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

Robert Kosowski, Allan Timmermann, Russ Wermers, and Hal White*

May 2005

*Kosowski is from INSEAD (Robert.Kosowski@insead.edu, Tel: +33-01-6072-4000), Timmermann and White are from University of California, San Diego (atimmerm@ucsd.edu, Tel: 858-534-4860), and Wermers is from the University of Maryland at College Park (wermers@umd.edu, Tel: 301-405-0572). We thank an anonymous referee for providing us with many insightful comments. Also, we thank Wayne Ferson, Will Goetzmann, Rick Green, Andrew Metrick, and participants at the 2001 CEPR/JFI "Institutional Investors and Financial Markets: New Frontiers" symposium at INSEAD (especially Peter Bossaerts, the discussant), the 2001 European meeting of the Financial Management Association in Paris (especially Matt Morey, the discussant), the 2001 Western Finance Association annual meetings in Tucson, Arizona (especially David Hsieh, the discussant), the 2002 American Finance Association annual meetings in Atlanta, Georgia (especially Kent Daniel, 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-Universität zu Berlin; Imperial College (London); and the University of Pennsylvania.

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

Abstract

We apply a new bootstrap statistical technique to examine the performance of the U.S. open-end, domestic-equity mutual fund industry over the 1975 to 2002 period. Specifically, we bootstrap the joint distribution of performance measures ("alphas") across all funds to determine whether managers of high-alpha funds are simply the luckiest in a large field of managers, or whether they possess genuine stockpicking skills. This bootstrap approach is necessary because the cross-section of mutual fund alphas has a complex, non-normal distribution–due to heterogeneous risk-taking by funds as well as non-normalities in individual fund alpha distributions. Our bootstrap approach reveals findings that differ from many past studies. Specifically, we find that a sizable minority of managers really do pick stocks well enough to more than cover their costs. Moreover, our bootstrap indicates that the superior alphas of these managers *persist*.

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 fund performed well due to his unusual acumen in finding underpriced stocks. In addition, Marcus (1990) asserts that the prolonged superior performance of the Magellan fund is difficult to explain as a purely random outcome, where one assumes that Mr. Lynch and the other Magellan managers have no true stockpicking skills, and are merely the luckiest of a large group of fund managers. More recently, the Schroder Ultra Fund topped the field of 6,000 funds (across all investment objective categories) with a return of 107 percent per year over the three years ending in 2001–and closed to new investors in 1998 due to overwhelming demand, presumably because investors credited the fund manager with having extraordinary skills.

Recent research is supportive of the existence of subgroups of fund managers with superior stockpicking talents, even though most prior studies conclude that the average mutual fund underperforms its benchmarks, net of costs.[1] For example, 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. Furthermore, Wermers (2000) finds that high-turnover mutual funds hold stocks that substantially beat the Standard and Poor's 500 index over the 1975 to 1994 period.

The apparent superior performance of a small group of funds, such as Magellan or the Schroder Ultra Fund, raises the question of whether this is credible evidence of genuine stockpicking skills, or whether it simply reflects the extraordinary luck of a few individual fund managers. Hundreds of new funds are launched every year; by January 2005, over 4,500 equity mutual funds existed in the United States, holding assets valued at almost $4.4 trillion. In 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.[2] Indeed, the literature on performance persistence, to a

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

[2]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 may be effective in correcting for common variation in returns, they generally assume that idiosyncratic variation in returns is uncorrelated and normally distributed across funds.

large extent, is motivated by the need to measure performance during a period separate from the ranking period in an attempt to control for luck, but luck can also persist.

This paper conducts the first comprehensive examination of mutual fund performance ("alpha") that explicitly controls for luck without imposing an ex-ante parametric distribution from which fund returns are assumed to be drawn. In addition, our approach is robust to unknown forms of time-series and cross-sectional heteroskedasticity, as well as autocorrelations and cross-correlations in fund returns. Specifically, we apply several different bootstrap approaches to analyze the significance of the alphas of extreme funds–especially funds with large, positive alphas. In applying the bootstrap to analyze the significance of alpha estimates, we explicitly model and control for the expected idiosyncratic variation in mutual fund returns. As emphasized by Horowitz (2003), the bootstrap has been shown in Monte Carlo experiments to "spectacularly reduce" the difference between true and nominal probabilities of correctly rejecting a given null hypothesis (in our context, that no superior fund managers exist).[3] Furthermore, given the intractability in parametrically modelling the joint distribution of mutual fund performance across hundreds or thousands of funds, most of which are very sparsely overlapping, the bootstrap offers a very attractive alternative approach to analyzing ranked mutual funds.

To illustrate the problem that we address in this paper, suppose that we are told that a particular fund has an alpha of 10 percent per year over a five-year period. This would be, prima facie, an extremely impressive performance record. However, if this fund is, in fact, the best performer among a group of 1,000 funds, then its alpha may not appear to be so impressive. Clearly, when outlier funds are selected based on an *ex-post* ranking of a large cross-section, the separation of luck from skill is difficult. Moreover, any such analysis must account for non-normality in the joint distribution of fund alphas, which can result from (1) heterogenous risk-taking across funds and (2) non-normally distributed individual fund alphas.

Our objective in this paper is simple. In the face of mutual fund alphas that deviate significantly from normality, we address the following question. By random chance, how many funds from a large group will generate high alphas 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. open-end, domestic equity funds during the 1975 to 2002 period–one of the largest panels of fund returns ever analyzed. Across a wide array of performance measurement

[3]Indeed, we find that the bootstrap reduces the probability of a false rejection of the null of no mutual fund outperformance for the top funds, compared to inference based on the parametric t-distribution.

models, our bootstrap tests indicate that, controlling for sampling variability (luck), the large, positive alphas of the top ten percent of funds, net of costs, are extremely unlikely to be a result of sampling variability (luck). We find this result both when we bootstrap the cross-sectional distribution of fund alphas and when we bootstrap the cross-sectional distribution of the t-statistics of fund alphas, as well as with several bootstrap extensions that we employ.[4] Our tests also show strong evidence of mutual funds with negative and significant alphas, controlling for luck.

The results of our paper 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 1,788 mutual funds in our sample that exist for at least five years, our bootstrap indicates that, by luck alone, we would expect nine to achieve an estimated alpha greater than 10 percent per year (net of costs) over (at least) a five-year period. In reality, 29 funds exceed this alpha. As our analysis will show, this is sufficient, statistically, to provide overwhelming evidence that some fund managers have superior talents in picking stocks. Overall, our results provide compelling evidence that, net of all expenses and costs (except load fees and taxes), the superior alphas of star mutual fund managers survive, and are not an artifact of luck. The key to our study is the bootstrap analysis, which allows us to separate luck from skill in the complicated, non-normal cross-sectional distribution of ranked mutual fund alphas.

The above-mentioned bootstrap results are for net-of-cost performance. When we repeat our tests at the pre-cost level, ranking domestic equity mutual funds using the stockholdings-level performance measure of Daniel, Grinblatt, Titman, and Wermers (DGTW; 1997), we find that highly-ranked funds exhibit levels of pre-cost performance that are similar, but slightly higher than their net-of-cost performance; however, low-ranked funds exhibit insignificant pre-cost performance. Thus, much of the aforementioned variation in the cross-section of highly-ranked fund net-of-cost performance is due to differences in skills in choosing stocks, and not simply due to differences in expenses or trade costs. However, variation in low-ranked funds is mainly due to differences in costs. These findings are noteworthy, in that they implicate active-management skills in generating superior fund performance, and not simply superior cost efficiencies. That is, while most funds cannot compensate for their expenses and trade costs, a subgroup of funds exhibits stockpicking skills that more than compensate for such costs.

[4]Specifically, these extensions show that our results are not sensitive to a potential omitted factor from our models, or to possible cross-sectional correlations in idiosyncratic returns among mutual funds (since they may hold similar stocks due, perhaps, to herding behavior).

Further bootstrap results indicate that superior performance is found mainly among growth-oriented funds. This result adds to prior evidence at the stockholdings level that indicates that the average manager of a growth-oriented fund can pick stocks that beat their benchmarks, but the average manager of an income-oriented fund cannot (Chen, Jegadeesh, and Wermers (2000)). Our findings indicate that even seemingly well-performing income fund managers are merely lucky. In addition, we find stronger evidence of superior fund management during the first half of our sample period–after 1990, we must look further to the extreme right tail of the alpha distribution to find superior managers (i.e., managers who are not simply lucky). Thus, the huge growth in new funds over the past decade has apparently been driven by a growth in the number of active managers without talents, who often appear in the right tail by chance alone.[5]

One may wonder whether our results hold much in terms of economic significance, since our main evidence of superior performance occurs in the top ten percent of alpha-ranked funds. For example, if this bootstrap-based evidence of skill occurs primarily in small mutual funds, then our results may not have significant implications for the average investor in actively managed mutual funds. To address this issue, we measure the economic impact by examining the difference in value-added between all funds having a certain level of performance, including lucky and skilled funds, and those funds achieving that level of performance due to luck alone–this difference measures skill-based value-added. We estimate that about $1.2 billion per year in wealth is generated–in excess of benchmark returns, expenses, and trading costs–through active management by funds exceeding a performance level of four percent per year (through skill alone) during the five years preceding the end of 2002. Thus, subgroups of funds possessing skills add significantly to the wealth of fund shareholders. However, a similar computation shows that underperforming mutual funds destroy at least $1.5 billion per year in investor wealth.[6]

Perhaps the strongest motivation for employing the bootstrap is highlighted in our tests of performance persistence, since most investors look at past performance to infer the future. Here,

[5] Our bootstrap results provide useful guidelines for investors. For example, the bootstrap indicates that, among the subgroup of fund managers having an estimated alpha exceeding seven percent per year over a five-year (or longer) period, half have stockpicking talents, while the other half are simply lucky. This information is also useful as inputs to Bayesian models of performance evaluation–for example, Pastor and Stambaugh (2002) and Baks, Metrick, and Wachter (2001) show how prior views about manager skill combine with sample information in the investment decision. Our results provide a reasonable starting point in forming prior beliefs about manager skills for future tests of U.S. domestic-equity mutual fund performance.

[6] Wealth destroyed in a typical year exceeds wealth created by a greater ratio than these figures, as the average out-performing fund in our sample tends to be long-lived, and is generally much larger than the average underperforming fund.

we focus on reconstructing tests of persistence that are similar to those in Carhart (1997), while applying the bootstrap instead of the standard parametric t-tests used by Carhart and others. Notably, our findings overturn some important and widely quoted results from Carhart's paper: specifically, we find significant persistence in net return alphas (using bootstrapped p-values) for the top decile (and, sometimes, top two deciles) of managers, using several different past alpha ranking periods.

So, why does the cross-sectional bootstrap bring substantially different inference regarding high- and low-alpha funds, relative to the p-values of individual fund alphas in these regions? First, our bootstrap formally models the cross-sectional nature of ex-post sorts of funds by building the empirical joint distribution of their alphas. Further, as might be expected, higher moments play an important role in the explanation–indeed, we reject normality for about half of our individual fund alphas. However, and perhaps more surprisingly, we also find that clustering in idiosyncratic risk-taking, across funds, induces important non-normalities in the cross-section of fund alphas. Specifically, we find large variations, in the cross-section, in idiosyncratic risks taken by funds, with clusters of high- and lower-risk funds–perhaps due to the risk-shifting of Chevalier and Ellison (1997). In cases where risk-taking varies substantially, such as among the high-alpha growth funds or the top Carhart-ranked funds mentioned above, the empirical distribution (bootstrap) uncovers thinner tails in the cross-section of fund alphas than that imposed by assuming a parametric normal distribution because of the resulting mixture of individual fund alpha distributions with heterogeneous variances. That is, the presence of clusters of high-risk and lower-risk mutual funds (even when individual fund alphas are normally distributed), can create a cross-sectional distribution of alphas that has thin tails, relative to a normal distribution. In other cases, such as the high-alpha income-oriented funds also noted above, there is less clustered (more uniformly distributed) risk-taking across funds. The resulting cross-section of alphas exhibits thick-tails. Thus, heterogenous risk-taking, among funds, as well as higher moments in individual fund alphas may induce thin- or thick-tailed cross-sectional distributions of alphas, or even tails that are thinner at some percentile points, and thicker at others.

It is important to note that ranking funds by their alpha t-statistic controls for differences in risk-taking across funds. That is, the cross-section of fund t-statistics remains normally distributed in the presence of funds with normally distributed alphas, but with differential levels of idiosyncratic risk. Nevertheless, the higher moments that we find in individual fund alphas (i.e., skewness and kurtosis) result in a cross-section of t-statistics that is non-normal–indeed, we find that the bootstrap remains crucial in inference tests when analyzing the cross-section of t-statistics. Thus, our bootstrap provides improved inference in identifying funds with significant skills, because of (1) differential risk-taking among funds and (2) non-normalities in individual fund alphas.

In summary, when we account for the complex distribution of cross-sectional alphas, we confirm that superior managers with persistent talents do exist among growth funds, and that the bootstrap is crucial to uncovering the significance of these talents. Our results are also interesting in light of the model of Berk and Green (2004), who predict that net return persistence is competed away by fund inflows (since managers likely have decreasing returns-to-scale in their talents). Our evidence of persistence in the top decile of funds indicates that such a reversion to the mean in performance is somewhat slow to occur.

Our paper proceeds as follows. Section I describes our bootstrapping procedure, while Section II describes the mutual fund database used in our study. Section III provides empirical results, the robustness of which is further explored in Section IV. Section V examines performance persistence using the bootstrap, and Section VI concludes.

I Bootstrap Evaluation of Fund Alphas

A. Rationale for the Bootstrap Approach

We apply a bootstrap procedure to evaluate the performance of U.S. open-end, domestic-equity mutual funds.[7] In this setting, there are many reasons why the bootstrap is necessary for proper inference. These include the propensity of individual funds to exhibit non-normally distributed returns, as well as the cross-section of funds representing a complex mixture of these individual fund distributions. We begin by discussing individual funds, then progress to the central focus of this paper: evaluating the cross-sectional distribution of ranked mutual fund alphas, which involves evaluating a complex mixture of individual fund alpha distributions.

A.1 Individual Mutual Fund Alphas

As we will show in a later section of this paper, roughly half of our funds have alphas that are drawn from a distinctly non-normal distribution. There are several reasons for these non-normalities.

[7] The bootstrap is introduced by Efron (1979); details regarding the properties of the bootstrap are available in Efron and Tibshirani (1993) or Hall (1992).

First, individual stocks within the typical mutual fund portfolio have returns with non-negligible higher moments. Although the central limit theorem implies that an equal-weighted portfolio of such non-normally distributed stocks will approach normality, managers often hold heavy positions in relatively few stocks or industries, or they may implement dynamic strategies that involve changing their levels of risk-taking when the risk of the overall market portfolio changes. In addition, market benchmark returns may have non-normalities, and there may be co-skewness in benchmark and individual stock returns, as well as time-series correlations in the idiosyncratic return component of funds. Thus, normality may be a poor approximation in practice, even for a fairly large mutual fund portfolio. The bootstrap can substantially improve on this approximation, as shown by Bickel and Freedman (1984) and Hall (1986). For example, by recognizing the presence of thick tails in individual fund returns, the bootstrap is often found to reject abnormal performance for fewer mutual funds and we confirm this effect in our data.

A.2 The Cross-Section of Mutual Fund Alphas

While the intuition gained from the individual fund bootstrap results above is helpful, such intuition does not necessarily carry over to the cross-section of mutual funds. Specifically, the cross-sectional distribution of alphas also carries the effect of variations in risk-taking (as well as sample sizes) across funds.[8] Furthermore, cross-sectional correlations in residual (fund-specific) risk, although very close to zero on average, may be non-zero in the tails if some funds load on similar non-priced factors. These effects tend to be important because high-risk funds often hold concentrated portfolios that load on similar industries or individual stocks.

That is, fund-level non-normalities in alphas imply non-normalities in the cross-sectional distribution of alphas, but the reverse need not be true. Even funds that each have normally-distributed residuals can create non-normalities in the cross-section of ranked alphas. To illustrate, consider 1,000 mutual funds, each existing over 336 months (the time span of our sample period). Suppose

[8]It is noteworthy that, even if all funds had normally distributed returns with identical levels of risk, it would still be infeasible to apply standard statistical methods to assess the significance of extreme alphas drawn from a large universe of ranked funds. In this case, the best alpha is modeled as the maximum value drawn from a multivariate normal distribution whose dimension depends on the number of funds in existence. Modeling this joint normal distribution depends on estimating the entire covariance matrix across all individual mutual funds which is generally impossible to estimate with precision. Specifically, the difficulty in estimating the covariance matrix across several hundred or thousands of funds is compounded by the entry and exit of funds, which implies that many funds do not have overlapping return records with which to estimate their covariances. Although one might use long-history market indices to improve the covariance matrix estimation, as in Pastor and Stambaugh (2002), this method is not likely to improve the covariance estimation between funds taking extreme positions away from the indices.

each fund has IID standard-normal distributed model residuals, and that the true model intercept (alpha) equals zero for each fund–thus, measured fund alphas are simply the average realized residual over the 336 months. In this simple case, the cross-sectional distribution of fund alphas would be normally distributed.

However, consider the same 1,000 funds with heterogeneous levels of risk, such that, across funds, residual variances range uniformly between 0.5 and 1.5 (i.e., the average variance is unity). In this case, the tails of the cross-sectional distribution of alphas are now fatter than those of a normal distribution. The intuition here is clear: as we move further in the right tail, the probability of these extreme outcomes does not drop very quickly, as higher-risk funds more than compensate for the large drop in such extreme outcomes from low-risk funds. Conversely, consider a case where the distribution of risk levels is less evenly spread out–more clustered–with 1 percent of the funds having a residual standard deviation of 4, and the remaining 99 percent having a standard deviation of 0.92 (the average risk across funds remains at one). Here, tails of the cross-section of alphas are now thinner than that of a normal. The intuition for these unexpected results is quite simple: the presence of many funds with low risk levels means that these funds' realized residuals have a low probability of lying out in the far tails of the cross-sectional distribution of alpha estimates. At some point in the right tail, this probability drops off faster than can be compensated by the presence of a small group of high-risk funds.

Finally, consider a larger proportion of high-risk funds than the prior case–suppose 10 percent of the 1,000 funds have a residual standard deviation of 2, while the remaining 90 percent have a standard deviation of 0.82 (again, the risk averages one across all funds). In this case, the cross-section of alphas has 5- and 3-percentile points that are thinner, but a 1-percentile point that is thicker than that of a normal. Thus, the cross-section of alphas can have thick or thin tails, relative to a normal distribution, regardless of the distribution of individual fund returns, as long as risk-taking is heterogeneous across funds.

In unreported tests, we measure the heterogeneity in risk-taking by all U.S. domestic equity mutual funds existing between 1975 and 2002. We find a heavily skewed distribution of risk-taking among funds–most funds cluster together with similar levels of risk, while a significant minority of funds have much higher levels of risk. In further unreported tests, we bootstrap the cross-section of fund alphas, where each fund is assumed to have residuals drawn from a normal distribution having the same moments as those present in the actual (non-normal) fund residuals (using a four-factor model to estimate residuals).[9] The results show that the cross-section of bootstrapped alphas (assuming individual fund alpha normality) have thinner tails, relative to a normal distribution, except in the extreme regions of the tails, which are thicker. Therefore, the heterogeneity in risk-taking that we observe in our fund sample generates many unusual non-normalities in the cross-section of alphas, even before considering any non-normalities in individual fund alphas.

It is important to note that similar cross-sectional effects will not result when we assess the distribution of the t-statistic of the fund alphas. Since the t-statistic normalizes by standard deviation, heterogeneity in risk-taking across funds, by itself, will not bring about non-normalities in the cross-section.[10] However, non-normalities in individual fund residuals–which, as discussed in a later section, we find for about half of our funds–still imply non-normalities in the cross-section of t-statistics.[11]

Thus, many factors, including cross-sectional differences in sample sizes and risk-taking, as well as fat tails and skews in the individual fund residuals, influence the shape of the distribution of alphas across funds. Given the possible interactions between these effects–and the added complexity arising from parameter estimation errors–it is very difficult, using an ex-ante imposed distribution, to evaluate the significance of the observed alpha of funds, since the quantiles of the standard normal distribution and those of the bootstrap need not be the same in the center, shoulders, tails, and extreme tails. Instead, the bootstrap is required for proper inference involving the cross-sectional distribution of fund performance outcomes.

To summarize, it is only in the very special case where (1) the residuals of fund returns are drawn from a multivariate Gaussian distribution, (2) correlations in fund-specific returns are zero, (3) funds have identical risk levels, and (4) there is no parameter estimation error that we are guaranteed that the standard critical values of the normal distribution are appropriate in the cross-

[9] During each bootstrap iteration, 336 residuals are drawn with replacement from each fund's actual residuals, and the fund alpha for that iteration is measured as the average of these residuals. This generates one cross-sectional alpha outcome, which is repeated 1,000 times.

[10] In fact, this points to the superior properties of the t-statistic, relative to the alpha itself; because of these superior properties, we will use this measure extensively in our bootstrap tests to come in later sections.

[11] For example, suppose that funds have IID thick-tailed residuals that (for each fund) are drawn from a mixture of two normal distributions, the first representing a high volatility state−with a standard deviation of three and a probability of 10%−and a second state with a standard deviation of 1/3 and a probability of 90%–i.e., the average residual variance remains at unity. In this case, the cross-section of fund t-statistics is thin-tailed, compared to a normal distribution. These results, (details available upon request), show that the cross-sectional distribution of performance estimates can have thin tails, even when the individual fund returns are fat-tailed. More complex effects, such as small positive correlations in the tails of the return distribution (across funds) balanced off against small negative correlations in the central part of the return distribution to yield an overall correlation of zero, can also change the cross-sectional tail probabilities.

section. In all other cases, the cross-section will be a complicated mixture of individual fund return distributions, and must be evaluated with the bootstrap.[12]

B. Implementation

In our implementation we consider two test statistics, namely the estimated alpha, $\widehat{\alpha}$, and the estimated t-statistic of $\widehat{\alpha}$, $\widehat{t}_{\widehat{\alpha}}$. $\widehat{\alpha}$ measures the economic size of abnormal performance but suffers from a potential lack of precision in the construction of confidence intervals, whereas $\widehat{t}_{\widehat{\alpha}}$ is a pivotal statistic with better sampling properties.[13] In addition, $\widehat{t}_{\widehat{\alpha}}$ has another very attractive statistical property. Specifically, a fund having a short life or engaging in high risk-taking will have a high-variance estimated alpha distribution, thus, alphas for these funds will tend to be spurious outliers in the cross-section. In addition, these funds tend to be smaller funds that are more likely to be subject to survival bias, raising the concern that the extreme right tail of the cross-section of fund alphas is inflated. The t-statistic provides a correction for these spurious outliers by normalizing the estimated alpha by the estimated variance of the alpha estimate. Furthermore, the cross-sectional distribution of t-statistics has better properties than the cross-section of alphas, in the presence of heterogenous fund volatilities due to differing fund risk levels or lifespans. For these reasons, we propose an alternate bootstrap that is conducted using $\widehat{t}_{\widehat{\alpha}}$, rather than $\widehat{\alpha}$. Indeed, the bulk of our tests in this paper will be applied to the t-statistic.

We apply our bootstrap procedure to monthly mutual fund returns using several models of performance proposed by the past literature. These include the simple one-factor model of Jensen (1968), the three-factor model of Fama and French (1993), the timing models of Treynor and Mazuy (1966) and Merton and Henriksson (1981), and several models that include conditional factors based on the papers of Ferson and Schadt (1996) and Christopherson, Ferson, and Glassman (1998). We will present results for two representative models in this paper, although results for all other models are consistent with those presented, and are available upon request from the authors.[14] The first model–the main model that we present in this paper–is the Carhart (1997) four-factor regression

[12] 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 address the estimation of cross-sectional correlations 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.

[13] A pivotal statistic is one that is not a function of nuisance parameters, such as $Var(\varepsilon_{it})$.

[14] We consider 15 different models. In all cases, results are similar to those presented in the paper. The two representative models that we present are the "best fit" models, according to standard model selection criteria, such as the Schwarz information criterion.

model,

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

where $r_{i,t}$ is the month t excess return on managed portfolio i (net return minus 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 $PR1YR_t$ are 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 second representative model is a conditional version of the four-factor model that controls for time-varying factor loadings and factor return premia, using the technique of Ferson and Schadt (1996). Hence, we extend Equation (1) as follows:

$$r_{i,t} = \alpha_i + b_i \cdot RMRF_t + s_i \cdot SMB_t + h_i \cdot HML_t + p_i \cdot PR1YR_t + \sum_{j=1}^{K} B_{i,j}[z_{j,t-1} \cdot RMRF_t] + \varepsilon_{i,t} \ , \tag{2}$$

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 predictive value of $z_{j,t-1}$ in forecasting the following month's excess market return, $RMRF_t$. This model computes the alpha of a managed portfolio, controlling for strategies that dynamically tilt the portfolio's beta in response to the predictable component of market returns.[15]

We now illustrate the bootstrap implementation with the Carhart (1997) four-factor model of Equation (1). The application of the 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 OLS-estimated alphas, factor loadings, and residuals using the time-series of monthly net returns (minus T-bills)

[15] We use three public information variables that best predict market returns: (1) the lagged level of the one-month Treasury bill yield, (2) the lagged dividend yield of the CRSP value-weighted New York Stock Exchange and American Stock Exchange index, and (3) the lagged yield on a constant-maturity 10-year Treasury bond less the lagged yield on three-month Treasury bills. These variables have been shown to be useful for predicting stock returns and risks over time by other studies as well (see, for example, Pesaran and Timmermann (1995)).

for fund i (r_{it}):

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

For fund i, the coefficient estimates, $\{\widehat{\alpha}_i,\ \widehat{\beta}_i,\ \widehat{s}_i,\ \widehat{h}_i,\ \widehat{p}_i\}$, as well as the time-series of estimated residuals, $\{\widehat{\epsilon}_{i,t}\ ,\ t = T_{0i}, ..., T_{1i}\}$ and the t-statistic of alpha, $\widehat{t}_{\widehat{\alpha}_i}$, are saved, where T_{0i} and T_{1i} are the dates of the first and last monthly returns available for fund i, respectively.

B.1 The Baseline Bootstrap Procedure: Residual Resampling

For our baseline bootstrap, we draw, for each fund, i, a sample with replacement from the fund residuals that are saved in the first step above, creating a pseudo time-series of resampled residuals, $\{\widehat{\epsilon}^b_{i,t},\ t = s^b_{T_{0i}}, ..., s^b_{T_{1i}}\}$, where $b = 1$ (for bootstrap resample number one), and where $s^b_{T_{0i}}, ..., s^b_{T_{1i}}$ is the time reordering resulting from resampling the same number of residuals as in the original sample for fund i.

Next, a time-series of pseudo monthly excess returns is constructed for this fund, imposing the null hypothesis of zero true performance ($\alpha_i = 0$, or, equivalently, $\widehat{t}_{\widehat{\alpha}_i} = 0$):

$$\{r^b_{i,t} = \widehat{\beta}_i RMRF_t + \widehat{s}_i SMB_t + \widehat{h}_i HML_t + \widehat{p}_i PR1YR_t + \widehat{\epsilon}^b_{i,t}\ ,\ \ t = s^b_{T_{0i}}, ..., s^b_{T_{1i}}\}. \tag{3}$$

Note that factor returns are drawn from the same time-period as resampled residual returns–we relax this in a later version of our bootstrap. As indicated by Equation (3), this sequence of artificial returns has a true alpha (and t-statistic of alpha) that is zero by construction. However, when we next regress the returns for a given bootstrap sample, b, on the Carhart factors, a positive estimated alpha (and t-statistic) may result, since that bootstrap may have drawn an abnormally high number of positive residuals, or, conversely, a negative alpha (and t-statistic) may result if an abnormally high number of negative residuals are drawn.

Repeating the above steps across all funds, $i = 1, ..., N$, we arrive at a draw from the cross-section of bootstrapped alphas. Repeating for all bootstrap iterations, $b = 1, ..., B$, we then build the distribution of these cross-sectional draws of alphas, $\widehat{\alpha}^b_i$, or their t-statistics, $\widehat{t}^b_{\widehat{\alpha}_i}$, resulting purely from sampling variation, while imposing the null of no true performance. For example, the distribution of alphas for the top fund is constructed as the distribution of the maximum alpha generated across all bootstraps.[16] Bootstrapping the distribution of the $\widehat{t}_{\widehat{\alpha}}$ proceeds similarly. As noted in Section I.A, this cross-sectional distribution can be non-normal, even if individual fund alphas are normally distributed. If we find that our bootstrap iterations generate far fewer extreme positive values of $\widehat{\alpha}$ or $\widehat{t}_{\widehat{\alpha}}$, compared to those observed in the actual data, then we conclude that sampling variation (luck) is not the sole source of high alphas, but that genuine stockpicking skills actually exist. In all of our bootstrap tests, we execute 1,000 bootstrap iterations ($B = 1,000$).

B.2 Bootstrapping Extensions

We will implement some other straightforward extensions of this bootstrap for our universe of funds as well. These extensions, which are described in more detail in Section IV, include residual and factor resampling, as well as a procedure that demonstrates that our results are robust to the presence of a potential omitted factor in our models. We also allow for the possibility of cross-sectional dependence between fund residuals that may, for example, be due to funds holding similar (or the same) stocks at the same time. That is, funds with very high measured alphas might have similar holdings. In addition, we implement a procedure that allows for the possibility that the residuals are correlated over time for a given fund, perhaps due to time-series patterns in stock returns that are not properly specified by our performance models.

II Data

We examine monthly returns from the Center for Research in Security Prices (CRSP) mutual fund files. The CRSP database contains monthly data on net returns for each shareclass of every open-end mutual fund existing after January 1, 1962, 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 some investment objective information is available from the CRSP database, we supplement these data with investment objective and other fund information from the CDA-Spectrum mutual fund files obtained from Thomson Financial, Inc., of Rockville, Maryland.[17] We use Thomson data since these investment objectives are more complete and consistent than the correspond-

[16] Of course, this maximum alpha can potentially be associated with a different fund during each bootstrap iteration, depending on the outcome of the draw from each fund's residuals.

[17] The Thomson database, and the technique for matching it with the CRSP database are described in Wermers (1999, 2000). These links are available from Wharton Research Data Services (WRDS).

ing information from CRSP.[18] For each open-end, U.S. domestic-equity fund, we compute monthly fund-level net returns by weighting shareclass-level returns by the proportion of fund total net assets represented by each shareclass at the beginning of each month. All shareclasses that exist at the beginning of a given month are included in this computation for that month; specifically, no-load, load, and institutional classes. Thus, our computed fund-level returns represent the experience of the average dollar invested in that fund (across all shareclasses), although most shareclass-level returns are not substantially different (ignoring loads) from fund-level returns. 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.[19]

Our final database contains fund-level monthly net returns data on 2,118 U.S. open-end, domestic equity funds that exist for at least a portion of the period from January 31, 1975 to December 31, 2002.[20] We study the performance of the full sample of funds, as well as funds in each investment-objective category. Namely, our sample consists of aggressive-growth funds (285), growth funds (1,227), growth-and-income funds (396), and balanced or income funds (210).[21] Since balanced funds and income funds allocate a significant fraction of assets to non-equity investments, we require that such funds hold at least 50 percent domestic equities during the majority of their existence to be included in our tests.

Table I shows counts of funds and their average returns during five-year subintervals. Panel A presents counts for the entire fund dataset, while Panels B through E present counts, segregated by investment objective. For example, Panel A shows that 322 funds (both surviving and non-

[18]CRSP investment objective data is often missing for at least some years (and sometimes all years) of the funds' 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 Thomson files report investment objectives in a more consistent manner across funds and over time. In any case, all investment objective information (CRSP and Thomson) is considered, as well as examining the name of the fund, when we classify it into a domestic-equity category. Further, we use the first available investment objective for that fund to classify a fund throughout its life–only 16 percent of funds change objectives.

[19]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.

[20]In an earlier version of this paper, we analyzed fund returns starting on January 1, 1962–the beginning of the CRSP dataset. However, we were forced to only consider funds that exist on January 1, 1975 (the beginning of the Thomson investment objective data) due to the investment objective data problem in CRSP–this backfilling induced survival bias. However, all major results were found to be consistent with those reported here.

[21]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 each category are available in Grinblatt, Titman, and Wermers (1995).

surviving) exist during 1971 to 1975; this figure grows to 1,824 during 1998 to 2002. This rapid growth in numbers is mainly driven by a large expansion in the number of growth funds (Panel C), although other types also show substantial increases. In order to check whether our fund universe is representative of all U.S. open-end, domestic equity funds that exist at each point in time, we compare our counts with those obtained from the Investment Company Institute (2004). In general, we find that, allowing for a lag in the inclusion of new funds by Thomson (as noted in Wermers (2000)), our counts track the counts of the ICI quite closely.[22]

We include only funds having a minimum of 60 monthly net return observations in our baseline bootstrap tests, although we relax this in later tests.[23] Panel A of Table I compares, for each five-year subperiod, counts of all existing funds with counts of funds having all 60 monthly returns. In addition, the panel shows average monthly excess returns and four-factor model alphas (both annualized to percent per year) for equal-weighted portfolios of funds in these two groups. Although our count of funds is substantially lower when we require 60 monthly returns, excess returns and four-factor alphas are only slightly higher than those for the full sample. Specifically, excess returns and alphas of funds surviving for the full 60 months are roughly 20 basis points per year higher, during each subperiod, than those for all existing funds (see Panel A). Slightly higher (but still small) differences exist for growth-oriented funds, which generally undertake riskier strategies (see Panels B and C). Overall, our results show that short-lived funds do not have substantially different average returns than longer-lived funds (consistent with the evidence of Carhart, Carpenter, Lynch, and Musto (2002)). Nevertheless, as a robustness check, we apply the bootstrap (using some of our simpler models) with a minimum history requirement of 18, 30, 90, and 120 months in extensions of our bootstrap. The results (which will be presented in a later section) generally show that survival

[22]Specifically, our dataset has 312, 306, 401, 745, 1,175, 1,704, and 1,558 funds at the end of 1975, 1980, 1985, 1990, 1995, 2000, and 2002, respectively (note that these counts differ from those shown in the first column of Panel A, since Panel A includes funds that did not survive until the end of the period shown). Our count drops after 1999, because we did not add funds that started up after that year (since these funds would not have sufficient returns by 2002 for our baseline regression tests). In order to check our counts of funds against the ICI totals, we first adjust our counts to exclude balanced funds and income funds to make our counts comparable to those provided in Table 6 of ICI (2004). For example, we subtract the 54 balanced and income funds from our total of 401 funds to arrive at 347 funds at the beginning of 1985; this compares with 430 funds (in the categories of "capital appreciation" and "total return") counted by the ICI at the same time. The lower number of funds in our database, relative to the ICI, is due to our counts lagging those of the ICI by one to three years. For example, our January 1995 count of 1,045 (excluding balanced/income) roughly matches the ICI count of 1,086 at January 1993; our January 1990 count of 651 roughly matches the ICI count of 621 at the beginning of 1987.

[23]This minimum data requirement is necessary to generate more precise regression parameter estimates for our more complex models of performance. 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) after the last non-missing return observation (and, thus, cannot be used in our regressions).

bias has almost no impact on our bootstrap results.

III Empirical Results

A. The Normality of Alphas

Before progressing to our bootstrap tests, we analyze (in unreported tests) the distribution of individual fund alphas generated by our models of Section I.C, as well as alphas generated by many other commonly used performance models. We find that normality is rejected for 48 percent of funds when using either the unconditional or conditional four-factor model; similar results are found with all other models that we test. Moreover, we also find that the rejections tend to be very large for many of the funds, especially funds with extreme estimated alphas (either positive or negative). This strong finding of non-normal alpha estimates challenges the validity of earlier research that relies on the normality assumption–this challenge to standard t- and F-tests of the significance of fund alphas strongly indicates the need to bootstrap, especially in the tails, to determine whether significant alphas are due to manager skills (or lack thereof). As we apply our bootstrap in the following sections, we will highlight the significant changes in inference that result.

B. Bootstrap Analysis of the Significance of Alpha Outliers

We first apply our residual resampling method, described in Section I.B.1, to analyze the significance of mutual fund alphas. In these tests, we rank all mutual funds, having at least 60 months of return observations during the 1975 to 2002 period, on their model alphas. It is important to note that, in all of our bootstrap results to come, we compare p-values generated from our cross-sectional bootstrap, for each ranked fund, with standard p-values corresponding to the t-statistics of these individual ranked funds–these individual fund t-tests, of course, do not consider the joint nature of the ex-post sorting that we implement. As discussed in Section I.A., the cross-sectional nature of our bootstrap, along with its ability to model non-normalities in fund alphas, provide benefits over the casual use of standard t-tests applied to individual funds. Since investors and researchers usually examine funds without considering the joint nature of ex-post sorts, we use this approach to inference as a benchmark against which to compare the bootstrap. As we will see, the bootstrap, in many cases, provides substantially different conclusions about the significance of individual ranked-fund performance.

B.1 Baseline Bootstrap Tests: Residual Resampling

Panel A of Table II shows several points in the resulting cross-section of alphas, using the unconditional and conditional four-factor models, and presents bootstrapped p-values ("Cross-sectionally bootstrapped p-value"), as well as standard p-values corresponding to the t-statistic of the individual fund at each percentile point of the distribution ("Parametric (standard) p-value"). For example, consistent with the results of Carhart (1997), the median fund in our sample has an unconditional four-factor alpha of -0.1 percent per month (-1.2 percent, annualized), while the bottom and top funds have alpha estimates of -3.6 and 4.2 percent per month, respectively.[24] Also, as further examples, the fifth-ranked fund and the fund at the one-percentile point in our sample have alphas of 1.3 and one percent per month, respectively.[25] Ranking funds by their conditional four-factor alphas results in alphas, and p-values (both cross-sectionally bootstrapped and parametric normal) that are remarkably similar to those from the unconditional four-factor alpha sort. This finding indicates that mutual funds do not substantially time the overall market factor through the use of macroeconomic variables. Therefore, for the remainder of this paper, we will present results only for the unconditional four-factor model; however, in all cases, the conditional four-factor model exhibits similar results, which are available upon request from the authors.

Overall, our results of Panel A show that funds with alphas ranked in the top decile (10^{th}-percentile and above) generally exhibit significant bootstrapped p-values, whether we use the unconditional or conditional four-factor model. However, this is not always the case. For example, the second-ranked fund under the unconditional model displays a large but insignificant alpha; this alpha simply is insufficiently large to reject (based on the empirical distribution of alphas) that the manager achieved it through luck alone. Thus, our bootstrap highlights that extreme alphas are not always significant, and that the bootstrap is important in testing for significance in the tails, which can have quite complex distributional properties.

No funds between (and including) the 20^{th}-percentile and the median exhibit alphas sufficient to beat their benchmarks, net of costs–using either the unconditional or conditional versions of

[24] This top-ranked fund is the Schroeder Ultra Fund, which was prominently featured in the media as a fund with extraordinary performance. Although the fund eventually closed to new investments, there were many cases of investors wishing to purchase some shares from current shareholders at exorbitant prices in order to be allowed to add further to those holdings, which is allowed by this fund.

[25] As an example, the cross-sectionally bootstrapped p-value of 0.02 is the probability that the fifth-ranked fund (from repeated *ex-post* alpha sorts) generates an alpha of at least 1.3 percent per month, purely by sampling variation (i.e., with a true alpha of zero). In contrast, the parametric (standard) p-value of 0.01 for this fund is obtained from a simple time-series t-test for this alpha, without regard to the rank of this fund in the cross-section.

the four-factor model. When we examine funds below the median, using a null hypothesis that these funds do not underperform their benchmarks (net of costs), we find that all bootstrapped p-values strongly reject this null. This finding of significantly negative alphas for below-median funds indicates that these funds may very well be inferior to low-cost index funds. In unreported results available from the authors upon request, we arrive at the same conclusions with all other performance models, including complex models with both market timing measures and multiple risk factors.

As discussed in Section I.A, we also rank with a second measure of fund performance–the t-statistic for the estimated alpha.[26] As mentioned, the t-statistic has some advantageous statistical properties when constructing bootstrapped cross-sectional distributions, since it scales alpha by its standard error (which tends to be larger for shorter-lived funds and for funds taking higher levels of risk). In addition, it is related to the Treynor and Black (1973) appraisal ratio, which is prescribed by Brown, Goetzmann, Ibbotson, and Ross (1992) for helping to mitigate survival bias problems in datasets. Thus, the distribution of bootstrapped t-statistics in the tails is likely to have better properties (fewer problems with high variance or survival bias) than the distribution of bootstrapped alpha estimates in that region.

Panel B presents results for funds ranked by their t-statistics. In general, right-tail funds continue to exhibit significant performance under a t-statistic ranking, as they did in the alpha ranking of Panel A. Most importantly, note that our inference about fund manager talent is somewhat different with the cross-sectional bootstrap than with the standard parametric normal assumption applied to individual fund alpha distributions (shown as "parametric (standard) p-value" in Panel B).[27] Namely, most of the top five funds have bootstrapped p-values that are higher than the parametric p-values–for both unconditional and conditional model alpha t-statistics. In this extreme right tail of the cross-section, the bootstrap uncovers more probability mass (fatter extreme right tail) than expected under a parametric normal assumption–as a result of the complex interaction between non-normal individual fund alphas as well as the complexity of the mixture of these distributions imposed by the cross-sectional draws. The same applies to funds closer to the median. For

[26] Reported t-statistics use Newey-West (1987) heteroskedasticity and autocorrelation consistent standard errors.

[27] Again, in addition to their normality assumption, one should note that these individual fund parametric p-values do not account for the ex-post, cross-sectional nature of our ranking of funds. For instance, one would not conclude that the three-percentile fund shown in Panel A has a signficant alpha simply because its p-value is three percent–this would be expected in the absence of any true alpha under an assumption of IID multivariate normal fund returns. However, since ex-post individual fund p-values are often used to infer talent (rather than a full cross-sectional test statistic across funds), we compare inference under the cross-sectional bootstrap with this often-used approach.

example, the standard parametric p-value for the t-statistic (the one-tailed p-value for t=1.36 is roughly nine percent) indicates that the fund at the 10^{th}-percentile exhibits a significant t-statistic, under the unconditional four-factor model. However, the bootstrap does not find this t-statistic to be significant, and does not reject the null of no manager talent at the 10^{th}-percentile (this p-value equals 25 percent).

To explore further, Figure I presents distributions of unconditional four-factor alphas for funds at various points in the cross-section. For example, Panel A1 shows the bootstrapped distribution of the alpha of the bottom-ranked fund, across all bootstrap iterations. While the mode of this distribution lies at roughly -1.7 percent per month, bootstrapped alphas vary from about -1 percent per month to (in rare cases) less than -6 percent per month. It is easy to see that the actual bottom-fund alpha of -3.6 percent per month (shown as the dashed line in Panel A1) lies well within the left-tail rejection region of the distribution; this rejection is so strong, that a standard t-test also rejects. However, Panel B4 shows a case where the bootstrap rejects the null, while the simple t-test does not. In general, as we proceed to the center of the cross-sectional distribution, (Panels A1 to A4 and Panels B1 to B4), alpha distributions become more symmetric, but remain markedly non-normal.

As noted earlier, the extreme deviation from normality observed in the extreme top and bottom funds, ranked by alpha (e.g., Panels A1 and B1), is due to those positions generally being occupied by funds having very risky strategies–which motivates our t-statistic ranking procedure shown in Panel B of Table II. However, we find that bootstrapped t-statistics for funds at various points in the cross-sectional distribution also deviate substantially from normality–this is illustrated by Figure II.

Panel A of Figure II compares the cross-sectional distribution of actual fund t-statistic estimates with the distribution generated by the bootstrap.[28] The two densities in Panel A have quite different shapes–the distribution of actual t-statistics has more probability mass in the left and right tails, and far less mass in the center than the bootstrapped distribution. However, this is not the whole story–the distribution of actual t-statistics also exhibits several complex features, such as "shoulders" in the tail regions. Thus, it is not simply that the bootstrap more adequately measures fat or thin

[28]This panel shows the bootstrapped cross-sectional distribution under the unconditional Carhart model. The distributions are smoothed with a kernel density estimator–this estimator replaces the "boxes" in a histogram by "bumps" that are smooth–the kernel function is a weighting function that determines the shape of the bumps. The plot was generated using a Gaussian kernel function. The optimal bandwidth controls the smoothness of the density estimate, and is calculated according to Silverman (1986).

tails of the actual distribution that makes our bootstrap inference different from inference based on a normality assumption, but also that the bootstrap more adequately captures the complex shape of the entire cross-sectional distribution of t-statistics (and, especially that of the tails) under the null. The 95 percent standard error bands around the bootstrapped distribution confirm that the differences between the two distributions are statistically significant.

Overall, this figure illustrates that our sample of funds exhibits actual t-statistics with a very non-normal cross-sectional distribution, and that the tails of this actual distribution are not well-explained by random sampling error (which is represented by the bootstrapped distribution). These observations add to our prior evidence that many superior and inferior funds exist in our sample. Since our interest is the actual number of funds exceeding a certain level of alpha, compared to the bootstrapped distribution, we plot the cumulative density function in Panel B. The results confirm our observations from Panel A–in the far right tail, the actual probability distribution has more weight than the bootstrapped distribution. In addition, as Panel B of Table II indicates, t-statistics above 1.96 are generally significant; this results in the actual cumulative density function lying below the bootstrapped cumulative density function in that region.

We can also use the bootstrapped distribution of alphas to calculate how many funds (out of the total set of funds with a track record of at least five years) would be expected, by chance alone, to exceed a given level of performance–this number can be compared to the number of funds that actually exceed this level of performance in our sample. Panel A of Figure III plots the cumulative number of funds from the original and (imputed) from the bootstrapped distribution that perform above each level of alpha, while Panel B plots the cumulative numbers that perform below each level. For example, Panel A indicates that nine funds should have an alpha estimate higher than 10 percent per year by chance–in reality, 29 funds achieve this alpha. Further, Panel B indicates that 128 funds exhibit an alpha estimate less than -6 percent per year, compared to an expected number of 63 funds by random chance.

Overall, the results in this section provide strong evidence that many of the extreme funds in our sample exhibit significant positive (or negative) alphas and alpha t-statistics. For example, Panel A of Figure III indicates that, among the subgroup of fund managers having an alpha exceeding seven percent per year over a five-year (or longer) period, about half have stockpicking talents sufficient to exceed their costs, while the other half are simply lucky.

To evaluate the overall potential economic impact of our findings, we approximate the value-added of skilled managers. This is important as, for example, our bootstrap-based evidence of skill might occur primarily in small mutual funds, which might tend to lie further in the tails of the alpha distribution. If so, then our results may not have significant implications for the average investor in actively managed mutual funds. To address this issue, we measure the economic impact by examining the difference in value-added between all funds having a certain level of performance, including lucky and skilled funds, and those funds achieving that level of performance due to luck alone (estimated by the bootstrap)–this difference estimates skill-based value-added.[29] Panels A and B of Figure III show the cumulative value-added (value-destroyed) above (below) each point in the alpha distribution. We estimate that about $1.2 billion per year in wealth is generated–in excess of expenses, and trading costs–through true active management skills by funds in the right tail of the cross-section of alphas over the 1975 to 2002 period. By contrast, truly underperforming left-tail funds destroy a total of $1.5 billion per year by their inability to compensate for fees and trading costs. It should be noted that wealth destroyed in a typical year exceeds wealth created by a greater ratio than these figures, as the average outperforming fund in our sample tends to be long-lived, and is generally much larger than the average underperforming fund.

One should note, at this point, that many issues could complicate the interpretation of our baseline bootstrap results presented above. For example, funds may have cross-sectionally correlated residuals. If true, this could bias our bootstrap results, which (so far) have rested on the assumption of independent residuals. We will explore these and other concerns in Section IV of this paper.

B.2 Baseline Bootstrap Tests for Subperiods

To examine whether the cross-sectional distribution of mutual fund performance has changed over our sample period, we examine two subperiods of roughly equal lengths, namely, 1975-1989 and 1990-2002. The results are reported in Table III. We find that outperforming fund managers have become scarcer after 1990, according to our bootstrap. Either markets have become more efficient, or competition among the large number of new funds has reduced the gains from trading (or perhaps

[29]For example, we found that there are 29 funds with alpha estimates greater than or equal to 10 percent per year, while there are only nine funds expected to achieve those alphas by chance. To approximate the value-added of the additional 20 funds (since we do not know which 20 funds out of the 29 are skilled), we compute value-added as the alpha of funds in each one percent per year interval in the right tail above (and including) the 10 percent per year point, multiplied by the average size for funds lying in that alpha interval. We repeat this computation, using an average of the largest funds in each alpha interval.

these two are related). Nevertheless, we do find substantial performance in the top five percent of funds, post-1990. Note, also, that inference based on the bootstrap differs from that of the standard parametric normal at many more points in the cross-sectional distribution for subperiods, relative to the whole sample of Table II. In fact, as we will see in the next section, the bootstrap becomes more important for correct inference when we move to smaller numbers of funds, or to funds having shorter lives.

B.3 Baseline Bootstrap Tests 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 fund buys stocks with abnormal returns that are two percent per year higher than stocks the fund sells. By contrast, the average income-oriented fund does not exhibit any stockpicking talents. However, it is not clear whether stockpicking talent translates into superior net return performance. Accordingly, we next divide our sample of funds by investment objective to see whether the tails of the alpha and t-statistic distributions are affected by the investment style of a fund.

Table IV reports bootstrap results for each investment-objective subgroup, each ranked with the unconditional four-factor alpha, as well as the corresponding t-statistic of this alpha. We focus on discussing the t-statistic rankings, although the alpha rankings show similar results. For example, Panel A of Table IV shows the results of our t-statistic ranking, and associated bootstrapped p-values, applied only on those funds having an investment objective of "growth." For this subgroup of funds, our bootstrapped p-values indicate that all funds above (and including) the five-percentile point have managers with stockpicking skills that outweigh their expenses and trading costs. However, the results also lead us to conclude that the inferior performance of funds in the left tail is significant; all funds below (and including) the 20^{th}-percentile in the left tail have managers who cannot recover costs through superior stockpicking talents. Note that inference based on a standard t-test follow our cross-sectional bootstrap-based inferences fairly closely–with the exception of the 10-percentile point–for this subgroup of funds. Thus, other than the important result that the cross-sectional bootstrap finds outperformance in only half of the right tail, compared to standard individual fund t-tests, differences between the two different test results are not large.

Panel B reports results for aggressive-growth funds (see the t-statistic ranking). Here, right tail bootstrapped p-values, above (and including) the 5-percentile fund, are statistically significant, with the exception of the best and second-best fund. Note that, using standard t-tests, we would conclude that these two funds, and the 10-percentile fund, exhibit significant performance. Again, as with growth funds, standard t-tests would lead us to believe that superior performance is about twice as prevalent among funds as it actually proves to be. Even more compelling evidence on the importance of the bootstrap is evident in Panels C and D, which examine funds having an investment objective of growth and income, and funds having an investment objective of either balanced or income, respectively. While the large bootstrapped p-values lead us to conclude that high-alpha funds in either of these groups are simply lucky, standard t-tests would lead us to conclude otherwise for funds above (and including) the 10-percentile fund.

Thus, among smaller samples of funds, our bootstrap departs most strongly from the standard normality assumption when we examine the riskiest funds (extreme aggressive-growth funds) as well as fund groups where the alpha of top funds is not very large (income-oriented funds). Figure IV further explores the shape of the bootstrapped distributions of t-statistics in the right tail of each ranked investment-objective category. These figures clearly show the complex, non-normal shapes of these distributions, which explains why we arrive at different conclusions about manager talent with the bootstrap. Note that, with more extreme actual measured t-statistics (i.e., Panels A1 through D1), bootstrapped distributions become much more non-normal, exhibiting substantial "shoulders" and skewness. In addition, note that the outperformance for growth-oriented funds is much more dramatic than for income-oriented funds, making the bootstrap less crucial to arrive at appropriate inferences for growth funds.

For the sake of brevity, we do not include results for the conditional four-factor model, segregated by investment-objective categories, but these results are very similar to the unconditional model results shown above (and are available from the authors upon request). On the whole, we find that inference using the bootstrap is substantially different from that of the normality assumption of past studies of performance. Specifically, in some very important cases, we arrive at exactly the opposite conclusion about manager talent–especially in examining the right tails of funds with somewhat concentrated strategies (aggressive-growth funds) or funds with less dramatic levels of performance (income-oriented funds).

IV Sensitivity Analysis

In the last section, we conducted an extensive set of resampling tests to determine the significance of alpha and t-statistic outliers. In this section, we test whether our results are sensitive to changes in the nature of the bootstrap procedure, to the assumed return-generating process, or to the set of mutual funds included in the bootstrap tests. In general, we will show that our main findings in this paper are robust to changes in these parameters.

A. Time-Series Dependence

Our bootstrap 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 that apply if the IID assumption is violated (see, for example, Hall (1992)).

Nevertheless, as a robustness check, we explicitly allow for dependence in return residuals over time by adopting the stationary bootstrap suggested by Politis and Romano (1994) which resamples data blocks (returns) of random length. Specifically, the Politis and Romano approach draws a sequence of IID variables from a geometric distribution to determine the length of the blocks, and draws a sequence from a uniform distribution to arrange the blocks to yield a stationary pseudo-time series.

To explore the sensitivity of our results, we compare the bootstrap results under a block length of one monthly return (which is the same as our previous independent resampling) and for larger block lengths–up to a maximum block length of 10 monthly returns. In unreported results, we find that, with all block lengths greater than one, estimated alphas, t-statistics, and bootstrapped p-values are almost identical to the results from our baseline block length of one (in Section III).

B. Residual and Factor Resampling

We next implement a bootstrap with independent resampling of regression residuals and factor returns. This approach breaks any correlation between these two components. Such a correlation would occur, for example, if a fund manager holds stocks having a return co-skewness with the market, or with other factor returns. Co-skewness may also occur if the manager has market- or factor-timing abilities that are not properly specified in the performance model.

When resampling factor returns, the same draw is used across all funds (to preserve the correlation effect of factor returns on all funds), giving the following data for bootstrap iteration b for fund i. Thus, the independently resampled factors and residuals are represented by

$$\{RMRF_t^b, SMB_t^b, HML_t^b, PR1YR_t^b,\ t = \tau_{T_{0i}}^b, ..., \tau_{T_{1i}}^b\} \text{ and } \{\hat{\epsilon}_{i,t}^b,\ t = s_{T_{0i}}^b, ..., s_{T_{1i}}^b\}.$$

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 ($\alpha_i = 0$):

$$\{r_{i,t}^b = \widehat{\beta}_i RMRF_{t_F}^b + \widehat{s}_i SMB_{t_F}^b + \widehat{h}_i HML_{t_F}^b + \widehat{p}_i PR1YR_{t_F}^b + \widehat{\epsilon}_{i,t_\epsilon}^b\},$$

for $t_F = \tau_{T_{0i}}^b, ..., \tau_{T_{1i}}^b$ and $t_\varepsilon = s_{T_{0i}}^b, ..., s_{T_{1i}}^b$, the independent time reorderings imposed by resampling the factor returns and residuals, respectively, in bootstrap iteration b. Again, in unreported results, we find that this approach exhibits almost identical results (for both left-tail and right-tail funds) to those of our residual-only approach of Section III.

C. *Cross-Sectional Bootstrap*

In empirical tests, we find the cross-sectional correlation between fund residuals to be very low–the average residual correlation is 0.09 for the four-factor Carhart model.[30] Nevertheless, we refine our bootstrap procedure to capture any potential cross-sectional correlation in residuals by implementing an extension that draws residuals, across all funds, during identical time periods. For example, funds may herd into, or otherwise hold the same stocks at the same time, inducing correlations in their residuals. This herding may be especially important in the tails of the cross-section of alphas.

In this procedure, 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. Since some funds may, as a result, be allocated bootstrap index entries from periods when they did not exist, or otherwise have a return observation, we drop a fund if it does not have at least 60 observations during the reindexed time sequence. Again, unreported tests show that these results are almost identical to our baseline results of Section III.

[30]Since our data is an unbalanced panel of funds, we calculate the average correlation by matching each fund with the time series of funds that existed during the same time. We then calculate the average of all these pairwise correlations.

D. Portfolios of Funds

In order to determine whether our analysis of individual fund alphas in the cross-section, or t-statistics of individual fund alphas, is substantially affecting our inference tests, we next consider the corresponding average statistics for portfolios of funds in each tail of the alpha (or t-statistic) distribution. This cross-sectional smoothing enables a robustness check of our results. For example, in the left tail of the alpha distribution, our first portfolio consists of all funds having an alpha (or, alternatively, t-statistic) that lies in the lowest one-percentile of the alpha (or, t-statistic) distribution of all funds. Our second portfolio consists of all funds having an alpha that lies in the lowest two-percentile of the alpha (or t-statistic) distribution, etc. Analogous portfolios are formed for the right-tail of the alpha (or, t-statistic), distributions. We calculate the average alpha, and average t-statistic, for funds in each of these portfolios, as well as the associated bootstrapped p-value for each portfolio. For these portfolio tests, the p-value tells us the probability of observing the average t-statistic that we actually do observe, under the imposed assumption of a zero true t-statistic. This test is similar to bootstrapping an F-test to determine whether we can jointly reject that all funds in a certain portion of the tail do not have a performance measure that deviates from zero. Again, in unreported tests, we find strong evidence that both the left-tail and right-tail alpha t-statistics observed for the portfolios of funds cannot be attributed simply to luck.

E. 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 in the cross-section of alpha estimates. In an attempt to correct for this effect, the baseline bootstrap results imposed a minimum of 60 observations in order to exclude funds that are very short-lived; in addition, we based most tests on the t-statistic, which is less sensitive to these variance outliers. However, it is possible that this minimum return requirement may impose a survival bias on our results–perhaps bootstrapping is less (or more) necessary for proper inference when we do not impose such a requirement.

To explore this concern, we vary the return requirement to include funds having at least 18, 30, 90, and 120 months, respectively. The results, shown in Table V, strongly indicate that our inference about the tails of the performance distribution remains qualitatively similar as we move from a requirement of 18 monthly returns to a requirement of 120 returns. Note that a requirement of 120 months eliminates the extreme left and right tails, but the bootstrap responds by moving deeper into the distribution to identify outperforming funds. Thus, the bootstrap performs consistently through different types of data inclusion rules.

F. Persistent Omitted Factor in Fund Residuals

Our return generating models assume that there is not a persistent source of variation in fund residuals. If a persistent non-priced factor (e.g., an oil shock factor) is missing from our model, then perhaps we might erroneously conclude that the alpha of a fund with a loading on this factor (e.g., an energy fund) is extreme due to manager talent. Thus, it is important to test for the possibility of a persistent missing factor in our models.

We provide two pieces of evidence against such a missing non-priced factor. First, the argument of a missing factor is only plausible for a relatively short-lived fund that exists only during the period when this factor affected fund returns. Our results described in the prior section show that our broad conclusions hold, whether we require a fund to exist for 18 or 120 months.

Second, we apply a variation of the bootstrap that uses a Monte Carlo simulation of a hypothetical omitted factor in the residuals. To capture the persistent nature of the omitted factor, we model it as a slowly mean-reverting AR(1) process and choose parameter values that are consistent with the observed fund data. Details of the Monte Carlo simulation are available from the authors upon request, but they are broadly consistent with our baseline results; thus, we find it very unlikely that an omitted factor is driving our main performance results.

G. Bootstrap Tests for Stockholdings-Based Alphas

To further examine the robustness of our results, we examine the importance of the bootstrap for evaluating the stockholdings-based performance measure of Daniel, Grinblatt, Titman and Wermers (1997; DGTW). This "Characteristic-Selectivity Measure," computed by matching each stock held by a fund with a value-weighted portfolio of stocks having similar size, book-to-market, and momentum characteristics, is described as

$$CS_t = \sum_{j=1}^{N} \widetilde{w}_{j,t-1}(\widetilde{R}_{j,t} - \widetilde{R}_t^{b_{j,t-1}}),$$

where $\widetilde{w}_{j,t-1}$ is the portfolio weight on stock j at the end of month $t-1$, $\widetilde{R}_{j,t}$ is the month t buy-and-hold return of stock j, and $\widetilde{R}_t^{b_{j,t-1}}$ is the month t buy-and-hold return of the value-weighted matching benchmark portfolio. The construction of the benchmarks follows the procedure in Daniel, Grinblatt, Titman and Wermers (1997).

This measure, besides providing an analysis of fund returns before all costs, provides an alternative benchmarking approach. Specifically, if portfolios of certain types of stocks, such as small-capitalization, value stocks are able to outperform the Carhart four-factor model, we would observe funds that predominantly hold these stocks in the right tail of our cross-section of alphas. If this outperformance is due to non-linear factor return premia, then the DGTW matching procedure will provide an improved benchmark for such stocks.

Analogous to our prior application of the bootstrap, we bootstrap the CS performance measure by subtracting the time-series average CS measure, for each fund, from each month's measure to arrive at a demeaned CS residual. Then, bootstrapping the fund return is simple–we resample these demeaned residuals to generate a bootstrapped sequence of monthly residuals, then compute the bootstrapped fund performance as the average residual and the t-statistic of the average residual. This procedure is repeated for 1,000 bootstrapped iterations for each fund, and the cross-sectional distribution of CS measures is constructed from these bootstrap outcomes. This process is repeated 1,000 times to construct the cross-sectional empirical distribution of the time-series t-statistics of the CS measures.

Table VI reports bootstrap results for the CS performance measure. Again, we focus on discussing the distribution of t-statistics. Although (as we would expect from its pre-cost nature) the distribution of CS measures is shifted slightly to the right, compared to the unconditional four-factor model of Table II, the right-tail bootstrap results are very consistent with our prior results that are after costs. Specifically, significant performance, according to the bootstrapped t-statistic, now extends to all funds at or above the 20-percentile point (see Panel A), as opposed to the 5-percentile cut-off for our after-cost results of Table II. Thus, funds between the 5^{th} and 20^{th} percentiles are skilled, but cannot generate performance sufficient to overcome their expenses and trading costs. It is important to note that the t-statistic bootstrap finds no evidence of underperforming mutual funds, gross of expenses and trading costs–it is easy to understand this outcome, as one cannot imagine a fund manager who perversely attempts to underperform her benchmarks. Thus, the significant underperformance documented in Table II is entirely due to funds that cannot pick stocks well enough to cover their costs, and not to funds that somehow consistently choose underperforming stocks. Again, outperformance is much more prevalent among growth-oriented than income-oriented funds. In addition, inference based on the bootstrap deviates substantially in both the left and right tails of the distributions.

V Performance Persistence

Our analysis in Sections III and IV demonstrates that the performance of the top aggressive-growth and growth managers is not an artifact of luck. This finding implies that some level of persistence in performance is present as well, although the extent and duration of such persistence is not yet known. Persistence is also an interesting issue in light of the paper of Lynch and Musto (2002), which predicts persistence among winning funds, but not among losers; and the paper of Berk and Green (2004), which predicts negligible persistence among winning funds. Following these papers, we measure persistence in fund performance, net of trading costs and fees.

Perhaps the most influential paper on performance persistence is Carhart (1997)–that paper tests whether the alpha from the unconditional four-factor model persists over one- to three-year periods. Carhart's general results are that persistence in superior fund performance is very weak, or even nonexistent.[31] To test the robustness of Carhart's results, implement a bootstrap analysis of the Carhart sorting procedure (rather than Carhart's standard t-tests) to evaluate the significance of the future alphas of past winning and losing funds. The application of the bootstrap will sharpen the estimates of p-values, but will not change the alpha point estimates from Carhart's.[32]

In our baseline persistence tests, we rank funds using the alpha (intercept) of the unconditional four-factor model, measured over the three-years prior to a given year-end. For example, funds are first ranked on January 1, 1978 by their four-factor alphas over the period 1975 to 1977, and the excess returns of funds are measured over the following year (1978, in this case).[33] This process

[31]In general, other studies find similar results: Gruber (1996) finds weak persistence among superior funds; Bollen and Busse (2002) find evidence of very short-term persistence (at the quarterly frequency); Teo and Woo (2001) find that losing funds strongly persist for up to six years; and Wermers (2004) finds strong evidence of multi-year persistence in superior growth funds, but at the stockholdings level (pre-expenses and trading costs).

[32]However, two further differences in our study are also important to note, and they will affect the point estimates. First, our dataset covers the 1975 to 2002 (inclusive) period, while Carhart's covers the 1962 to 1993 (inclusive) period. Second, and more importantly, we combine shareclasses into portfolios before ranking funds on net returns at the portfolio level, while Carhart ranks shareclasses directly. Our approach, therefore, reduces the influence of small shareclasses, especially during the latter years of our sample period. In addition, shareclasses of a single portfolio, by construction, have almost perfectly correlated net returns–the only difference being due to uneven changes in expense ratios across the shareclasses during the time-period under consideration. This invalidates the assumption of independent residuals in cross-sectional regressions. This consideration is only important during the post-1990 period, when multiple shareclasses became significant in the U.S. fund industry.

[33]Funds are required to have 36 monthly return observations during the four years prior to the ranking date, but need not have complete return information during the test year (to minimize survival bias). Weights of portfolios are

is repeated through our last ranking date, January 1, 2002. Four-factor alphas are then computed for equal-weighted, ranked portfolios in the cross-section, which will consist of different funds over time. That is, to remain consistent with Carhart, we form equal-weighted portfolios of funds rather than examining individual funds (as we did in prior sections of this paper), with the exception of the very top and bottom funds.[34]

Panel A of Table VII shows that the top-ranked fund (the identity of which changes over the years), ranked by its lagged three-year alpha, generates a test-year alpha of 0.48 percent per month, which is significant using either the cross-sectionally bootstrapped right-tail p-value, or the right-tail p-value corresponding to a simple t-test (labeled as "one-tailed parametric p-value of alpha). In fact, the standard t-test provides inference similar to that of the bootstrap for many ranked-fund fractiles, except that the bootstrap provides the very important insight that the top decile consists of skilled funds. In agreement with Carhart, our standard t-test does not reject the null of no performance for these funds, while the bootstrap strongly rejects.[35]

Some statistics that describe these fractile portfolio distributions are helpful in understanding why the bootstrap differs from the standard t-test. Panel A provides a normality test (Jarque-Bera) as well as a measure of standard deviation, skew, and kurtosis for each portfolio. Note that the standard deviations indicate heterogeneity in risk-taking in the cross-section, and the skew and kurtosis measures indicate some important non-normalities in portfolio returns. Either of these factors can explain why the cross-sectional bootstrap differs from the standard t-tests implemented by Carhart.

Since our alphas (and t-statistics) are computed net of expenses (and security-level trading costs), one might wonder what level of pre-expense alphas our ranked funds generate. This question addresses whether stockpicking skills are present, whether or not expenses effectively capture all

readjusted whenever a fund disappears during the test year.

[34]To generate bootstrapped p-values of the $t-$statistic, we follow a procedure analogous to the bootstrap algorithm described in Section I.B. Specifically, we bootstrap fund excess returns during the test year using factor loadings estimated during the prior three-year period, under the null of a zero true alpha during the test year for each ranked fund or portfolio of funds. This process is repeated for each test year to build a full time-series of test-year bootstrapped excess fund (or portfolio of fund) returns over all test years for each ranking point in the alpha distribution. We next estimate the alpha and t-statistic of alpha for each fractile portfolio (or individual fund), then repeat the above for all remaining bootstrap iterations. Thus, funds are ranked on their three-year lagged alphas, but inferences are made based on the bootstrapped distribution of their test-year t-statistics.

[35]In general, inference using standard t-tests agree with those of the bootstrap in the left tail, mainly because their underperformance is so large. These past losing funds exhibit even higher levels of skewness and kurtosis than past winning funds. In addition, we find positive and significant alphas for the spread portfolios (e.g., the top minus bottom 10 percent, labeled "sprd 10%"), which is not surprising, based on the strong results for high- and low-ranked funds.

of the consumer surplus that is generated. In Panel A, we report the time-series average expense ratio for ranked fractile portfolios. The top decile of funds averages a test-year expense ratio of 97 basis points, which increases our point estimate of alpha to roughly two percent per year (since time-series variability in expense ratios is trivial, this point estimate is also bootstrap-significant).[36] Interestingly, near-median funds (deciles five and six) exhibit slightly negative alphas when expenses are added back. Thus, these fund managers appear to have some stockpicking skills, but are too inefficient at picking stocks to justify their costs. With our bottom decile of funds, even adding back expenses does not bring them above water–their trading costs apparently far outweigh any selectivity abilities.

Panel B of Table VII repeats our tests, using the one-year past unconditional four-factor alpha as the ranking variable. Shortening the ranking period is important, not from any attempt to try to mine the data for persistence. Rather, it demonstrates that shorter ranking periods result in highly-ranked funds with alpha distributions that are much more non-normal than those from longer ranking periods, as well as resulting in greater heterogeneity in risk-taking in the cross-section. Specifically, as shown in Panel B, the skew and kurtosis deviate much more strongly from normality among almost all highly ranked funds and fund portfolios. Thus, funds with extreme positive lagged alphas, based on short-term rankings, are much more likely to hold stocks, perhaps purposely, with a temporarily high standard deviation, skewness, and kurtosis–this leads us to conclude that the bootstrap is more important when evaluating short-term persistence. Nevertheless, standard t-tests agree with most of the bootstrapped results, except that the bootstrap shows that the top three deciles of funds have significant alphas, rather than the top two (as shown by the t-test).

In unreported tests, we repeat these persistence tests among investment-objective subgroups, finding that growth-oriented funds exhibit strong persistence, while income-oriented funds exhibit little. These results are, again, consistent with our baseline (non-persistence) bootstrap tests of performance of prior sections.

To summarize, we find some important differences from Carhart (1997) with the bootstrap applied to fund persistence. Extreme funds, based on their past alphas, exhibit portfolio returns that deviate very significantly from normality. Controlling for these non-normal distributions reverses one of the central results of Carhart–that is, we find that performance does indeed strongly persist

[36] In unreported tests, we repeat our baseline persistence tests using gross returns–that is, monthly net returns with expense ratios (divided by 12) added back. The bootstrap confirms what we would suspect–all funds, past winners and losers, have higher alphas, but bootstrap p-values do not change very much.

among the top decile of funds, ranked on their three-year past four-factor alphas.

VI Conclusion

We test whether the four-factor alphas of "star" mutual fund managers are due to luck or genuine stockpicking skills. In particular, we examine the statistical significance of the performance and performance persistence of the "best" and "worst" funds by means of a flexible bootstrap procedure applied to a variety of unconditional and conditional factor models of performance. Our findings indicate that the performance of the best and worst managers is not solely due to luck; that is, it cannot be explained solely by sampling variability. We also uncover large differences between the performance of funds with different investment objectives. While there is strong evidence of superior performance and performance persistence among growth-oriented funds, using bootstrap tests for significance, there is no evidence of ability among managers of income-oriented funds.

The methods presented in this study may also prove useful in addressing other questions in finance, especially 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 eliminating the need: to specify the exact shape of the distribution from which alphas are drawn, to estimate correlations between fund returns (which is infeasible in the presence of non-overlapping fund data), and to explicitly control for potential 'data snooping' biases arising from an ex-post sort on alphas.

In summary, our evidence points to the need for the bootstrap in future rankings of mutual fund performance. At the very least, without the bootstrap, rankings on the appraisal ratio or the t-statistic of the alpha help to reduce (but do not eliminate) problems with ex-post sorts as described above.

References

Baks, K., 2004, On the performance of mutual fund managers, Manuscript, Emory University.

Baks, K.P., A. Metrick and J. Wachter, 2001, Should investors avoid all actively managed mutual funds? A study in Bayesian performance evaluation, *Journal of Finance* 56, 45-85.

Berk, J. and R. Green, 2004, Mutual fund flows and performance in rational markets, Manuscript, Carnegie-Mellon University.

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

Bollen, N. and J.A. Busse, 2002, Short-term persistence in mutual fund performance, *Review of Financial Studies*, forthcoming.

Bollerslev, T., R. Chou and K. Kroner, 1992, ARCH modeling in finance: A review of the theory and empirical evidence, *Journal of Econometrics* 52, 5-59.

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.

Busse, J., 1999, Volatility timing in mutual funds: Evidence from daily returns, *Review of Financial Studies* 12, 1009-1041.

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

Carhart, M., Carpenter, J., Lynch, A.W., and Musto, D., 2002, Mutual fund survivorship, *Review of Financial Studies* 15, 1439-1463.

Chen, H.L., N. Jegadeesh and R. Wermers, 2000, An examination of the stockholdings and trades of fund managers, *Journal of Financial and Quantitative Analysis* 35, 343-368.

Chevalier, J, and G. Ellison, 1997, Risk taking by mutual funds as a response to incentives, *Journal of Political Economy*, 105, 1167-1200.

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* 11, 111-142.

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

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

Efron, B. and R. J. Tibshirani, 1993, *An introduction to the bootstrap*, Chapman and Hall.

Fama, E.F. and K.R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3-56.

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

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., 1996, 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.

Horowitz, J.L., 2003, Bootstrap methods for Markov processes, *Econometrica* 71, 1049-1082.

Investment Company Institute, 2004, *Mutual fund fact book* (Washington, D.C.).

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

Lynch, A. and D. Musto, 2003, How investors interpret past fund returns, *Journal of Finance* 58, 5, 2033-2058,

Marcus, A.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.

Pastor, L. and R.F. Stambaugh, 2002, Mutual fund performance and seemingly unrelated assets, *Journal of Financial Economics* 63, 315–349.

Prather, L., W.J. Bertin and T. Henker, 2004, Mutual fund characteristics, managerial attributes, and fund performance, Review of Financial Economics, Volume 13, Issue 4, 2004, Pages 305-326.

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.

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

Teo, Melvyn and Sung-Jun Woo, 2001, Persistence in style-adjusted mutual fund returns, Manuscript, Harvard University.

Treynor, J.L. and F. 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., 1999, Mutual fund herding and the impact on stock prices, *Journal of Finance 55, 581-622.*

Wermers, R., 2000, Mutual Fund Performance: An empirical decomposition into stock-picking talent, style, transaction costs, and expenses, *Journal of Finance 55,* 1655 - 1703.

Wermers, R., 2004, Is money really smart? New evidence on the relation between mutual fund flows, manager behavior, and performance persistence. Manuscript, University of Maryland.

Table I

Summary Statistics On Mutual Fund Database

This table reports the number, returns and performance of U.S. open-end equity funds in existence during the 1975-2002 period. Panel A reports results for all investment objectives, Panel B for aggressive growth funds, Panel C for growth funds, Panel D for growth and income funds and Panel E for balanced or income funds (these two categories are combined). The first column in each panel of this table shows the number of funds in existence in each subperiod. The second column of each panel reports the number of funds in existence with at least 60 monthly net return observations during the subperiod. The row '1971-1975', for example, shows that during the 1971-75 subperiod there existed 322 funds, but only 231 of them had 60 monthly observations during the five year subperiod. The third column reports the excess return in percent per year (12 times the average monthly return) of an equally weighted portfolio of funds during the subperiod. The fourth column reports the excess return in percent per year for an equally-weighted portfolio of funds with at least 60 monthly observations during the subperiod. Column five reports the unconditional four-factor model alpha in percent per year for the equally-weighted portfolio of funds during the subperiod. Column six reports the four-factor alpha in percent per year for funds that had at least 60 monthly observations during the subperiod. The fourth column reports the excess return in percent per year for an equally-weighted portfolio of funds with at least 60 monthly observations during the subperiod. Column five reports the unconditional four-factor model alpha in percent per year for the equally-weighted portfolio of funds during the subperiod. Column six reports the four-factor alpha in percent per year for funds that had at least 60 monthly observations during the subperiod.

Panel A: All Investment Objectives

	Number of Funds		Excess Return of Equally-Weighted Portfolio (pct/year)		4-Factor-Alpha of Equally-Weighted Portfolio (pct/year)	
Min. # of Obs.	≥1	≥60	≥1	≥60	≥1	≥60
1971-1975	322	231	-9.8	-9.5	-0.1	0.1
1976-1980	342	282	11.9	12.1	-1.0	-0.7
1981-1985	459	300	4.1	4.0	0.2	0.1
1986-1990	796	421	9.5	9.6	0.2	0.0
1991-1995	1428	681	4.7	4.9	0.0	0.2
1996-2000	1929	1115	15.1	15.0	-1.8	-1.8
1998-2002	1824	1304	5.9	6.1	-1.0	-0.6
1975-2002	2118	1788	6.8	7.0	-0.5	-0.4

Panel B: Aggressive Growth Funds

	Number of Funds		Excess Return of Equally-Weighted Portfolio (pct/year)		4-Factor-Alpha of Equally-Weighted Portfolio (pct/year)	
Min. # of Obs.	≥1	≥60	≥1	≥60	≥1	≥60
1971-1975	84	65	-12.6	-12.0	-0.4	0.0
1976-1980	88	76	15.8	16.0	-1.6	-1.3
1981-1985	132	79	3.8	3.5	-0.1	-0.7
1986-1990	208	122	9.8	9.8	0.5	0.1
1991-1995	240	177	7.1	7.4	1.0	1.2
1996-2000	242	191	17.4	17.3	-2.4	-2.3
1998-2002	232	200	6.6	7.0	-1.1	-0.8
1975-2002	285	264	8.0	8.0	-0.4	-0.4

Panel C: Growth Funds

	Number of Funds		Excess Return of Equally-Weighted Portfolio (pct/year)		4-Factor-Alpha of Equally-Weighted Portfolio (pct/year)	
Min. # of Obs.	≥1	≥60	≥1	≥60	≥1	≥60
1971-1975	99	65	-11.2	-10.8	0.2	0.8
1976-1980	108	84	12.6	13.1	-1.2	-0.8
1981-1985	158	91	4.8	4.7	1.2	1.2
1986-1990	292	142	10.0	9.8	0.5	0.1
1991-1995	692	251	4.2	4.4	-0.5	-0.2
1996-2000	1145	527	16.1	16.1	-1.6	-1.7
1998-2002	1079	697	6.5	7.0	-0.7	-0.1
1975-2002	1227	985	7.2	7.4	-0.4	-0.1

Panel D: Growth and Income Funds

	Number of Funds		Excess Return of Equally-Weighted Portfolio (pct/year)		4-Factor-Alpha of Equally-Weighted Portfolio (pct/year)	
Min. # of Obs.	≥1	≥60	≥1	≥60	≥1	≥60
1971-1975	89	64	-8.2	-7.8	-0.1	-0.1
1976-1980	93	76	10.3	10.7	-0.2	-0.1
1981-1985	112	82	4.0	3.8	0.0	-0.1
1986-1990	195	105	9.6	9.7	-0.4	-0.4
1991-1995	330	169	3.7	4.0	-0.2	0.0
1996-2000	355	265	13.6	13.8	-1.8	-1.6
1998-2002	335	270	4.6	4.7	-1.4	-1.2
1975-2002	396	353	6.2	6.2	-1.1	-1.0

Panel E: Balanced or Income Funds

	Number of Funds		Excess Return of Equally-Weighted Portfolio (pct/year)		4-Factor-Alpha of Equally-Weighted Portfolio (pct/year)	
Min. # of Obs.	≥1	≥60	≥1	≥60	≥1	≥60
1971-1975	50	37	-5.6	-5.9	-0.6	-0.7
1976-1980	53	46	6.8	6.7	-0.8	-1.0
1981-1985	57	48	3.8	3.7	-0.4	-0.6
1986-1990	101	52	7.9	7.8	0.5	0.0
1991-1995	166	84	3.4	3.4	-0.2	-0.2
1996-2000	187	132	10.1	10.2	-1.9	-1.9
1998-2002	178	137	3.2	3.5	-1.1	-0.8
1975-2002	210	186	4.9	4.9	-0.7	-0.6

Table II

The Cross-Section of Mutual Fund Alphas

In Panel A of this table, all U.S. open-end domestic equity funds that have at least 60 monthly net return observations during the 1975 to 2002 period are ranked on their four-factor alpha. The first to third rows report results when funds are ranked on their unconditional four-factor alphas. The fourth to sixth rows report results when funds are ranked on their conditional four-factor alphas. In Panel A, the first and second rows report the OLS estimate of alphas (in percent per month) and the cross-sectionally bootstrapped p-values of the alphas. For comparison, the third row reports the p-values of the t-statistic of alphas based on standard critical values of the t-statistic. The fourth and fifth rows of Panel A report conditional alphas (in percent per month) and the cross-sectionally bootstrapped p-values of conditional alphas. The sixth row reports the p-values of the t-statistic of alphas, based on standard critical values of the t-statistic. In Panel B, all U.S. open-end domestic equity funds that have at least 60 monthly net return observations during the 1975 to 2002 period are ranked on the t-statistics of their four-factor alpha (first row). The second row of Panel B shows the cross-sectionally bootstrapped p-values of the t-statistics. For comparison, the third row shows the p-values of the t-statistic based on standard critical values. Rows four to six show results when funds are ranked based on the t-statistic of their conditional four-factor model alpha. Rows four and five report the t-statistics of the conditional alpha and the cross-sectionally bootstrapped p-value of the t-statistic. Row six shows the p-values of the t-statistic of the conditional alphas, based on standard critical values. In each panel, the first columns on the left (right) report results for funds with the five lowest (highest) alphas or t-statistics, followed by results for marginal funds at different percentiles in the left (right) tail of the distribution. The cross-sectionally bootstrapped p-value is based on the distribution of the best (worst) funds in 1000 bootstrap resamples. The t-statistics of alpha are based on heteroskedasticity and autocorrelation consistent standard errors.

Panel A: Funds Ranked on Four-Factor Model Alphas

	Worst Funds													Best Funds											
	Bottom	2.	3.	4.	5.	1%	3%	5%	10%	20%	30%	40%	Median	40%	30%	20%	10%	5%	3%	1%	5.	4.	3.	2.	Top
Unconditional Alpha (%)	-3.6	-2.7	-2.0	-1.7	-1.5	-0.8	-0.6	-0.5	-0.4	-0.2	-0.2	-0.1	-0.1	0.0	0.1	0.1	0.3	0.4	0.6	1.0	1.3	1.4	1.5	1.6	4.2
Cross-Sectionally Bootstrapped p-value	0.04	<0.01	<0.01	<0.01	<0.01	0.02	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01		1.00	1.00	0.99	0.02	<0.01	<0.01	<0.01	0.02	0.06	0.08	0.19	0.02
Parametric (standard) p-value	0.04	<0.01	<0.01	<0.01	<0.01	<0.01	0.06	0.01	0.09	<0.01	0.40	0.36		0.50	0.31	0.29	0.12	0.03	0.03	0.01	0.01	<0.01	0.12	0.01	<0.01
Conditional Alpha (%)	-4.8	-2.2	-2.0	-1.9	-1.8	-1.0	-0.6	-0.5	-0.4	-0.3	-0.2	-0.1	-0.1	0.0	0.0	0.1	0.3	0.4	0.6	1.1	1.4	1.5	1.6	1.8	4.2
Cross-Sectionally Bootstrapped p-value	<0.01	0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01		1	1	0.88	<0.01	<0.01	<0.01	<0.01	0.01	0.02	0.05	0.09	0.01
Parametric (standard) p-value	0.01	<0.01	<0.01	<0.01	<0.01	<0.01	0.01	0.05	0.03	0.02	0.12	0.37		0.58	0.40	0.13	0.25	0.09	0.02	0.02	0.15	<0.01	0.01	0.01	<0.01

Panel B: Funds Ranked on the t-Statistics of Four-Factor Model Alpha

	Worst Funds													Best Funds											
	Bottom	2.	3.	4.	5.	1%	3%	5%	10%	20%	30%	40%	Median	40%	30%	20%	10%	5%	3%	1%	5.	4.	3.	2.	Top
t-Unconditional Alpha (%)	-7.9	-6.1	-6.0	-4.8	-4.5	-3.6	-3.0	-2.4	-1.9	-1.4	-0.9	-0.6	-0.4	0.0	0.3	0.8	1.4	2.0	2.3	2.8	3.5	3.5	4.1	4.2	6.6
Cross-Sectionally Bootstrapped p-value	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01		1.00	1.00	1.00	0.25	<0.01	<0.01	<0.01	0.04	0.08	0.01	0.03	<0.01
Parametric (standard) p-value	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	0.01	0.03	0.09	0.17	0.26		0.49	0.37	0.23	0.09	0.03	0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
t-Conditional Alpha (%)	-8.5	-5.8	-5.8	-5.7	-5.2	-3.8	-3.2	-2.7	-2.1	-1.6	-1.1	-0.7	-0.4	-0.1	0.3	0.8	1.4	2.0	2.3	2.9	3.5	3.7	3.9	4.6	5.8
Cross-Sectionally Bootstrapped p-value	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01		1.00	1.00	1.00	0.04	<0.01	<0.01	<0.01	0.01	0.02	0.02	0.01	0.01
Parametric (standard) p-value	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	0.02	0.06	0.13	0.23		0.55	0.40	0.22	0.08	0.03	0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01

Table III

The Cross-Section of Mutual Fund Alphas, for Subperiods

This table reports subperiod results for the cross-section of funds' performance measures. In Panel A, all U.S. open-end domestic equity funds that have at least 60 monthly net return observations during the 1975 to 1989 period are ranked on their unconditional four-factor model alphas. The first and second rows report the OLS estimate of alphas (in percent per month) and the cross-sectionally bootstrapped *p* -values of alpha. For comparison, the third row reports the p-values of the t-statistic of alphas, based on standard critical values of the t-statistic. In Panel B, all U.S. open-end domestic equity funds that have at least 60 monthly net returns during the 1975 to 1989 period are ranked on the t-statistics of their unconditional four-factor model alpha. The first row shows the t-statistics of the unconditonal alpha. The second row reports the cross-sectionally bootstrapped p-values of the t-statistic of alphas. For comparison, the third row shows the p-values of the t-statistic, based on standard critical values. Panels C and D report the same measures as Panels A and B, but for the subperiod 1990-2002. In each panel, the first columns on the left (right) report results for funds with the five lowest (highest) alphas or t-statistics, followed by marginal funds at different percentiles in the left (right) tail of the distribution. The cross-sectionally bootstrapped p-value is based on the distribution of the best (worst) funds in 1000 bootstrap resamples. The t-statistics of alpha are based on heteroskedasticity and autocorrelation consistent standard errors.

Panel A: Funds Ranked on Four-Factor Model Alphas (1975-1989)

						Worst Funds													Best Funds						
	Bottom	2.	3.	4.	5.	1%	3%	5%	10%	20%	30%	40%	Median	40%	30%	20%	10%	5%	3%	1%	5.	4.	3.	2.	Top
Unconditional Alpha (%)	-2.7	-1.4	-1.2	-1.1	-1.0	-1.0	-0.6	-0.5	-0.3	-0.2	-0.1	-0.1	0.0	0.1	0.1	0.2	0.3	0.5	0.6	0.7	0.7	0.8	0.9	1.0	1.0
Cross-Sectionally Bootstrapped p-value	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	0.03	0.10		<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	0.01	0.01	0.01	0.02	0.05	0.36
Parametric (standard) p-value	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	0.10	0.02	0.01	0.28	0.16	0.36		0.33	0.35	0.08	0.01	0.09	0.03	<0.01	<0.01	<0.01	0.01	<0.01	0.01

Panel B: Funds Ranked on t-Statistics of Four-Factor Model Alphas (1975-1989)

						Worst Funds													Best Funds						
	Bottom	2.	3.	4.	5.	1%	3%	5%	10%	20%	30%	40%	Median	40%	30%	20%	10%	5%	3%	1%	5.	4.	3.	2.	Top
t-Unconditional Alpha (%)	-6.1	-5.2	-4.7	-4.6	-4.0	-4.0	-2.9	-2.3	-1.8	-1.0	-0.6	-0.4	0.0	0.4	0.8	1.2	1.9	2.5	2.8	3.4	3.4	3.5	3.5	4.4	4.6
Cross-Sectionally Bootstrapped p-value	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	0.03	0.14	0.03		0.03	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	0.01	<0.01	0.02
Parametric (standard) p-value	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	0.01	0.03	0.16	0.27	0.35		0.36	0.21	0.11	0.03	0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01

Panel C: Funds Ranked on Four-Factor Model Alphas (1990-2002)

						Worst Funds													Best Funds						
	Bottom	2.	3.	4.	5.	1%	3%	5%	10%	20%	30%	40%	Median	40%	30%	20%	10%	5%	3%	1%	5.	4.	3.	2.	Top
Unconditional Alpha (%)	-3.8	-3.6	-2.7	-1.5	-1.4	-0.9	-0.6	-0.5	-0.4	-0.3	-0.2	-0.1	-0.1	0.0	0.1	0.1	0.3	0.4	0.6	1.2	1.4	1.5	1.6	1.6	4.2
Cross-Sectionally Bootstrapped p-value	0.04	<0.01	<0.01	0.02	0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01		1.00	1.00	0.99	0.23	<0.01	<0.01	<0.01	0.02	0.01	0.07	0.19	0.02
Parametric (standard) p-value	<0.01	0.04	<0.01	<0.01	<0.01	0.01	0.01	0.04	0.05	0.07	0.24	0.25		0.49	0.44	0.07	<0.01	0.03	0.16	<0.01	<0.01	0.12	<0.01	0.01	<0.01

Panel D: Funds Ranked on t-Statistics of Four-Factor Model Alphas (1990-2002)

						Worst Funds													Best Funds						
	Bottom	2.	3.	4.	5.	1%	3%	5%	10%	20%	30%	40%	Median	40%	30%	20%	10%	5%	3%	1%	5.	4.	3.	2.	Top
t-Unconditional Alpha (%)	-7.3	-6.6	-6.0	-5.0	-4.7	-3.6	-2.9	-2.5	-1.9	-1.4	-1.0	-0.6	-0.4	-0.1	0.3	0.7	1.3	1.9	2.2	2.8	3.4	3.5	4.1	4.2	6.6
Cross-Sectionally Bootstrapped p-value	0.02	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01		1.00	1.00	1.00	0.50	<0.01	<0.01	0.01	0.04	0.12	<0.01	0.03	<0.01
Parametric (standard) p-value	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	0.01	0.03	0.09	0.17	0.27		0.48	0.38	0.23	0.09	0.03	0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01

Table IV

The Cross-Section of Mutual Fund Alphas, by Investment Objective

This table reports the cross-section of alphas, by investment objective. In Panel A all growth funds that have at least 60 monthly net return observations during the 1975 to 2002 period are ranked on their unconditional four-factor model alphas. The first and second rows report the OLS estimate of alphas (in percent per month) and the cross-sectionally bootstrapped p-value of the unconditional four-factor alpha. For comparison, the third row reports the p-values of the t-statistic of alphas, based on standard critical values of the t-statistic. In rows four to six of Panel A all growth funds that have at least 60 monthly net return observations during the 1975 to 2002 period are ranked on the t-statistics of their unconditional four-factor model alphas. The fourth row shows the t-statistics of the alpha. The fifth row reports the cross-sectionally bootstrapped p-values of the t-statistic of alphas. For comparison, the sixth row shows the p-values of the t-statistic, based on standard critical values. Panels B, C and D report the same measures as Panel A, but for the investment objectives of aggressive growth, growth & income, and balanced or income. In each panel, the first columns on the left (right) report results for funds with the five lowest (highest) alphas or t-statistics, followed by results for marginal funds at different percentiles in the left (right) tail of the distribution. The cross-sectionally bootstrapped p-value is based on the distribution of the best (worst) funds in 1000 bootstrap resamples. The t-statistics of alpha are based on heteroskedasticity and autocorrelation consistent standard errors.

Panel A: Growth

Funds Ranked on Four-Factor Model Alphas

	Worst Funds																			Best Funds
	Bottom	2.	3.	4.	5.	1%	3%	5%	10%	20%	20%	10%	5%	3%	1%	5.	4.	3.	2.	Top
Unconditional Alpha (%)	-3.6	-2.7	-1.7	-1.5	-1.2	-0.8	-0.6	-0.5	-0.4	-0.3	0.1	0.3	0.4	0.6	1.0	1.3	1.4	1.5	1.6	4.2
Cross-Sectionally Bootstrapped p-value	0.05	<0.01	0.01	0.01	0.07	0.25	0.02	<0.01	<0.01	<0.01	0.99	0.02	<0.01	<0.01	<0.01	0.02	0.06	0.08	0.19	0.02
Parametric (standard) p-value	0.04	<0.01	<0.01	<0.01	<0.01	0.06	<0.01	0.18	0.08	0.31	0.29	0.12	0.03	0.03	0.01	0.01	0.00	0.12	0.01	<0.01

Funds Ranked on t-Statistics of Four-Factor Model Alphas

	Bottom	2.	3.	4.	5.	1%	3%	5%	10%	20%	20%	10%	5%	3%	1%	5.	4.	3.	2.	Top
t-Unconditional Alpha (%)	-7.9	-4.8	-4.2	-4.0	-3.9	-3.6	-2.9	-2.4	-1.8	-1.3	0.8	1.4	2.0	2.3	2.8	3.5	3.5	4.1	4.2	6.6
Cross-Sectionally Bootstrapped p-value	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	1.00	0.25	<0.01	<0.01	<0.01	0.04	0.08	0.01	0.03	<0.01
Parametric (standard) p-value	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	0.01	0.04	0.10	0.23	0.09	0.03	0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01

Panel B: Aggressive Growth

Funds Ranked on Four-Factor Model Alphas

	Bottom	2.	3.	4.	5.	1%	3%	5%	10%	20%	20%	10%	5%	3%	1%	5.	4.	3.	2.	Top
Unconditional Alpha (%)	-2.0	-1.2	-1.1	-1.1	-0.9	-1.1	-0.7	-0.6	-0.4	-0.3	0.2	0.3	0.6	0.7	0.9	0.9	0.9	0.9	0.9	1.1
Cross-Sectionally Bootstrapped p-value	<0.01	0.02	0.01	<0.01	<0.01	0.01	<0.01	<0.01	<0.01	<0.01	0.33	<0.01	<0.01	<0.01	0.05	<0.01	0.02	0.05	0.20	0.37
Parametric (standard) p-value	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	0.01	0.03	0.09	0.20	0.07	0.01	0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01

Funds Ranked on t-Statistics of Four-Factor Model Alphas

	Bottom	2.	3.	4.	5.	1%	3%	5%	10%	20%	20%	10%	5%	3%	1%	5.	4.	3.	2.	Top
t-Unconditional Alpha (%)	-6.1	-3.8	-1.5	-2.1	-2.4	-1.5	-4.5	-1.7	-1.5	-1.2	0.7	2.4	2.6	2.9	2.2	3.1	1.2	2.2	2.3	3.5
Cross-Sectionally Bootstrapped p-value	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	0.63	0.13	<0.01	<0.01	0.04	0.02	0.06	0.04	0.12	0.19
Parametric (standard) p-value	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	0.01	0.03	0.09	0.20	0.07	0.01	0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01

Panel C: Growth and Income

Funds Ranked on Four-Factor Model Alphas

	Bottom	2.	3.	4.	5.	1%	3%	5%	10%	20%	20%	10%	5%	3%	1%	5.	4.	3.	2.	Top
Unconditional Alpha (%)	-0.9	-0.8	-0.8	-0.7	-0.6	-0.7	-0.5	-0.4	-0.3	-0.2	0.1	0.1	0.3	0.3	0.5	0.4	0.5	0.5	0.6	0.8
Cross-Sectionally Bootstrapped p-value	0.11	0.03	0.01	0.01	<0.01	0.01	<0.01	<0.01	<0.01	<0.01	0.97	1.00	0.22	0.35	0.27	0.35	0.27	0.20	0.39	0.26
Parametric (standard) p-value	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	0.03	0.06	0.15	0.24	0.31	0.17	0.01	0.09	0.02	0.03	0.02	0.09	0.03	0.05

Funds Ranked on t-Statistics of Four-Factor Model Alphas

	Bottom	2.	3.	4.	5.	1%	3%	5%	10%	20%	20%	10%	5%	3%	1%	5.	4.	3.	2.	Top
t-Unconditional Alpha (%)	-4.4	-4.0	-3.8	-3.6	-3.6	-3.6	-3.0	-2.5	-2.1	-1.4	0.7	1.3	1.7	1.9	2.5	2.4	2.5	2.6	2.7	3.5
Cross-Sectionally Bootstrapped p-value	0.03	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	0.99	0.68	0.59	0.69	0.46	0.35	0.46	0.49	0.63	0.27
Parametric (standard) p-value	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	0.01	0.02	0.08	0.25	0.10	0.05	0.03	0.01	0.01	0.01	0.01	<0.01	<0.01

Panel D: Balanced or Income

Funds Ranked on Four-Factor Model Alphas

	Bottom	2.	3.	4.	5.	1%	3%	5%	10%	20%	20%	10%	5%	3%	1%	5.	4.	3.	2.	Top
Unconditional Alpha (%)	-0.6	-0.6	-0.5	-0.4	-0.4	-0.6	-0.3	-0.3	-0.3	-0.2	0.1	0.1	0.2	0.3	0.4	0.3	0.3	0.4	0.4	0.5
Cross-Sectionally Bootstrapped p-value	0.37	0.05	0.01	0.06	0.03	0.05	0.01	<0.01	<0.01	<0.01	0.95	0.40	0.46	0.22	0.37	0.17	0.35	0.17	0.37	0.47
Parametric (standard) p-value	0.01	0.01	0.02	0.15	0.02	0.01	0.04	0.05	<0.01	0.04	0.17	0.21	0.05	0.11	<0.01	0.10	0.09	0.01	<0.01	0.05

Funds Ranked on t-Statistics of Four-Factor Model Alphas

	Bottom	2.	3.	4.	5.	1%	3%	5%	10%	20%	20%	10%	5%	3%	1%	5.	4.	3.	2.	Top
t-Unconditional Alpha (%)	-6.0	-3.6	-3.3	-3.3	-3.2	-3.6	-3.1	-2.6	-2.3	-1.6	0.7	1.2	1.7	2.0	2.4	2.1	2.1	2.2	2.4	2.8
Cross-Sectionally Bootstrapped p-value	<0.01	0.01	<0.01	<0.01	<0.01	0.01	<0.01	<0.01	<0.01	<0.01	0.98	0.81	0.65	0.51	0.71	0.50	0.69	0.74	0.71	0.64
Parametric (standard) p-value	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	0.01	0.01	0.05	0.26	0.11	0.05	0.03	0.01	0.02	0.02	0.02	0.01	<0.01

Table V

Sensitivity Analysis of the Cross-Section of Mutual Fund Alphas to Minimum Number of Observations

In Panel A, all U.S. open-end domestic equity funds that have at least 18 monthly net return observations during the 1975 to 2002 period are ranked on the t-statistics of their unconditional four-factor model alphas. The first row shows the t-statistics of the unconditional alphas. The second row reports the cross-sectionally bootstrapped p-values of the t-statistics. For comparison, the third row shows the p-value of the t-statistic, based on standard critical values. Panels B, C, D and E report the same measures as Panel A, but for funds with a minimum number of observations of 30, 60, 90 and 120 monthly net return observations, respectively. In each panel, the first columns on the left (right) report results for funds with the five lowest (highest) t-statistics, followed by results for marginal funds at different percentiles in the left (right) tail of the distribution. The cross-sectionally bootstrapped p-value is based on the distribution of the best (worst) funds in 1000 bootstrap resamples. The t-statistics of alpha are based on heteroskedasticity and autocorrelation consistent standard errors.

Panel A: Minimum of 18 Observations Per Fund

	Bottom	2.	3.	4.	5.	1%	3%	5%	10%	20%	20%	10%	5%	3%	1%	5.	4.	3.	2.	Top
t-Unconditional Alpha (%)	-7.9	-6.1	-6.0	-4.8	-4.5	-3.8	-3.1	-2.5	-2.1	-1.4	0.7	1.3	1.9	2.2	2.7	3.5	3.5	4.1	4.2	6.6
Cross-Sectionally Bootstrapped p-value	0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	1.00	0.93	0.01	<0.01	0.04	0.20	0.34	0.06	0.15	0.03
Parametric (standard) p-value	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	0.01	0.02	0.08	0.25	0.10	0.03	0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01

Panel B : Minimum of 30 Observations Per Fund

	Bottom	2.	3.	4.	5.	1%	3%	5%	10%	20%	20%	10%	5%	3%	1%	5.	4.	3.	2.	Top
t-Unconditional Alpha (%)	-7.9	-6.1	-6.0	-4.8	-4.5	-3.7	-3.1	-2.5	-2.1	-1.4	0.7	1.3	1.9	2.3	2.7	3.5	3.5	4.1	4.2	6.6
Cross-Sectionally Bootstrapped p-value	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	1.00	0.83	<0.01	<0.01	0.02	0.11	0.20	0.03	0.09	0.01
Parametric (standard) p-value	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	0.01	0.02	0.08	0.24	0.10	0.03	0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01

Panel C: Minimum of 60 Observations Per Fund

	Bottom	2.	3.	4.	5.	1%	3%	5%	10%	20%	20%	10%	5%	3%	1%	5.	4.	3.	2.	Top
t-Unconditional Alpha (%)	-7.9	-6.1	-6.0	-4.8	-4.5	-3.6	-3.0	-2.4	-1.9	-1.4	0.8	1.4	2.0	2.3	2.8	3.5	3.5	4.1	4.2	6.6
Cross-Sectionally Bootstrapped p-value	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	1.00	0.25	<0.01	<0.01	<0.01	0.04	0.08	0.01	0.03	<0.01
Parametric (standard) p-value	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	0.01	0.03	0.09	0.23	0.09	0.03	0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01

Panel D: Minimum of 90 Observations Per Fund

	Bottom	2.	3.	4.	5.	1%	3%	5%	10%	20%	20%	10%	5%	3%	1%	5.	4.	3.	2.	Top
t-Unconditional Alpha (%)	-7.9	-6.1	-6.0	-4.8	-4.5	-3.6	-2.6	-2.4	-1.9	-1.4	0.8	1.4	1.9	2.2	2.6	3.1	3.2	3.4	3.5	3.5
Cross-Sectionally Bootstrapped p-value	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	0.97	0.06	0.01	<0.01	0.13	0.13	0.16	0.11	0.30	0.64
Parametric (standard) p-value	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	0.01	0.03	0.09	0.22	0.08	0.03	0.01	0.01	<0.01	<0.01	<0.01	<0.01	<0.01

Panel E: Minimum of 120 Observations Per Fund

	Bottom	2.	3.	4.	5.	1%	3%	5%	10%	20%	20%	10%	5%	3%	1%	5.	4.	3.	2.	Top
t-Unconditional Alpha (%)	-7.9	-6.1	-4.8	-4.5	-3.8	-3.6	-2.6	-2.4	-2.0	-1.3	0.9	1.5	2.0	2.3	2.7	3.1	3.2	3.4	3.5	3.5
Cross-Sectionally Bootstrapped p-value	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	0.46	0.01	<0.01	<0.01	0.01	0.02	0.02	0.02	0.12	0.40
Parametric (standard) p-value	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	0.01	0.03	0.10	0.19	0.07	0.02	0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01

Table VI

The Cross-Section of Stockholdings-Based Performance Measures

In Panel A, all U.S. open-end domestic equity funds that have holdings available for at least five years during the 1975 to 2002 period are ranked on the CS measure introduced by Daniel, Grinblatt, Titman and Wermers (1997). The first and second rows of Panel A report the CS measures (in percent per month) and the cross-sectionally bootstrapped p-values of the CS measures. For comparison, the third row reports the p-values of the t-statistic of the CS measures based on standard critical values of the t-statistic. In rows four to six funds are ranked on the t-statistics of their CS measures. The fourth rows shows the t-statistics. The fifth row reports the cross-sectionally bootstrapped p-values of the t-statistics. For comparison, the sixth row shows the p-values of the t-statistic, based on standard critical values. Panels B, C, D and E report the same measures as Panel A but for growth, aggressive growth, growth and income and balanced or income funds, respectively. In each panel, the first columns on the left (right) report results for funds with the five lowest (highest) CS measures or t-statistics, followed by results for marginal funds at different percentiles in the left (right) tail of the distribution. The cross-sectionally bootstrapped p-value is based on the distribution of the best (worst) funds in 1000 bootstrap resamples. The t-statistics of the CS measure are based on heteroskedasticity and autocorrelation consistent standard errors.

Panel A: All Investment Objectives

Funds Ranked on CS Measure

	Bottom	2.	3.	4.	5.	1%	3%	5%	10%	20%	20%	10%	5%	3%	1%	5.	4.	3.	2.	Top
CS (pct/month)	-2.4	-2.0	-1.6	-1.5	-1.5	-0.7	-0.4	-0.3	-0.2	-0.1	0.2	0.4	0.6	0.7	0.9	1.2	1.3	1.4	1.6	1.7
Cross-Sectionally Bootstrapped p-value	0.29	0.14	0.16	0.09	0.05	0.76	1.00	1.00	1.00	1.00	<0.01	<0.01	<0.01	<0.01	0.19	0.29	0.35	0.40	0.51	0.74
Parametric (standard) p-value	0.03	<0.01	0.16	<0.01	0.09	0.07	0.06	0.19	0.12	0.14	0.17	0.08	0.01	0.03	<0.01	<0.01	0.21	0.05	0.05	0.07

Funds Ranked on t-Statistics of CS Measure

	Bottom	2.	3.	4.	5.	1%	3%	5%	10%	20%	20%	10%	5%	3%	1%	5.	4.	3.	2.	Top
t-Unconditional CS	-4.3	-3.2	-3.1	-3.0	-3.0	-2.1	-1.6	-1.4	-1.0	-0.5	1.2	1.7	2.0	2.3	2.9	3.1	3.2	3.3	3.5	3.7
Cross-Sectionally Bootstrapped p-value	0.13	0.76	0.60	0.49	0.49	1.00	1.00	1.00	1.00	1.00	<0.01	<0.01	<0.01	<0.01	<0.01	0.07	0.09	0.18	0.21	0.34
Parametric (standard) p-value	<0.01	<0.01	<0.01	<0.01	<0.01	0.02	0.06	0.09	0.16	0.32	0.12	0.05	0.02	0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01

Panel B: Growth

Funds Ranked on CS Measure

	Bottom	2.	3.	4.	5.	1%	3%	5%	10%	20%	20%	10%	5%	3%	1%	5.	4.	3.	2.	Top
CS (pct/month)	-2.4	-2.0	-1.5	-1.0	-1.0	-0.8	-0.5	-0.4	-0.2	-0.1	0.3	0.4	0.6	0.7	0.9	1.1	1.2	1.4	1.6	1.7
Cross-Sectionally Bootstrapped p-value	0.14	0.05	0.11	0.78	0.65	0.90	0.94	1.00	1.00	1.00	<0.01	<0.01	<0.01	<0.01	0.25	0.25	0.26	0.21	0.31	0.57
Parametric (standard) p-value	0.03	<0.01	<0.01	0.18	0.16	0.30	0.03	0.20	0.17	0.22	0.28	0.18	0.03	0.04	0.03	0.01	<0.01	0.05	0.05	0.07

Funds Ranked on t-Statistics of CS Measure

	Bottom	2.	3.	4.	5.	1%	3%	5%	10%	20%	20%	10%	5%	3%	1%	5.	4.	3.	2.	Top
t-Unconditional CS	-4.3	-3.2	-3.0	-3.0	-2.8	-2.4	-1.7	-1.4	-1.0	-0.5	1.3	1.7	2.0	2.2	2.9	3.1	3.1	3.2	3.3	3.5
Cross-Sectionally Bootstrapped p-value	0.11	0.51	0.40	0.31	0.29	0.78	1.00	1.00	1.00	1.00	<0.01	<0.01	<0.01	0.01	<0.01	0.01	0.05	0.11	0.25	0.42
Parametric (standard) p-value	<0.01	<0.01	<0.01	<0.01	<0.01	0.01	0.05	0.08	0.17	0.31	0.10	0.05	0.02	0.02	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01

Panel C: Aggressive Growth

Funds Ranked on CS Measure

	Bottom	2.	3.	4.	5.	1%	3%	5%	10%	20%	20%	10%	5%	3%	1%	5.	4.	3.	2.	Top
CS (pct/month)	-1.5	-0.9	-0.7	-0.6	-0.4	-0.7	-0.4	-0.3	-0.2	-0.1	0.2	0.4	0.5	0.7	0.9	0.8	0.9	0.9	0.9	1.0
Cross-Sectionally Bootstrapped p-value	0.2	0.32	0.53	0.82	0.99	0.53	0.97	1.00	1.00	1.00	<0.01	<0.01	<0.01	0.01	0.17	0.06	0.05	0.17	0.39	0.68
Parametric (standard) p-value	0.09	0.05	0.04	0.20	0.08	0.04	0.06	0.10	0.29	0.40	0.01	0.02	0.09	0.07	0.07	0.01	<0.01	0.07	<0.01	0.07

Funds Ranked on t-Statistics of CS Measure

	Bottom	2.	3.	4.	5.	1%	3%	5%	10%	20%	20%	10%	5%	3%	1%	5.	4.	3.	2.	Top
t-Unconditional CS	-1.8	-1.6	-1.6	-1.6	-1.4	-1.6	-1.3	-1.2	-0.8	-0.4	1.2	1.7	2.1	2.3	2.8	2.5	2.7	2.8	2.9	3.7
Cross-Sectionally Bootstrapped p-value	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	<0.01	<0.01	<0.01	0.02	0.04	0.07	0.04	0.04	0.13	0.05
Parametric (standard) p-value	0.04	0.05	0.05	0.06	0.08	0.05	0.09	0.12	0.21	0.34	0.12	0.04	0.02	0.01	<0.01	0.01	<0.01	<0.01	<0.01	<0.01

Panel D: Growth and Income

Funds Ranked on CS Measure

	Bottom	2.	3.	4.	5.	1%	3%	5%	10%	20%	20%	10%	5%	3%	1%	5.	4.	3.	2.	Top
CS (pct/month)	-1.6	-0.9	-0.4	-0.4	-0.3	-0.4	-0.3	-0.2	-0.2	-0.1	0.1	0.2	0.4	0.5	0.5	0.5	0.5	0.5	0.5	0.7
Cross-Sectionally Bootstrapped p-value	0.15	0.07	0.92	0.92	0.92	0.92	0.98	0.91	0.89	1.00	0.08	0.18	<0.01	0.01	0.40	0.12	0.26	0.40	0.65	0.74
Parametric (standard) p-value	0.16	<0.01	<0.01	0.17	0.08	<0.01	0.10	0.03	0.22	0.16	0.15	0.06	0.08	0.01	0.06	0.01	0.06	0.06	0.01	0.04

Funds Ranked on t-Statistics of CS Measure

	Bottom	2.	3.	4.	5.	1%	3%	5%	10%	20%	20%	10%	5%	3%	1%	5.	4.	3.	2.	Top
t-Unconditional CS	-3.1	-2.6	-2.6	-2.0	-2.0	-2.6	-1.6	-1.5	-1.1	-0.6	1.0	1.6	2.0	2.4	2.7	2.4	2.4	2.7	2.9	3.0
Cross-Sectionally Bootstrapped p-value	0.42	0.55	0.37	0.94	0.86	0.37	0.99	0.97	0.96	1.00	0.02	0.02	0.03	0.02	0.13	0.13	0.26	0.13	0.20	0.49
Parametric (standard) p-value	<0.01	<0.01	0.01	0.02	0.02	0.01	0.06	0.07	0.13	0.28	0.15	0.06	0.02	0.01	<0.01	0.01	0.01	<0.01	<0.01	<0.01

Panel E: Balanced or Income

Funds Ranked on CS Measure

	Bottom	2.	3.	4.	5.	1%	3%	5%	10%	20%	20%	10%	5%	3%	1%	5.	4.	3.	2.	Top
CS (pct/month)	-0.4	-0.3	-0.3	-0.2	-0.2	-0.4	-0.4	-0.3	-0.3	-0.1	0.2	0.3	0.4	1.3	1.3	0.2	0.2	0.3	0.4	1.3
Cross-Sectionally Bootstrapped p-value	0.69	0.42	0.34	0.49	0.40	0.69	0.69	0.42	0.34	0.79	0.13	0.27	0.39	0.22	0.22	0.13	0.25	0.27	0.39	0.22
Parametric (standard) p-value	0.23	0.10	0.01	0.17	0.21	0.23	0.23	0.10	0.01	0.24	0.05	0.10	0.03	0.21	0.21	0.17	0.05	0.10	0.03	0.21

Funds Ranked on t-Statistics of CS Measure

	Bottom	2.	3.	4.	5.	1%	3%	5%	10%	20%	20%	10%	5%	3%	1%	5.	4.	3.	2.	Top
t-Unconditional CS	-2.4	-1.3	-1.0	-0.8	-0.7	-2.4	-2.4	-1.3	-1.0	-0.7	1.1	1.7	1.9	2.4	2.4	1.3	1.7	1.7	1.9	2.44
Cross-Sectionally Bootstrapped p-value	0.32	0.81	0.88	0.86	0.79	0.32	0.32	0.81	0.88	0.69	0.17	0.18	0.23	0.23	0.23	0.16	0.05	0.18	0.23	0.23
Parametric (standard) p-value	0.01	0.1	0.17	0.21	0.23	0.01	0.01	0.10	0.17	0.24	0.14	0.05	0.03	0.01	0.01	0.10	0.05	0.1	0	0.01

Table VII

Bootstrap Performance Persistence Tests - All Investment Objectives

In Panel A, mutual funds are sorted on January 1 each year (from 1978 until 2002) into decile portolios, based on their unconditional four-factor model alphas estimated over the prior three years. We require a minimum of 36 monthy net return observations for this estimate. For funds that have missing observations during these prior three years, observations from the 12 months preceding the three-year window are added to obtain 36 observations. This assures that funds with missing observations are not excluded. The portfolios are equally weighted monthly, so the weights are readjusted whenever a fund disappears. Funds with the highest past three-year return comprise decile 1, and funds with the lowest comprise decile 10. The '5%ile' portfolio is a equally-weighted portfolio of the top five percent funds. The last four rows represent the difference in returns between top and bottom decile, 10 percentile, 5 percentile, 1 percentile, as well as between the ninth and tenth deciles. In Panel B, the portfolios are formed based on past one year alphas, and funds are held for one year. Column five reports the one-tailed parametric p-value of alpha. Columns six and seven report the cross-sectionally bootstrapped p-values for the t-statistic of alpha. Column six reports the probability that the bootstrapped t-statistic of alpha is lower than (-|t(alpha)|), i.e. the left tail of the bootstrapped distribution. Column seven reports the probability that the bootstrapped t-statistic of alpha is higher than (+|t(alpha)|), i.e. the right tail of the bootstrapped distribution. Columns 12 and 13 report the adjusted R-squared and the annual expense ratio. The expense ratio is calculated as the time-series average of the cross-sectional average of the expense ratios of the funds in the portfolios. The last three columns report the skewness, kurtosis and the p-value of the Jarque-Bera non-normality statistics. RMRF, SMB and HML are Fama and French's (1993) market proxy and factor-mimicking portfolios for size and book-to-market equity. PR1YR is a factor-mimicking portfolio for one-year return momentum. Alpha (in percent per month) is the intercept of the model.

Panel A: Three Year Ranking Periods, One Year Holding Period

Fractile	Excess Ret. (pct/ month)	Std. Dev.	Alpha (pct/ month)	t-stat of alpha	One-tailed parametric p-value of alpha	bootstr. p-value of t(alpha) (left tail)	bootstr. p-value of t(alpha) (right tail)	RMRF	SMB	HML	PR1YR	Adj. R^2	Exp. Ratio	Skew.	Kur.	p-value (JB-Test)
top	1.04	7.64	0.48	1.5	0.06	0.05	0.04	0.91	0.46	-0.56	0.19	0.65	1.04	0.2	5.2	<0.01
1%ile	0.54	6.79	0.11	0.7	0.24	0.38	0.20	1.03	0.53	-0.45	-0.06	0.89	0.98	-0.2	4.6	<0.01
5%ile	0.58	5.65	0.12	1.3	0.09	0.13	0.03	0.97	0.40	-0.27	-0.04	0.95	1.01	-0.1	5.0	<0.01
1.Dec	0.57	5.17	0.08	1.1	0.13	0.23	0.05	0.95	0.33	-0.15	-0.04	0.96	0.97	-0.2	4.7	<0.01
2.Dec	0.53	4.42	0.02	0.3	0.39	0.60	0.23	0.90	0.15	-0.02	-0.01	0.97	0.87	0.3	4.6	<0.01
3.Dec	0.53	4.21	-0.02	-0.4	0.35	0.45	0.27	0.90	0.10	0.06	0.01	0.98	0.86	0.3	4.7	<0.01
4.Dec	0.50	4.06	-0.02	-0.3	0.38	0.57	0.15	0.88	0.08	0.06	-0.01	0.98	0.84	0.1	4.4	<0.01
5.Dec	0.49	4.13	-0.04	-1.0	0.16	0.32	0.04	0.90	0.05	0.08	-0.01	0.98	0.86	0.3	5.3	<0.01
6.Dec	0.46	4.04	-0.08	-2.3	0.01	0.03	<0.01	0.89	0.03	0.08	0.01	0.98	0.85	0.4	5.8	<0.01
7.Dec	0.48	4.12	-0.09	-1.9	0.03	0.03	<0.01	0.90	0.05	0.10	0.02	0.97	0.88	0.2	4.8	<0.01
8.Dec	0.52	4.08	-0.07	-1.7	0.05	0.07	0.01	0.89	0.10	0.11	0.04	0.97	0.94	0.3	4.7	<0.01
9.Dec	0.50	4.25	-0.09	-1.6	0.06	0.10	0.03	0.90	0.14	0.08	0.03	0.97	1.03	0.5	6.2	<0.01
10.Dec	0.33	4.54	-0.29	-4.2	<0.01	<0.01	<0.01	0.93	0.26	0.08	0.04	0.96	1.30	0.7	5.5	<0.01
95%ile	0.15	4.76	-0.49	-6.0	<0.01	<0.01	<0.01	0.95	0.31	0.06	0.04	0.95	1.50	0.3	4.3	<0.01
99%ile	-0.22	5.66	-0.89	-5.2	<0.01	<0.01	<0.01	1.04	0.40	0.09	0.00	0.81	2.47	1.0	8.2	<0.01
bottom	-1.02	11.69	-1.38	-2.6	<0.01	<0.01	<0.01	0.87	0.98	-0.42	-0.11	0.32	5.26	2.4	20.5	<0.01
sprd10%	0.24	1.57	0.37	3.8	<0.01	<0.01	<0.01	0.02	0.08	-0.23	-0.07	0.36	-0.32	-0.3	4.7	<0.01
sprd5%	0.42	2.10	0.61	4.7	<0.01	<0.01	<0.01	0.01	0.09	-0.33	-0.08	0.36	-0.49	-0.4	4.4	<0.01
sprd1%	0.76	3.89	1.00	4.0	<0.01	<0.01	<0.01	-0.01	0.14	-0.55	-0.06	0.25	-1.49	-0.8	5.7	<0.01
Dec9_10	0.17	0.85	0.20	4.5	<0.01	<0.01	<0.01	-0.03	-0.12	0.01	0.00	0.27	-0.27	0.0	6.5	<0.01

Panel B: One Year Ranking Periods, One Year Holding Period

Fractile	Excess Ret. (pct/ month)	Std. Dev.	Alpha (pct/ month)	t-stat of alpha	One-tailed parametric p-value of alpha	bootstr. p-value of t(alpha) (left tail)	bootstr. p-value of t(alpha) (right tail)	RMRF	SMB	HML	PR1YR	Adj. R^2	Exp. Ratio	Skew.	Kur.	p-value (JB-Test)
top	0.84	9.14	0.14	0.4	0.36	0.39	0.31	1.10	0.85	-0.27	<0.01	0.55	1.06	0.3	8.7	<0.01
1%ile	0.79	6.50	0.14	0.9	0.19	0.13	0.24	1.04	0.61	-0.26	0.04	0.87	1.26	0.2	3.7	0.01
5%ile	0.82	5.69	0.14	1.5	0.06	0.05	0.06	1.00	0.49	-0.17	0.09	0.93	1.09	0.6	5.1	<0.01
1.Dec	0.78	5.30	0.14	1.9	0.03	0.04	0.01	0.97	0.41	-0.14	0.07	0.95	1.04	0.7	5.9	<0.01
2.Dec	0.66	4.49	0.07	1.5	0.07	0.11	<0.01	0.92	0.20	-0.01	0.03	0.97	0.92	0.4	4.9	<0.01
3.Dec	0.58	4.21	0.03	0.6	0.27	0.53	0.08	0.90	0.12	0.04	0.01	0.97	0.88	0.6	5.4	<0.01
4.Dec	0.54	4.08	<0.01	-0.1	0.46	0.80	0.15	0.89	0.09	0.06	<0.01	0.98	0.86	0.3	5.7	<0.01
5.Dec	0.53	4.01	-0.01	-0.3	0.37	0.62	0.17	0.88	0.07	0.07	<0.01	0.98	0.85	0.5	5.0	<0.01
6.Dec	0.46	3.97	-0.06	-1.8	0.04	0.10	<0.01	0.88	0.06	0.08	<0.01	0.98	0.87	0.2	4.0	<0.01
7.Dec	0.45	4.00	-0.08	-2.0	0.02	0.08	<0.01	0.88	0.06	0.07	<0.01	0.98	0.90	0.3	5.0	<0.01
8.Dec	0.42	4.12	-0.15	-3.1	<0.01	0.01	<0.01	0.90	0.09	0.07	0.02	0.97	0.92	0.0	4.1	<0.01
9.Dec	0.40	4.17	-0.18	-3.2	<0.01	<0.01	<0.01	0.91	0.11	0.09	0.01	0.96	0.95	-0.1	4.2	<0.01
10.Dec	0.28	4.39	-0.30	-3.7	<0.01	<0.01	<0.01	0.92	0.23	0.10	-0.03	0.93	1.22	-0.1	4.5	<0.01
95%ile	0.22	4.57	-0.39	-4.5	<0.01	<0.01	<0.01	0.94	0.28	0.11	-0.03	0.92	1.35	-0.1	4.1	<0.01
99%ile	-0.03	4.96	-0.66	-5.1	<0.01	<0.01	<0.01	0.93	0.41	0.14	-0.04	0.82	1.87	-0.2	4.5	<0.01
bottom	-0.59	9.96	-1.31	-2.7	<0.01	<0.01	<0.01	1.12	0.25	0.50	-0.17	0.23	2.20	2.8	24.8	<0.01
sprd10%	0.51	2.22	0.44	3.9	<0.01	<0.01	<0.01	0.05	0.18	-0.24	0.10	0.39	-0.18	0.4	4.5	<0.01
sprd5%	0.61	2.67	0.53	4.0	<0.01	<0.01	<0.01	0.06	0.21	-0.28	0.12	0.36	-0.26	0.5	4.3	<0.01
sprd1%	0.82	3.82	0.80	3.9	<0.01	<0.01	<0.01	0.11	0.21	-0.40	0.08	0.28	-0.60	0.2	3.1	0.34
Dec9_10	0.12	0.86	0.12	2.2	0.01	0.01	<0.01	-0.01	-0.12	-0.01	0.04	0.24	-0.26	0.8	7.6	<0.01

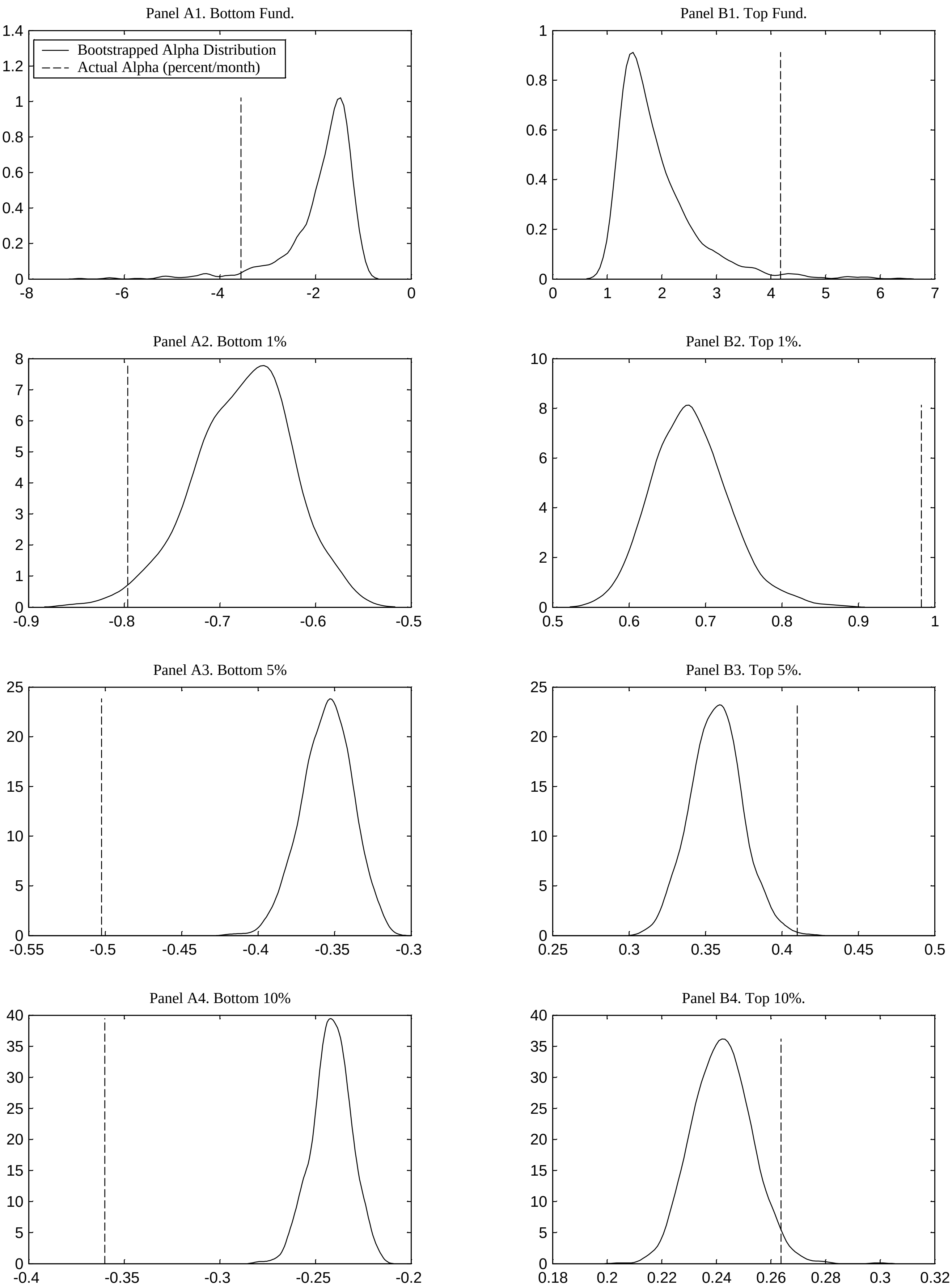
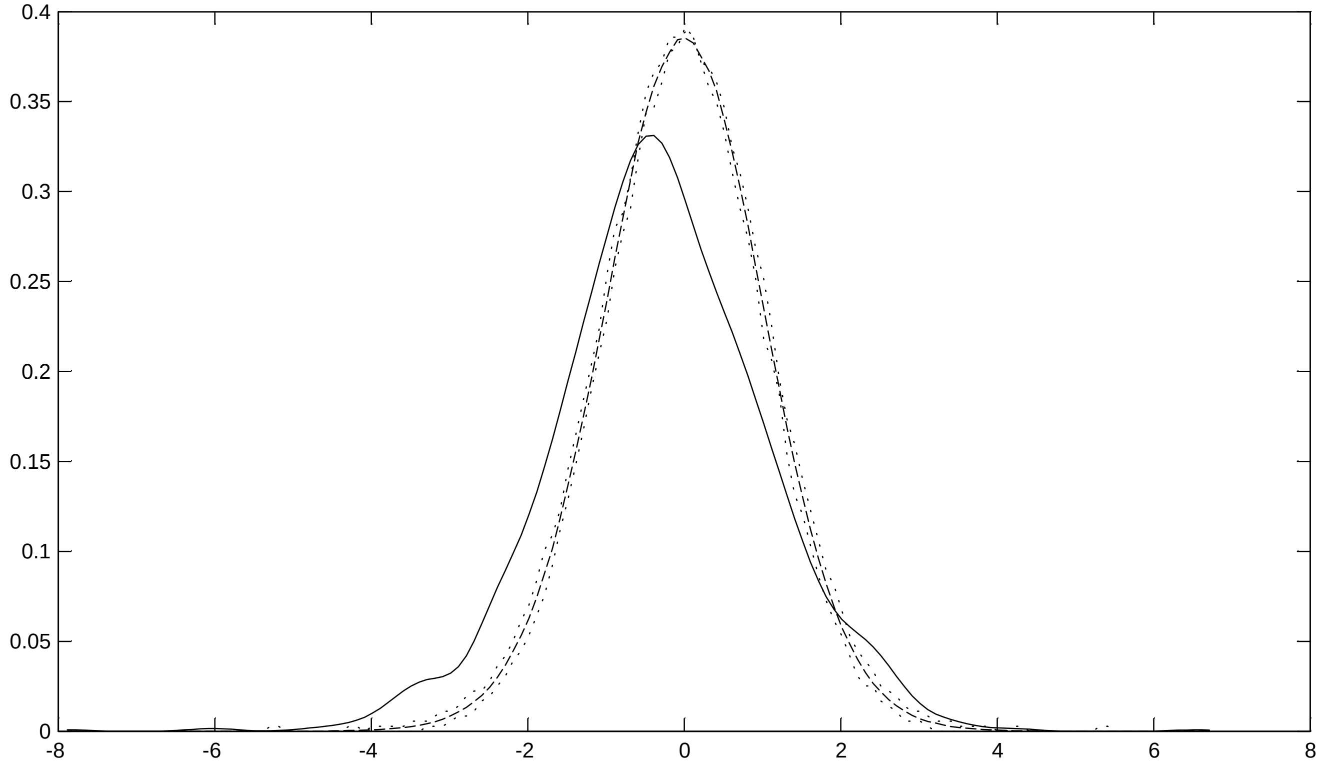


Figure I. This figure plots kernel density estimates of the bootstrapped unconditional four-factor model alpha distribution (solid line) for all U.S. equity funds with at least 60 monthly net return observations during the 1975-2002 period. The x-axis shows the alpha performance measure in percent per month, and the y-axis the kernel density estimate. The actual fund alpha is represented by the dashed vertical line. Panels A1-A4 show marginal funds in the left tail of the distribution. Panels B1-B4 show marginal funds in the right tail of the distribution. "Top 1%" in Panel B2, for example, refers to the marginal alpha at the top 1 percentile of the distribution.

Panel A. PDF of Actual (Solid Line) and Bootstrapped (Dashed Line) Cross-sectional Distribution of t-Statistic of Alphas with S.E. Bands (Dotted Lines).



Panel B. CDF of Actual (Solid Line) and Bootstrapped (Dashed Line) Cross-sectional Distribution of t-Statistic of Alphas with S.E. Bands (Dotted Lines).

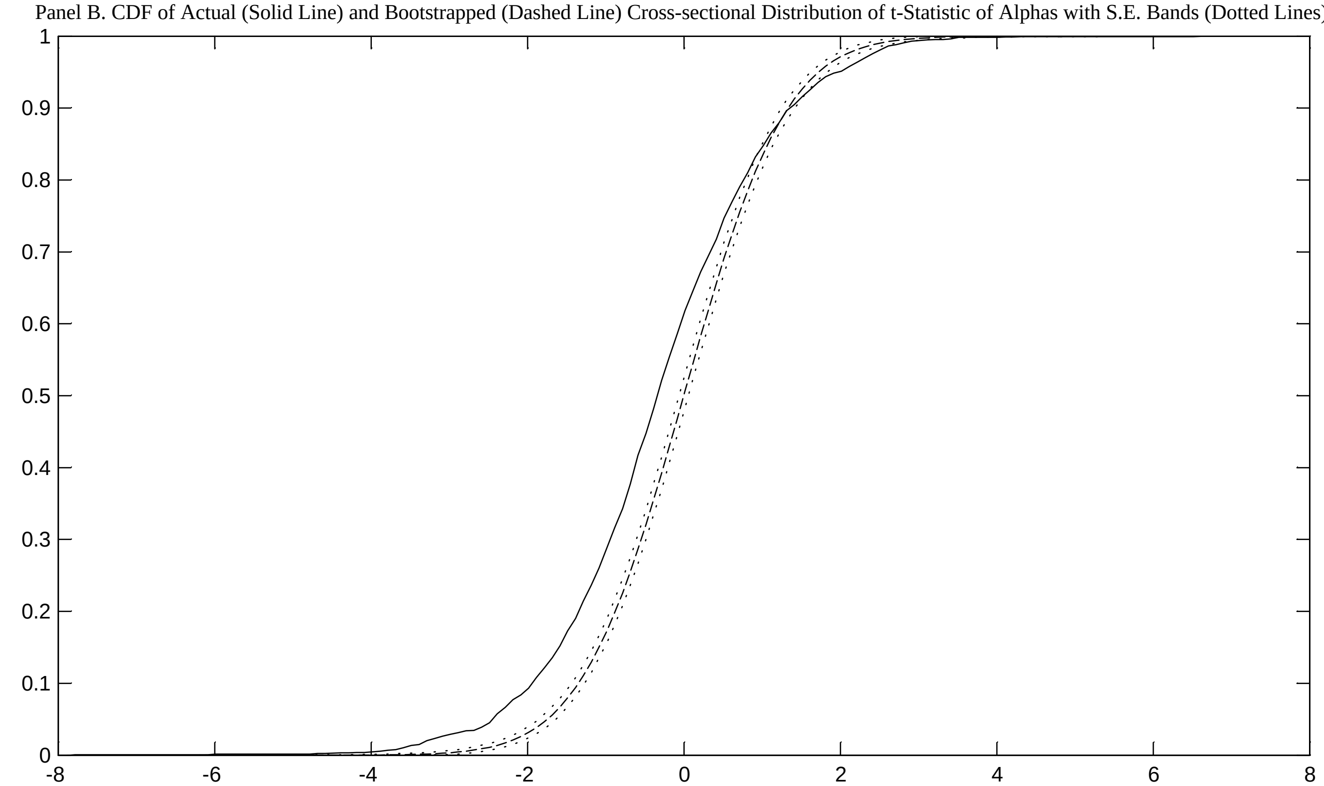
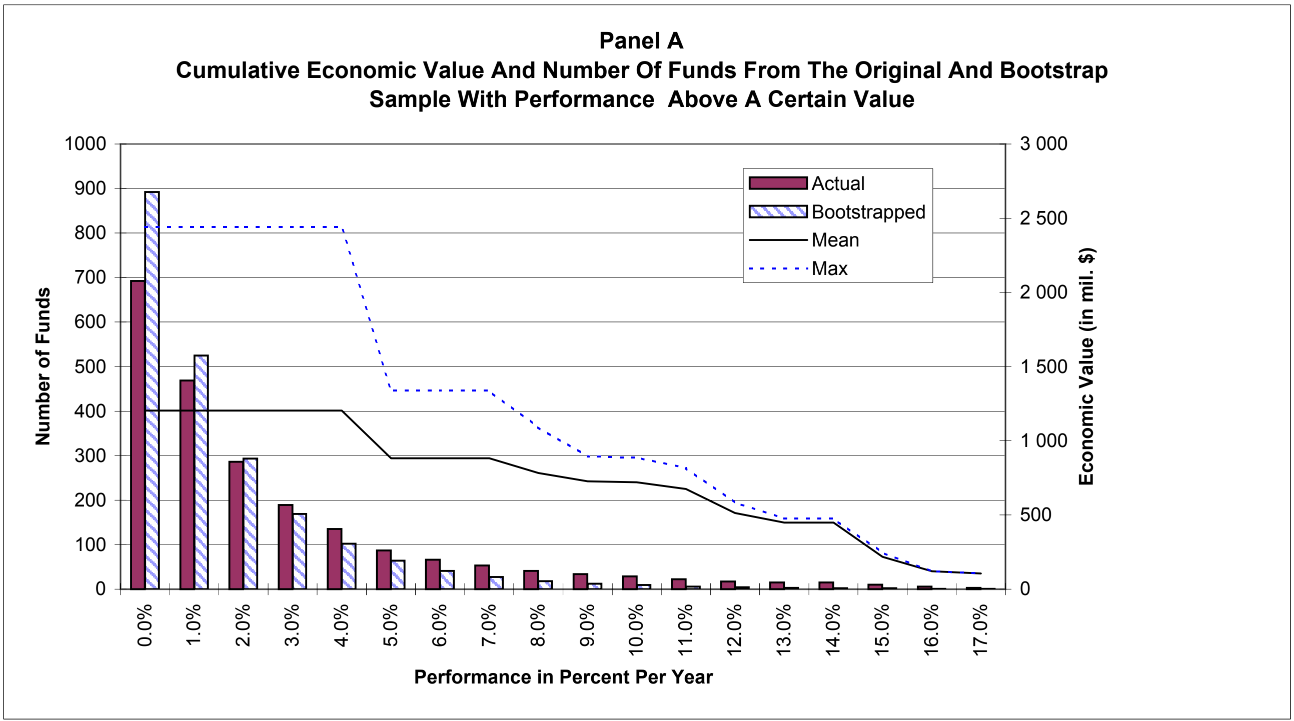


Figure II. This figure plots kernel density estimates of the actual (solid line) and bootstrapped (dashed line) cross-sectional distributions of the t-statistic of mutual fund alphas. Panel A shows the kernel density estimate of the probability density function (PDF) of the distributions, and Panel B the kernel density estimate of the cumulative density function (CDF) of the distributions. The alpha estimates are based on the unconditional four-factor model applied to all U.S. equity funds with at least 60 monthly net return observations during the 1975-2002 period. The standard error bands of the bootstrapped distribution are shown as dotted lines.



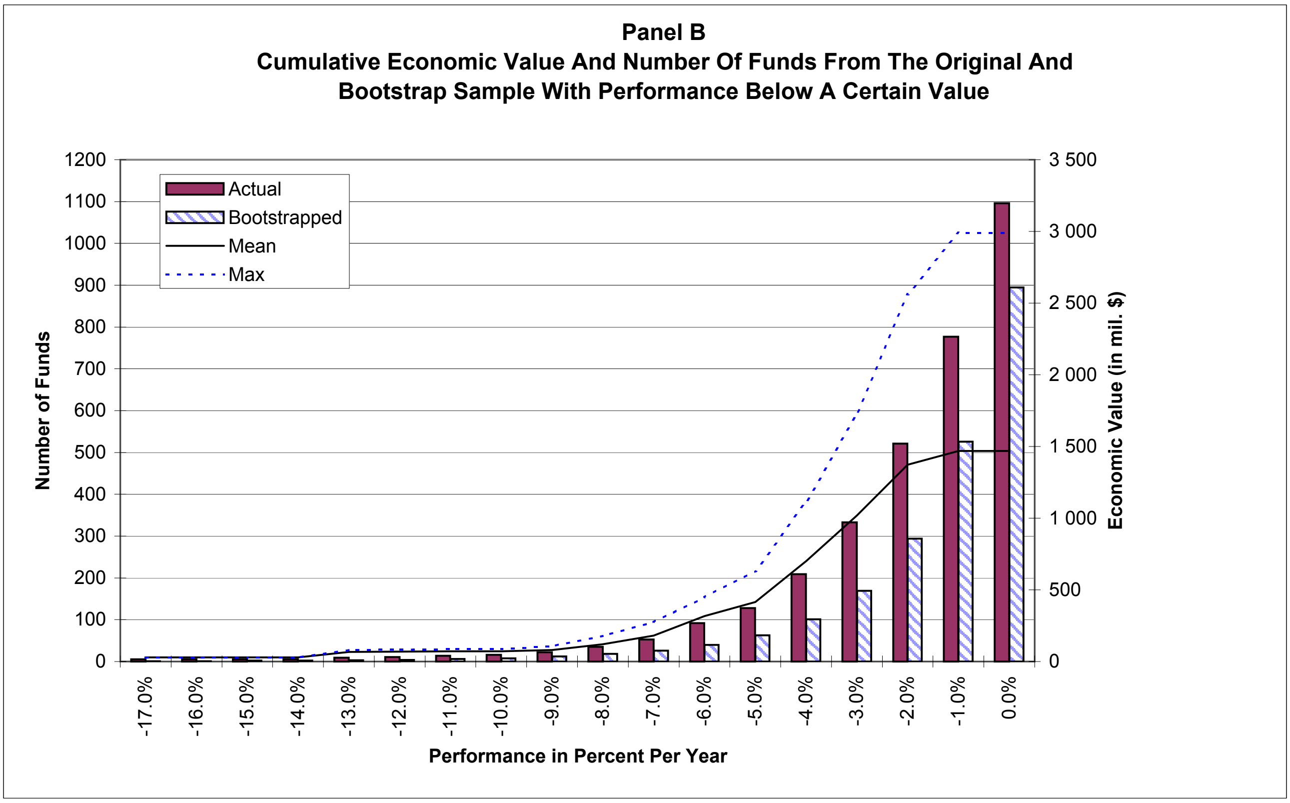


Figure III. This figure presents the number of funds from the original and the bootstrapped cross-sectional distributions (as vertical bars) that surpass various values of performance (in percent per year). The vertical bars in Panel A show the number of actual and bootstrapped funds above a certain level of performance. Panel B shows the number of funds below a certain level of performance. In Panel A, the solid and dashed lines show the cumulative economic value that a hypothetical investor could potentially gain by investing in the difference between the actual and the bootstrapped number of funds in all higher performance brackets. The solid (dashed) line is based on the average (average of a subgroup of the largest funds') total net assets in each performance bracket. In Panel B, the solid and dashed lines show the cumulative value that is potentially lost by the statistically significant underperformance of some funds. The results are based on all U.S. equity funds in existence in our sample between 1975 and 2002 with a minimum of 60 monthly net return observations. The figure is based on the unconditional four-factor model.

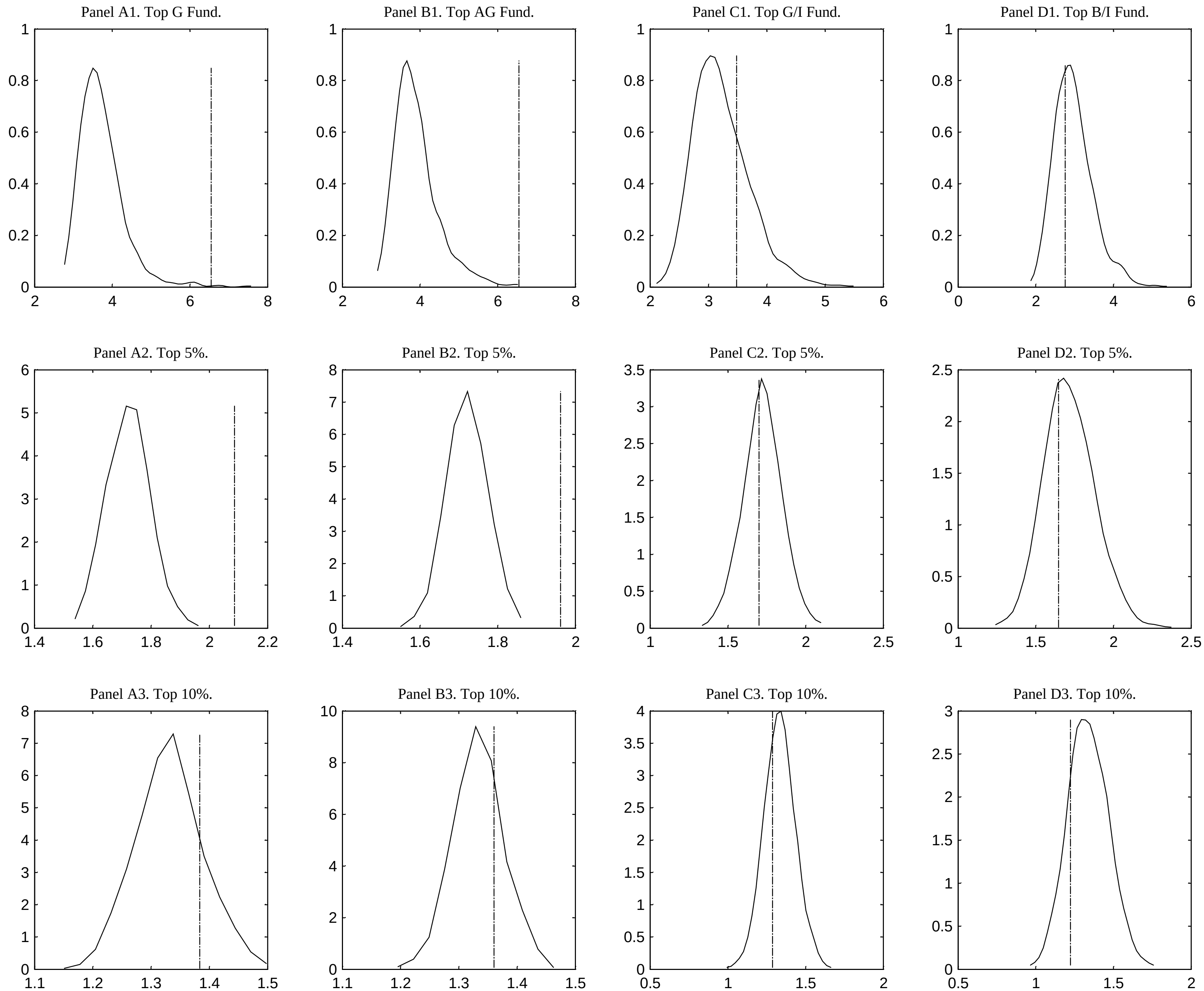


Figure IV. This figure plots kernel density estimates of the bootstrapped distribution of the t-statistic of alpha (solid line) for U.S. equity funds with at least 60 monthly net return observations during the 1975-2002 period. Panels A1-A3 show results for growth funds (G), Panels B1-B3 for aggressive growth funds (AG), Panels C1-C3 for growth and income funds and Panels D1-D3 for balanced or income funds. The x-axis shows the t-statistic, and the y-axis shows the kernel density. The actual t-statistic of alpha is represented by the dashed vertical line. The results are based on the unconditional four-factor model.