361
t-alpha	-4.499	-4.499	-4.499	-4.499	-4.499
p-value (bootstrapped)	0.157	0.296	0.178	0.157	0.276

Panel B. Sensitivity Analysis of the Best Fund's Performance

	Benchmark Scenario	Stationary Bootstrap q=0.1	Stationary Bootstrap q=0.5	Stationary Bootstrap q=1	Factor-Residual Resampling
alpha (%)	1.240	1.240	1.240	1.240	1.240
t-alpha	5.344	5.344	5.344	5.344	5.344
p-value (bootstrapped)	0.004	0.03	0.006	0.003	0.004

TABLE VII

Bootstrap Procedure Sensitivity Analysis - Conditional Model

This table reports a sensitivity analysis of the statistical significance of the worst and best fund's performance to various bootstrap procedures. The first row reports the OLS estimate of alpha in percent per month for each fund. The second and third row report the t-statistic of alpha based on heteroskedasticity and autocorrelation consistent standard errors as well as the bootstrapped p-value of the t-statistic. The benchmark bootstrap in column 1 uses the uniform bootstrap to resample residuals. Columns 2 to 4 report results for the Politis and Romano (1994) stationary bootstrap using various smoothing parameters q. Column 5 reports results for factor-residual resampling instead of residual resampling. The p-value is based on the distribution of the best (worst) funds in 1000 bootstrap resamples.The results in this table are based on the preferred conditional beta Fama-French model with three risk factors and one conditioning variable.

Panel A. Sensitivity Analysis of the Worst Fund's Performance

	Benchmark Scenario	Stationary Bootstrap q=0.1	Stationary Bootstrap q=0.5	Stationary Bootstrap q=1	Factor-Residual Resampling
alpha (%)	-1.884	-1.884	-1.884	-1.884	-1.884
t-alpha	-4.496	-4.496	-4.496	-4.496	-4.496
p-value (bootstrapped)	0.127	0.298	0.141	0.124	0.228

Panel B. Sensitivity Analysis of the Best Fund's Performance

	Benchmark Scenario	Stationary Bootstrap q=0.1	Stationary Bootstrap q=0.5	Stationary Bootstrap q=1	Factor-Residual Resampling
alpha (%)	1.176	1.176	1.176	1.176	1.176
t-alpha	5.830	5.830	5.830	5.830	5.830
p-value (bootstrapped)	0.001	0.014	<0.001	<0.001	<0.001

TABLE VIII

Statistical Significance of Equally-Weighted Portfolios of Funds - Unconditional Model

The first row reports the OLS estimate of alpha in percent per month for each porfolio of funds. The second and third row report the t-statistic of alpha based on heteroskedasticity and autocorrelation consistent standard errors as well as the bootstrapped p-value of the t-statistic. The first column on the left (right) reports results for the portfolio formed from funds in the lowest (highest) percentile of the alpha and t-statistic distribution followed by results for a portfolio that also includes funds in the next lowest (highest) percentile in the left (right) tail of the alpha and t-statistic. The p-value is based on the distribution of the best (worst) funds in 1000 bootstrap resamples. The results are based on the preferred unconditional three-factor Fama-French model. Panels A, B, C, D and E report results for different investment objectives, which were determined by the first CDA investment objective reported for a fund ignoring any subsequent changes in the objective.

Panel A: All Investment objectives

	min1%	min2%	min3%	min5%	min10%	max10%	max5%	max3%	max2%	max1%
alpha (%)	-0.982	-0.800	-0.712	-0.589	-0.441	0.595	0.724	0.808	0.867	0.970
t-alpha	-3.783	-3.377	-3.110	-2.764	-2.297	2.951	3.373	3.646	3.813	4.124
p-value (bootstrapped)	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

Panel B: Aggressive Growth

	min1%	min2%	min3%	min5%	min10%	max10%	max5%	max3%	max2%	max1%
alpha (%)	-0.948	-0.892	-0.839	-0.738	-0.552	0.802	0.895	0.945	0.979	1.032
t-alpha	-3.000	-2.735	-2.550	-2.251	-1.830	3.398	3.620	3.737	3.800	3.869
p-value (bootstrapped)	0.277	0.284	0.298	0.39	0.588	<0.001	<0.001	<0.001	<0.001	<0.001

Panel C: Growth

	min1%	min2%	min3%	min5%	min10%	max10%	max5%	max3%	max2%	max1%
alpha (%)	-1.025	-0.818	-0.714	-0.595	-0.440	0.533	0.638	0.722	0.796	0.911
t-alpha	-3.667	-3.353	-3.130	-2.782	-2.318	2.972	3.444	3.786	4.035	4.397
p-value (bootstrapped)	0.003	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

Panel D: Growth and Income

	min1%	min2%	min3%	min5%	min10%	max10%	max5%	max3%	max2%	max1%
alpha (%)	-0.695	-0.612	-0.547	-0.454	-0.371	0.320	0.416	0.467	0.503	0.571
t-alpha	-4.268	-3.896	-3.533	-3.048	-2.553	2.150	2.550	2.726	2.783	2.876
p-value (bootstrapped)	<0.001	<0.001	<0.001	<0.001	<0.001	0.008	0.014	0.046	0.109	0.216

Panel E: Balanced and Income

	min1%	min2%	min3%	min5%	min10%	max10%	max5%	max3%	max2%	max1%
alpha (%)	-0.534	-0.503	-0.471	-0.421	-0.330	0.270	0.319	0.353	0.382	0.399
t-alpha	-2.981	-2.801	-2.683	-2.544	-2.223	2.195	2.661	2.945	3.164	3.304
p-value (bootstrapped)	0.124	0.107	0.088	0.045	0.021	0.024	0.01	0.011	0.013	0.025

TABLE IX

Statistical Significance of Equally-Weighted Portfolios of Funds - Conditional Model

The first row reports the OLS estimate of alpha in percent per month for each porfolio of funds. The second and third row report the t-statistic of alpha based on heteroskedasticity and autocorrelation consistent standard errors as well as the bootstrapped p-value of the t-statistic. The first column on the left (right) reports results for the portfolio formed from funds in the lowest (highest) percentile of the alpha and t-statistic distribution followed by results for a portfolio that also includes funds in the next lowest (highest) percentile in the left (right) tail of the alpha and t-statistic. The p-value is based on the distribution of the best (worst) funds in 1000 bootstrap resamples. The results are based on the preferred conditional three-factor Fama-French model with one conditioning variable. Panels A, B, C, D and E report results for different investment objectives, which were determined by the first CDA investment objective reported for a fund ignoring any subsequent changes in the objective.

Panel A: All Investment objectives

	min1%	min2%	min3%	min5%	min10%	max10%	max5%	max3%	max2%	max1%
alpha (%)	-1.049	-0.849	-0.752	-0.616	-0.441	0.595	0.708	0.790	0.849	0.959
t-alpha	-3.581	-3.220	-3.020	-2.733	-2.297	2.951	3.396	3.699	3.915	4.365
p-value (bootstrapped)	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

Panel B: Aggressive Growth

	min1%	min2%	min3%	min5%	min10%	max10%	max5%	max3%	max2%	max1%
alpha (%)	-1.097	-1.008	-0.940	-0.837	-0.609	0.772	0.864	0.916	0.941	0.989
t-alpha	-3.033	-2.943	-2.809	-2.583	-1.995	3.305	3.585	3.721	3.831	4.000
p-value (bootstrapped)	0.187	0.086	0.061	0.033	0.188	<0.001	<0.001	<0.001	<0.001	0.001

Panel C: Growth

	min1%	min2%	min3%	min5%	min10%	max10%	max5%	max3%	max2%	max1%
alpha (%)	-1.026	-0.829	-0.735	-0.615	-1.026	0.929	0.645	0.736	0.805	0.929
t-alpha	-3.303	-3.029	-2.867	-2.653	-3.303	4.676	3.527	3.904	4.225	4.676
p-value (bootstrapped)	0.064	0.027	0.011	0.001	0.064	<0.001	<0.001	<0.001	<0.001	<0.001

Panel D: Growth and Income

	min1%	min2%	min3%	min5%	min10%	max10%	max5%	max3%	max2%	max1%
alpha (%)	-0.683	-0.583	-0.530	-0.452	-0.361	0.320	0.412	0.470	0.514	0.583
t-alpha	-4.005	-3.758	-3.506	-3.052	-2.538	2.192	2.593	2.862	2.978	3.100
p-value (bootstrapped)	0.003	<0.001	<0.001	<0.001	<0.001	0.004	0.004	0.014	0.032	0.083

Panel E: Balanced and Income

	min1%	min2%	min3%	min5%	min10%	max10%	max5%	max3%	max2%	max1%
alpha (%)	-0.415	-0.403	-0.373	-0.337	-0.285	0.258	0.305	0.343	0.373	0.394
t-alpha	-2.904	-2.661	-2.528	-2.295	-2.015	2.255	2.798	3.137	3.352	3.696
p-value (bootstrapped)	0.155	0.195	0.188	0.222	0.153	0.011	0.004	0.003	0.003	0.003

TABLE X

Sensitivity Analysis of Statistical Significance of Performance Measures to Minimum Number of Observations

This table shows the effect on the bootstrapped performance measures of varying the minimum number of observations required for a fund to be included in the sample. The first row reports the OLS estimate of alpha in percent per month for each fund. The second and third row report the t-statistic of alpha based on heteroskedasticity and autocorrelation consistent standard errors as well as the bootstrapped p-value of the t-statistic. The first column on the left (right) reports results for funds with the lowest (highest) alpha and t-statistic followed by results for the funds with the second lowest (highest) alpha and t-statistic and marginal funds at different percentiles in the left (right) tail of the distribution. The p-value is based on the distribution of the best (worst) funds in 1000 bootstrap resamples. The results are based on the unconditional Fama-French model and all investment objectives.

Panel A: Minimum of 18 observations per fund

	bottom	2.min	3.min	5.min	min1%	min3%	min5%	min10%	max10%	max5%	max3%	max1%	5.max	3.max	2.max	top
alpha (%)	-2.361	-1.656	-1.086	-1.028	-0.845	-0.603	-0.496	-0.316	0.414	0.583	0.717	1.092	1.428	1.745	2.131	2.637
t-alpha	-6.208	-5.320	-4.965	-4.499	-3.426	-2.769	-2.274	-1.738	2.072	2.678	2.995	3.655	4.243	4.715	5.344	6.136
p-value (bootstrapped)	0.042	0.011	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.004	0.004	0.011	0.061

Panel B : Minimum of 30 observations per fund

	bottom	2.min	3.min	5.min	min1%	min3%	min5%	min10%	max10%	max5%	max3%	max1%	5.max	3.max	2.max	top
alpha (%)	-2.361	-1.656	-1.086	-0.989	-0.765	-0.538	-0.429	-0.287	0.411	0.551	0.668	0.918	1.115	1.240	1.321	1.745
t-alpha	-5.320	-4.965	-4.631	-4.472	-3.417	-2.643	-2.238	-1.670	2.108	2.749	3.109	3.672	4.082	4.308	5.344	6.136
p-value (bootstrapped)	0.066	0.011	0.002	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.001	0.001	0.017

Panel C : Minimum of 60 observations per fund

	bottom	2.min	3.min	5.min	min1%	min3%	min5%	min10%	max10%	max5%	max3%	max1%	5.max	3.max	2.max	top
alpha (%)	-2.361	-1.086	-0.889	-0.862	-0.707	-0.480	-0.339	-0.252	0.411	0.544	0.650	0.848	0.954	1.027	1.115	1.240
t-alpha	-4.499	-4.472	-4.328	-3.803	-3.248	-2.381	-2.124	-1.621	2.245	2.824	3.174	3.725	4.007	4.243	4.308	5.344
p-value (bootstrapped)	0.157	0.018	0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.004	0.004

Panel D: Minimum of 90 observations per fund

	bottom	2.min	3.min	5.min	min1%	min3%	min5%	min10%	max10%	max5%	max3%	max1%	5.max	3.max	2.max	top
alpha (%)	-1.086	-0.889	-0.868	-0.796	-0.745	-0.462	-0.322	-0.232	0.426	0.547	0.648	0.848	0.879	0.954	1.027	1.240
t-alpha	-4.472	-4.328	-3.833	-3.542	-3.196	-2.354	-2.124	-1.621	2.335	2.851	3.284	3.826	3.947	4.082	4.308	5.344
p-value (bootstrapped)	0.138	0.009	0.011	0.002	<0.001	0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.001	<0.001

TABLE XI

Cross-Sectional Bootstrap - Statistical Significance of the Best and Worst Funds' Performance

This table reports results for a cross-sectional bootstrap where the bootstrap index is the same for all funds for each bootstrap. The first row reports the OLS estimate of alpha in percent per month for each fund. The second and third row report the t-statistic of alpha based on heteroskedasticity and autocorrelation consistent standard errors as well as the bootstrapped p-value of the t-statistic. The first column on the left (right) reports results for funds with the lowest (highest) alpha and t-statistic followed by results for the funds with the second lowest (highest) alpha and t-statistic and marginal funds at different percentiles in the left (right) tail of the distribution. The p-value is based on the distribution of the best (worst) funds in 1000 bootstrap resamples. The unconditional model in Panel A is based on the three-factor Fama-French model. Panels B and C report results based on the conditional beta Fama-French model and conditional alpha and beta Fama French model, respectively.

Panel A: Unconditional Model

	bottom	2.min	3.min	5.min	min1%	min3%	min5%	min10%	max10%	max5%	max3%	max1%	10.max	5.max	3.max	top
alpha (%)	-2.361	-1.086	-0.889	-0.862	-0.707	-0.480	-0.339	-0.252	0.411	0.544	0.650	0.848	0.848	0.954	1.027	1.240
t-alpha	-4.499	-4.472	-4.328	-3.803	-3.248	-2.381	-2.124	-1.621	2.245	2.824	3.174	3.725	3.725	4.007	4.243	5.344
p-value (bootstrapped)	0.264	0.073	0.037	0.037	0.039	0.115	0.084	0.138	<0.000	<0.000	<0.000	0.001	0.001	0.001	0.003	0.008

Panel B: Conditional Beta Model

	bottom	2.min	3.min	5.min	min1%	min3%	min5%	min10%	max10%	max5%	max3%	max1%	10.max	5.max	3.max	top
alpha (%)	-1.884	-1.190	-1.152	-0.949	-0.757	-0.489	-0.377	-0.236	0.400	0.521	0.632	0.853	0.853	0.972	1.028	1.176
t-alpha	-4.496	-3.987	-3.836	-3.527	-3.177	-2.483	-2.122	-1.629	2.362	2.788	3.131	3.903	3.903	4.245	4.358	5.830
p-value (bootstrapped)	0.251	0.119	0.057	0.046	0.026	0.046	0.068	0.115	<0.000	<0.000	<0.000	<0.000	<0.000	<0.000	0.001	0.001

Panel C: Conditional Alpha And Beta Model

	bottom	2.min	3.min	5.min	min1%	min3%	min5%	min10%	max10%	max5%	max3%	max1%	10.max	5.max	3.max	top
alpha (%)	-1.694	-1.234	-1.230	-1.190	-0.784	-0.546	-0.415	-0.278	0.438	0.574	0.692	0.973	0.973	1.345	1.423	1.649
t-alpha	-5.843	-4.008	-3.826	-3.565	-3.115	-2.498	-2.217	-1.645	2.338	2.862	3.238	3.810	3.810	4.297	4.416	5.894
p-value (bootstrapped)	0.038	0.187	0.122	0.077	0.062	0.05	0.032	0.09	<0.000	<0.000	<0.000	<0.000	<0.000	0.001	0.002	0.008

TABLE XII

Cross-Sectional Bootstrap - Statistical Significance of the Best and Worst Funds' Performance by Investment Objective

This table reports results for a cross-sectional bootstrap where the bootstrap index is the same for all funds for each bootstrap. The first row reports the OLS estimate of alpha in percent per month for each fund. The second and third row report the t-statistic of alpha based on heteroskedasticity and autocorrelation consistent standard errors as well as the bootstrapped p-value of the t-statistic. The first column on the left (right) reports results for funds with the lowest (highest) alpha and t-statistic followed by results for the funds with the second lowest (highest) alpha and t-statistic and marginal funds at different percentiles in the left (right) tail of the distribution. The p-value is based on the distribution of the best (worst) funds in 1000 bootstrap resamples. The results in this table are based on the unconditional Fama-French model. Panels A, B, C and D report results for different investment objectives, which were determined by the first CDA investment objective reported for a fund ignoring any subsequent changes in the objective.

Panel A: Aggressive Growth

	bottom	2.min	3.min	min1%	min2%	min3%	min4%	min5%	min10%	max10%	max5%	max4%	max3%	max2%	max1%	3.max	2.max	top
alpha (%)	-1.086	-0.889	-0.868	-0.868	-0.755	-0.666	-0.586	-0.529	-0.287	0.664	0.775	0.822	0.848	0.879	0.954	0.954	1.027	1.115
t-alpha	-3.259	-3.103	-2.636	-2.636	-2.198	-2.081	-1.710	-1.599	-1.105	2.975	3.310	3.485	3.499	3.671	3.826	3.826	3.833	3.947
p-value (bootstrapped)	0.395	0.174	0.299	0.423	0.499	0.409	0.636	0.643	0.762	<0.000	<0.000	<0.000	<0.000	<0.000	<0.000	<0.000	0.001	0.025

Panel B: Growth

	bottom	2.min	3.min	min1%	min2%	min3%	min4%	min5%	min10%	max10%	max5%	max4%	max3%	max2%	max1%	3.max	2.max	top
alpha (%)	-2.361	-0.796	-0.707	-0.603	-0.508	-0.477	-0.423	-0.348	-0.240	0.387	0.478	0.507	0.527	0.607	0.732	0.873	0.973	1.240
t-alpha	-4.328	-3.803	-3.542	-3.247	-2.772	-2.354	-2.214	-2.131	-1.644	2.245	2.761	2.894	3.100	3.366	4.007	4.243	4.308	5.344
p-value (bootstrapped)	0.092	0.048	0.037	0.033	0.045	0.136	0.12	0.083	0.11	<0.000	<0.000	<0.000	<0.000	<0.000	<0.000	<0.000	0.002	0.002

Panel C: Growth and Income

	bottom	2.min	3.min	min1%	min2%	min3%	min4%	min5%	min10%	max10%	max5%	max4%	max3%	max2%	max1%	3.max	2.max	top
alpha (%)	-0.749	-0.706	-0.629	-0.629	-0.417	-0.385	-0.322	-0.310	-0.265	0.169	0.307	0.347	0.374	0.391	0.419	0.419	0.624	0.670
t-alpha	-4.499	-4.472	-3.833	-3.833	-3.248	-2.458	-2.381	-2.274	-1.817	1.416	2.192	2.318	2.557	2.618	2.830	2.830	2.892	2.905
p-value (bootstrapped)	0.172	0.017	0.019	0.02	0.021	0.125	0.077	0.065	0.056	0.338	0.044	0.05	0.032	0.088	0.118	0.118	0.211	0.49

Panel D: Balanced and Income

	bottom	2.min	3.min	min1%	min2%	min3%	min4%	min5%	min10%	max10%	max5%	max4%	max3%	max2%	max1%	3.max	2.max	top
alpha (%)	-0.547	-0.522	-0.439	-0.522	-0.439	-0.376	-0.330	-0.312	-0.184	0.185	0.249	0.255	0.267	0.347	0.354	0.347	0.354	0.443
t-alpha	-3.114	-2.847	-2.442	-2.847	-2.442	-2.330	-2.293	-2.238	-1.712	1.401	2.072	2.110	2.290	2.884	3.168	2.884	3.168	3.440
p-value (bootstrapped)	0.309	0.19	0.244	0.213	0.252	0.214	0.162	0.129	0.188	0.31	0.1	0.138	0.117	0.026	0.031	0.024	0.027	0.086

TABLE XIII

Monte Carlo Study of the Power of the Bootstrap

This table reports a Monte Carlo study of the statistical significance of the best and worst funds' performance using artificially generated data for fund returns based on a simple single-factor model and an unbalanced panel in which the funds have the same life spans as in the original data. We impose the following underperformance or outperformance (in percent per month): (a) α=-0.2, (b) α=-0.02, (c) α=0, (d) α=0.02 and (e) α=0.2. We calculate p-values to test whether the statistical significance tests correctly identify the statistical significance of the true performance imposed in the data. The tables are based on 1000 Monte Carlo simulations and one bootstrap resample. The first row reports the OLS estimate of alpha in percent per month for each fund. The second and third row report the t statistic of alpha based on heteroskedasticity and autocorrelation consistent standard errors as well as the bootstrapped p-value of the t-statistic. The first column on the left (right) reports results for funds with the lowest (highest) alpha and t-statistic followed by results for the funds with the second lowest (highest) alpha and t-statistic and marginal funds at different percentiles in the left (right) tail of the distribution.

Panel A: (a) α = -0.2

	bottom	2.min	3.min	5.min	10.min	min1%	min3%	min5%	min10%	max10%	max5%	max3%	max1%	10.max	5.max	3.max	2.max	top
alpha (%)	-0.54	-0.50	-0.48	-0.46	-0.43	-0.43	-0.38	-0.35	-0.31	-0.09	-0.05	-0.02	0.03	0.03	0.06	0.08	0.10	0.14
t-alpha	-7.11	-6.70	-6.48	-6.22	-5.86	-5.86	-5.24	-4.91	-4.40	-0.93	-0.51	-0.24	0.26	0.26	0.55	0.77	0.95	1.31
p-value (bootstrapped)	<0.000	<0.000	<0.000	<0.000	<0.000	<0.000	<0.000	<0.000	<0.000	1	1	1	1	1	1	1	1	1

Panel B: α = -0.02

	bottom	2.min	3.min	5.min	10.min	min1%	min3%	min5%	min10%	max10%	max5%	max3%	max1%	10.max	5.max	3.max	2.max	top
alpha (%)	-0.36	-0.32	-0.30	-0.28	-0.25	-0.25	-0.20	-0.17	-0.13	0.09	0.13	0.16	0.21	0.21	0.24	0.26	0.28	0.32
t-alpha	-3.80	-3.47	-3.28	-3.05	-2.74	-2.74	-2.25	-1.98	-1.59	1.06	1.44	1.71	2.20	2.20	2.51	2.72	2.91	3.30
p-value (bootstrapped)	0.3	0.244	0.194	0.142	0.083	0.083	0.012	0.003	<0.000	0.999	0.994	0.98	0.916	0.916	0.847	0.811	0.775	0.654

Panel C: α = 0

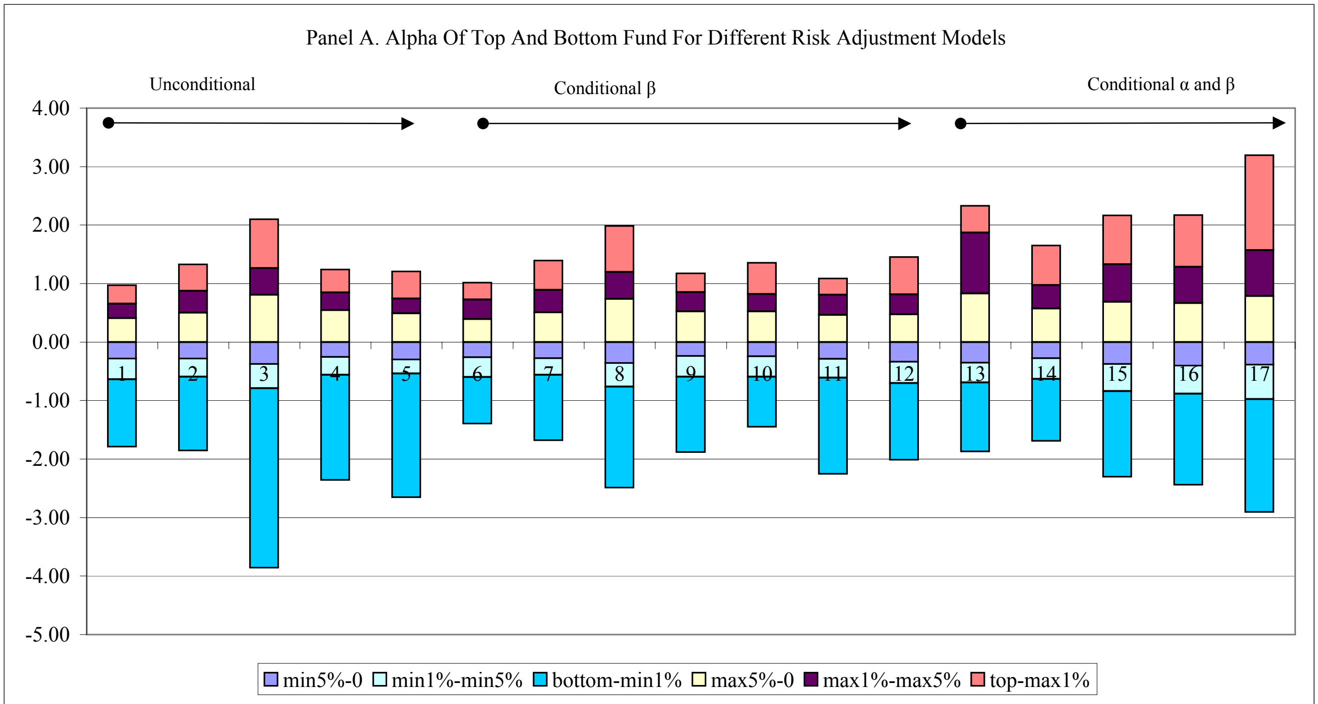
	bottom	2.min	3.min	5.min	10.min	min1%	min3%	min5%	min10	max10	max5%	max3%	max1%	10.max	5.max	3.max	2.max	top
alpha (%)	-0.34	-0.30	-0.28	-0.26	-0.23	-0.23	-0.18	-0.15	-0.11	0.11	0.15	0.18	0.23	0.23	0.26	0.28	0.30	0.34
t-alpha	-3.52	-3.19	-3.00	-2.77	-2.47	-2.47	-1.97	-1.71	-1.32	1.32	1.70	1.97	2.46	2.46	2.77	2.98	3.17	3.56
p-value (bootstrapped)	0.494	0.514	0.499	0.498	0.478	0.478	0.491	0.484	0.482	0.494	0.517	0.511	0.528	0.528	0.521	0.531	0.519	0.492

Panel D: α = 0.02

	bottom	2.min	3.min	5.min	10.min	min1%	min3%	min5%	min10%	max10%	max5%	max3%	max1%	10.max	5.max	3.max	2.max	top
alpha (%)	-0.32	-0.28	-0.26	-0.24	-0.21	-0.21	-0.16	-0.13	-0.09	0.13	0.17	0.20	0.25	0.25	0.28	0.30	0.32	0.36
t-alpha	-3.26	-2.92	-2.74	-2.51	-2.20	-2.20	-1.71	-1.45	-1.06	1.59	1.98	2.25	2.74	2.74	3.04	3.26	3.46	3.84
p-value (bootstrapped)	0.7	0.737	0.781	0.84	0.92	0.92	0.984	0.992	0.997	0.001	0.005	0.013	0.08	0.08	0.145	0.21	0.255	0.329

Panel E: α = 0.2

	bottom	2.min	3.min	5.min	10.min	min1%	min3%	min5%	min10%	max10%	max5%	max3%	max1%	10.max	5.max	3.max	2.max	top
alpha (%)	-0.14	-0.10	-0.08	-0.06	-0.03	-0.03	0.02	0.05	0.09	0.31	0.35	0.38	0.43	0.43	0.46	0.48	0.50	0.54
t-alpha	-1.30	-0.97	-0.79	-0.57	-0.27	-0.27	0.24	0.51	0.93	4.40	4.91	5.25	5.85	5.85	6.22	6.47	6.68	7.08
p-value (bootstrapped)	0.999	1	1	1	1	1	1	1	1	<0.000	<0.000	<0.000	<0.000	<0.000	<0.000	<0.000	<0.000	<0.000



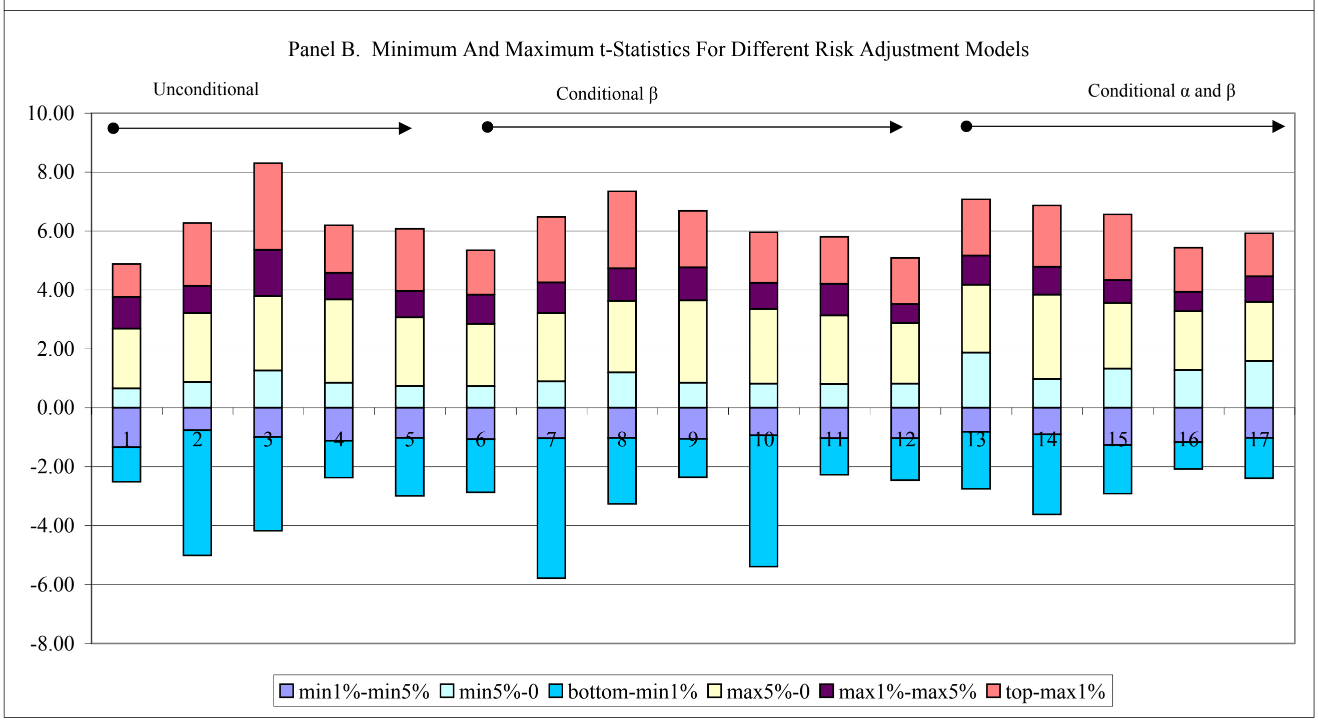


Figure I. For each factor model the columns in Panel A show the performance of the best and worst fund as measured by the monthly α (in percent). Each column also shows the performance of the fund with the marginal 1% and 5% alpha in the right (left)tail of the distribution as a percentage of the performance of the top (bottom) fund. Similarly, the columns in Panel B show the lowest and highest (marginal 1%, marginal 5%) t-statistic using heteroskedasticity and autocorrelation consistent standard errors. All models are estimated for funds that have at least 60 net return monthly observations. The models include unconditional, conditional beta and conditional alpha and beta versions of the following performance measures described in Table I: the Jensen model (models 1, 6 and 13), the Treynor-Mazuy model (models 2 and 7), the Henriksson-Merton model (model 3 and 8), the Fama-French model (models 4,9 and 13, 15) and the Carhart model (models 5, 11, 12, 16 and 17).

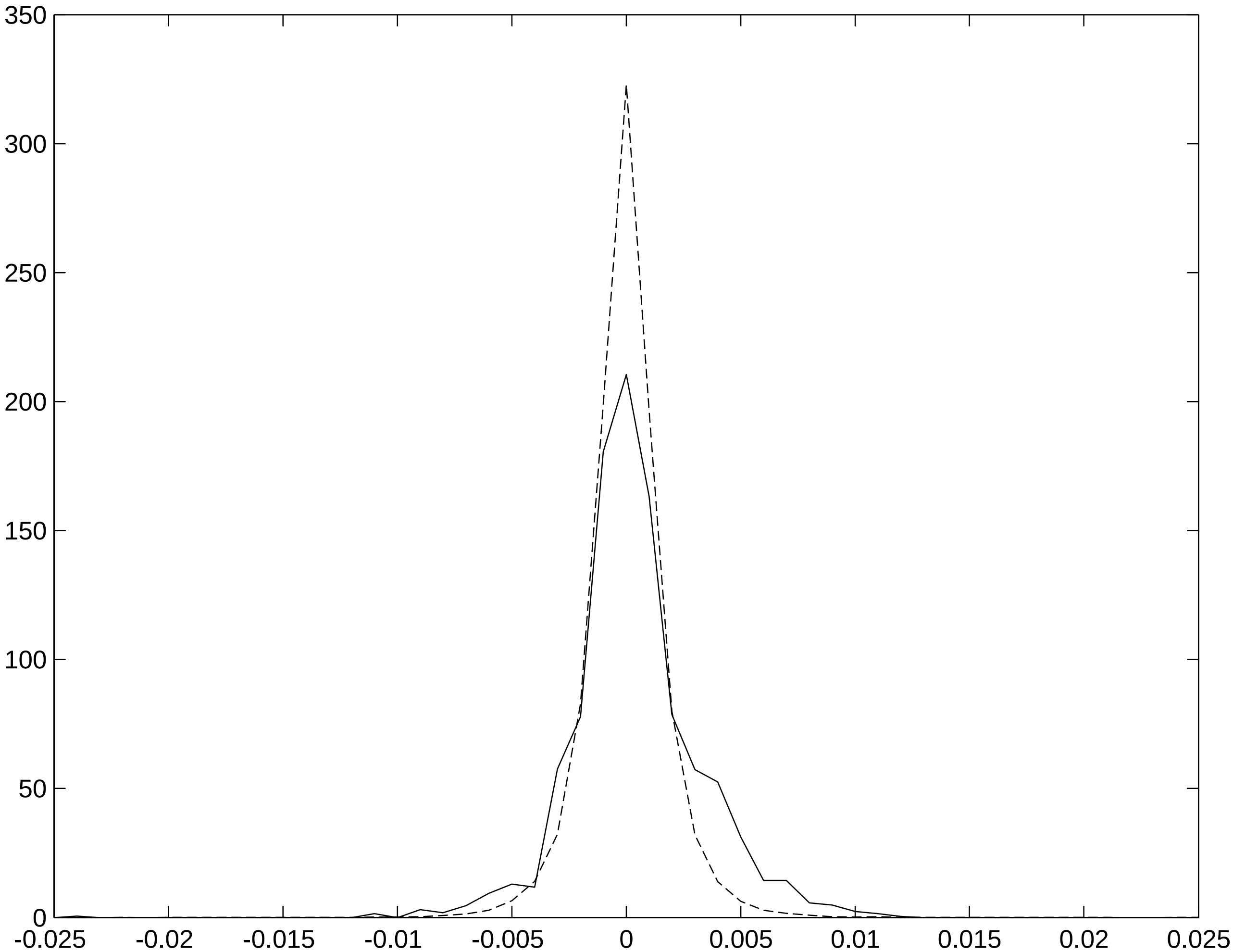
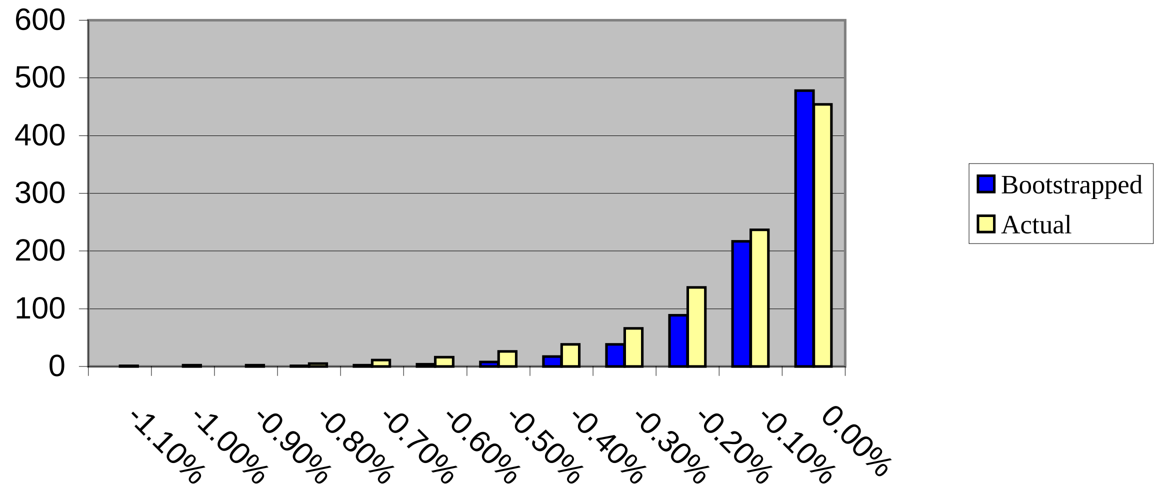


Figure II. Kernel Density Estimate. This figure shows the kernel density estimate of the bootstrapped (dashed line) and actual alpha distribution (solid line) for all investment objectives. This figure is based on the unconditional Fama-French model

Panel A
Number Of Funds From The Original And Bootstrap Sample With Performance Below A Certain Value



Panel B
Number Of Funds From The Original And Bootstrap Sample With Performance Above A Certain Value

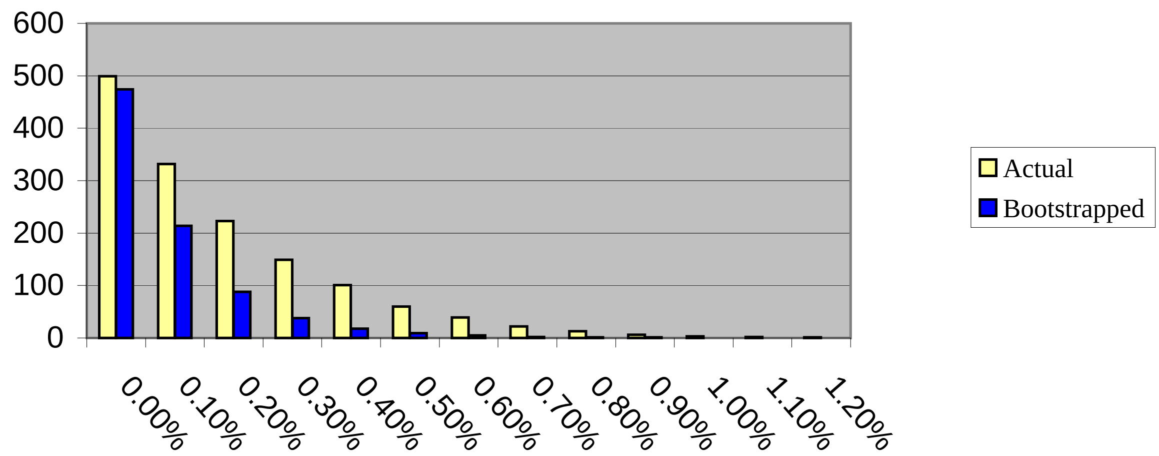


Figure III. This figure presents the number of funds from the original and the bootstrapped distribution against the various values of out- or underperformance that they surpass. The results are based on the number of funds in existence in our sample in 1994 with a minimum of 60 observations. The figure is based on the unconditional Fama-French model.

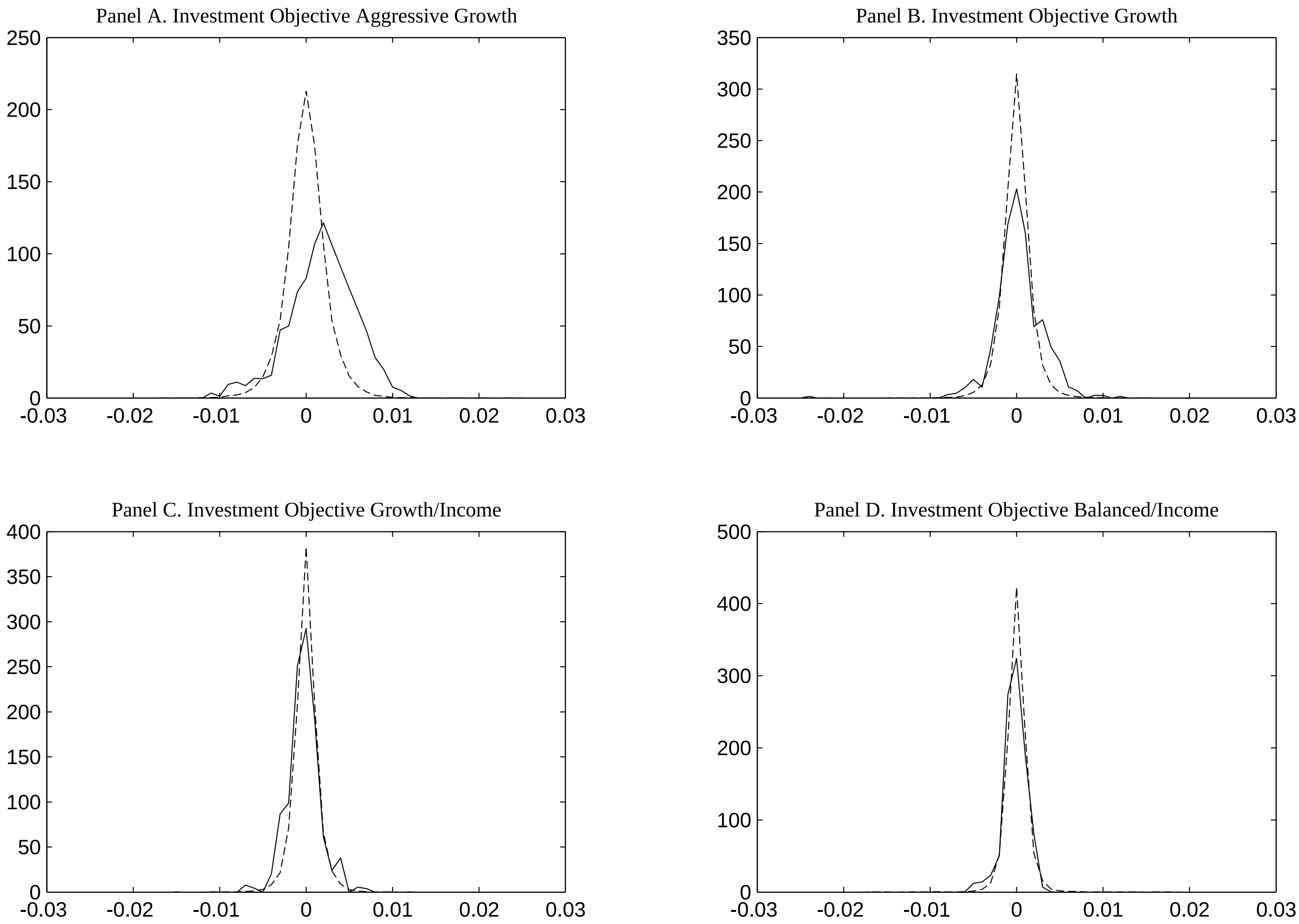


Figure IV. Kernel density estimates for each investment objective. This figure plots the kernel density estimates of the bootstrapped (dashed line) and actual alpha distribution (solid line) for investment objectives aggressive growth (Panel A), growth (Panel B), growth/income (Panel C) and balanced/income (Panel D). The results are based on the unconditional Fama-French model.