THE JOURNAL OF FINANCE • VOL. LXV, NO. 5 • OCTOBER 2010

Luck versus Skill in the Cross-Section of Mutual Fund Returns

EUGENE F. FAMA and KENNETH R. FRENCH*

ABSTRACT

The aggregate portfolio of actively managed U.S. equity mutual funds is close to the market portfolio, but the high costs of active management show up intact as lower returns to investors. Bootstrap simulations suggest that few funds produce benchmark-adjusted expected returns sufficient to cover their costs. If we add back the costs in fund expense ratios, there is evidence of inferior and superior performance (nonzero true α) in the extreme tails of the cross-section of mutual fund α estimates.

THERE IS A CONSTRAINT on the returns to active investing that we call equilibrium accounting. In short (details later), suppose that when returns are measured before costs (fees and other expenses), passive investors get passive returns, that is, they have zero α (abnormal expected return) relative to passive benchmarks. This means active investment must also be a zero sum game—aggregate α is zero before costs. Thus, if some active investors have positive α before costs, it is dollar for dollar at the expense of other active investors. After costs, that is, in terms of net returns to investors, active investment must be a negative sum game. (Sharpe (1991) calls this the arithmetic of active management.)

We examine mutual fund performance from the perspective of equilibrium accounting. For example, at the aggregate level, if the value-weight (VW) portfolio of active funds has a positive α before costs, we can infer that the VW portfolio of active investments outside mutual funds has a negative α. In other words, active mutual funds win at the expense of active investments outside mutual funds. We find that, in fact, the VW portfolio of active funds that invest primarily in U.S. equities is close to the market portfolio, and estimated before expenses, its α relative to common benchmarks is close to zero. Since the VW portfolio of active funds produces α close to zero in gross (pre-expense) returns, α estimated on the net (post-expense) returns realized by investors is negative by about the amount of fund expenses.

The aggregate results imply that if there are active mutual funds with positive true α, they are balanced by active funds with negative α. We test for the

*Fama is at the Booth School of Business, University of Chicago, and French is at the Amos Tuck School of Business Administration, Dartmouth College. We are grateful for the comments of Juhani Linnainmaa, Sunil Wahal, Jerry Zimmerman, and seminar participants at the University of Chicago, the California Institute of Technology, UCLA, and the Meckling Symposium at the University of Rochester. Special thanks to John Cochrane and the journal Editor, Associate Editor, and referees.

existence of such funds. The challenge is to distinguish skill from luck. Given the multitude of funds, many have extreme returns by chance. A common approach to this problem is to test for persistence in fund returns, that is, whether past winners continue to produce high returns and losers continue to underperform (see, e.g., Grinblatt and Titman (1992), Carhart (1997)). Persistence tests have an important weakness. Because they rank funds on short-term past performance, there may be little evidence of persistence because the allocation of funds to winner and loser portfolios is largely based on noise.

We take a different tack. We use long histories of individual fund returns and bootstrap simulations of return histories to infer the existence of superior and inferior funds. Specifically, we compare the actual cross-section of fund α estimates to the results from 10,000 bootstrap simulations of the cross-section. The returns of the funds in a simulation run have the properties of actual fund returns, except we set true α to zero in the return population from which simulation samples are drawn. The simulations thus describe the distribution of α estimates when there is no abnormal performance in fund returns. Comparing the distribution of α estimates from the simulations to the cross-section of α estimates for actual fund returns allows us to draw inferences about the existence of skilled managers.

For fund investors the simulation results are disheartening. When α is estimated on net returns to investors, the cross-section of precision-adjusted α estimates, $t(\alpha)$, suggests that few active funds produce benchmark-adjusted expected returns that cover their costs. Thus, if many managers have sufficient skill to cover costs, they are hidden by the mass of managers with insufficient skill. On a practical level, our results on long-term performance say that true α in net returns to investors is negative for most if not all active funds, including funds with strongly positive α estimates for their entire histories.

Mutual funds look better when returns are measured gross, that is, before the costs included in expense ratios. Comparing the cross-section of $t(\alpha)$ estimates from gross fund returns to the average cross-section from the simulations suggests that there are inferior managers whose actions reduce expected returns, and there are superior managers who enhance expected returns. If we assume that the cross-section of true α has a normal distribution with mean zero and standard deviation σ, then σ around 1.25% per year seems to capture the tails of the cross-section of α estimates for our full sample of actively managed funds.

The estimate of the standard deviation of true α, 1.25% per year, does not imply much skill. It suggests, for example, that fewer than 16% of funds have α greater than 1.25% per year (about 0.10% per month), and only about 2.3% have α greater than 2.50% per year (about 0.21% per month)—before expenses.

The simulation tests have power. If the cross-section of true α for gross fund returns is normal with mean zero, the simulations strongly suggest that the standard deviation of true α is between 0.75% and 1.75% per year. Thus, the simulations rule out values of σ rather close to our estimate, 1.25%. The power traces to the fact that a large cross-section of funds produces precise estimates of the percentiles of $t(\alpha)$ under different assumptions about σ, the standard deviation of true α. This precision allows us to put σ in a rather narrow range.

Readers suggest that our results are consistent with the predictions of Berk and Green (2004). We outline their model in Section II, after the tests on mutual fund aggregates (Section I) and before the bootstrap simulations (Sections III and IV). Our results reject most of their predictions about mutual fund returns. Given the prominence of their model, our contrary evidence seems an important contribution. The paper closest to ours is Kosowski et al. (2006). They run bootstrap simulations that appear to produce stronger evidence of manager skill. We contrast their tests and ours in Section V, after presenting our results. Section VI concludes.

I. The Performance of Aggregate Portfolios of U.S. Equity Mutual Funds

Our mutual fund sample is from the CRSP (Center for Research in Security Prices) database. We include only funds that invest primarily in U.S. common stocks, and we combine, with value weights, different classes of the same fund into a single fund (see French (2008)). To focus better on the performance of active managers, we exclude index funds from all our tests. The CRSP data start in 1962, but we concentrate on the period after 1983. During the period 1962 to 1983 about 15% of the funds on CRSP report only annual returns, and the average annual equal-weight (EW) return for these funds is 5.29% lower than for funds that report monthly returns. As a result, the EW average return on all funds is a nontrivial 0.65% per year lower than the EW return of funds that report monthly returns. Thus, during 1962 to 1983 there is selection bias in tests like ours that use only funds that report monthly returns. After 1983, almost all funds report monthly returns. (Elton, Gruber, and Blake (2001) discuss CRSP data problems for the period before 1984.)

A. The Regression Framework

Our main benchmark for evaluating fund performance is the three-factor model of Fama and French (1993), but we also show results for Carhart's (1997) four-factor model. To measure performance, these models use two variants of the time-series regression

$$R_{it} - R_{ft} = a_i + b_i(R_{Mt} - R_{ft}) + s_i SMB_t + h_i HML_t + m_i MOM_t + e_{it}. \quad (1)$$

In this regression, R_{it} is the return on fund i for month t, R_{ft} is the risk-free rate (the 1-month U.S. Treasury bill rate), R_{Mt} is the market return (the return on a VW portfolio of NYSE, Amex, and NASDAQ stocks), SMB_t and HML_t are the size and value-growth returns of Fama and French (1993), MOM_t is our version of Carhart's (1997) momentum return, a_i is the average return left unexplained by the benchmark model (the estimate of α_i), and e_{it} is the regression residual. The full version of (1) is Carhart's four-factor model, and the regression without MOM_t is the Fama–French three-factor model. The construction of SMB_t and HML_t follows Fama and French (1993). The momentum return,

MOM_t, is defined like HML_t, except that we sort on prior return rather than the book-to-market equity ratio. (See Table I below.)

Regression (1) allows a more precise statement of the constraints of equilibrium accounting. The VW aggregate of the U.S. equity portfolios of all investors is the market portfolio. It has a market slope equal to 1.0 in (1), zero slopes on the other explanatory returns, and a zero intercept—before investment costs. This means that if the VW aggregate portfolio of passive investors also has a zero intercept before costs, the VW aggregate portfolio of active investors must have a zero intercept. Thus, positive and negative intercepts among active investors must balance out—before costs.

There is controversy about whether the average SMB_t, HML_t, and MOM_t returns are rewards for risk or the result of mispricing. For our purposes, there is no need to take a stance on this issue. We can simply interpret SMB_t, HML_t, and MOM_t as diversified passive benchmark returns that capture patterns in average returns during our sample period, whatever the source of the average returns. Abstracting from the variation in returns associated with $R_{Mt} - R_{ft}$, SMB_t, HML_t, and MOM_t then allows us to focus better on the effects of active management (stock picking), which should show up in the three-factor and four-factor intercepts.

From an investment perspective, the slopes on the explanatory returns in (1) describe a diversified portfolio of passive benchmarks (including the risk-free security) that replicates the exposures of the fund on the left to common factors in returns. The regression intercept then measures the average return provided by a fund in excess of the return on a comparable passive portfolio. We interpret a positive expected intercept (true α) as good performance, and a negative expected intercept signals bad performance.[1]

Table I shows summary statistics for the explanatory returns in (1) for January 1984 through September 2006 (henceforth 1984 to 2006), the period used in our tests. The momentum factor (MOM_t) has the highest average return, 0.79% per month ($t = 3.01$), but the average values of the monthly market premium ($R_{Mt} - R_{ft}$) and the value-growth return (HML_t) are also large, 0.64% ($t = 2.42$) and 0.40% ($t = 2.10$), respectively. The size return, SMB_t, has the smallest average value, 0.03% per month ($t = 0.13$).

B. Regression Results for EW and VW Portfolios of Active Funds

Table II shows estimates of regression (1) for the monthly returns of 1984 to 2006 on EW and VW portfolios of the funds in our sample. In the VW portfolio, funds are weighted by assets under management (AUM) at the beginning of

[1] Formal justification for this definition of good and bad performance is provided by Dybvig and Ross (1985). Given a risk-free security, their Theorem 5 implies that if the intercept in (1) is positive, there is a portfolio with positive weight on fund i and the portfolio of the explanatory portfolios on the right of (1) that has a higher Sharpe ratio than the portfolio of the explanatory portfolios. Similarly, if the intercept is negative, there is a portfolio with negative weight on fund i that has a higher Sharpe ratio than the portfolio of the explanatory portfolios.

Table I

Summary Statistics for Monthly Explanatory Returns for the Three-Factor and Four-Factor Models

R_M is the return on a value-weight market portfolio of NYSE, Amex, and NASDAQ stocks, and R_f is the 1-month Treasury bill rate. The construction of SMB_t and HML_t follows Fama and French (1993). At the end of June of each year k, we sort stocks into two size groups. Small includes NYSE, Amex, and NASDAQ stocks with June market capitalization below the NYSE median and Big includes stocks with market cap above the NYSE median. We also sort stocks into three book-to-market equity (B/M) groups, Growth (NYSE, Amex, and NASDAQ stocks in the bottom 30% of NYSE B/M), Neutral (middle 40% of NYSE B/M), and Value (top 30% of NYSE B/M). Book equity is for the fiscal year ending in calendar year $k-1$, and the market cap in B/M is for the end of December of $k-1$. The intersection of the (independent) size and B/M sorts produces six value-weight portfolios, refreshed at the end of June each year. The size return, SMB_t, is the simple average of the month t returns on the three Small stock portfolios minus the average of the returns on the three Big stock portfolios. The value-growth return, HML_t, is the simple average of the returns on the two Value portfolios minus the average of the returns on the two Growth portfolios. The momentum return, MOM_t, is defined like HML_t, except that we sort on prior return rather than B/M and the momentum sort is refreshed monthly rather than annually. At the end of each month $t-1$ we sort NYSE stocks on the average of the 11 months of returns to the end of month $t-2$. (Dropping the return for month $t-1$ is common in the momentum literature.) We use the 30^{th} and 70^{th} NYSE percentiles to assign NYSE, Amex, and NASDAQ stocks to Low, Medium, and High momentum groups. The intersection of the size sort for the most recent June and the independent momentum sort produces six value-weight portfolios, refreshed monthly. The momentum return, MOM_t, is the simple average of the month t returns on the two High momentum portfolios minus the average of the returns on the two Low momentum portfolios. The table shows the average monthly return, the standard deviation of monthly returns, and the t-statistic for the average monthly return. The period is January 1984 through September 2006.

	Average Return				Standard Deviation				t-statistic			
	R_M-R_f	*SMB*	*HML*	*MOM*	R_M-R_f	*SMB*	*HML*	*MOM*	R_M-R_f	*SMB*	*HML*	*MOM*
1984–2006	0.64	0.03	0.40	0.79	4.36	3.38	3.17	4.35	2.42	0.13	2.10	3.01

Table II

Intercepts and Slopes in Variants of Regression (1) for Equal-Weight (EW) and Value-Weight (VW) Portfolios of Actively Managed Mutual Funds

The table shows the annualized intercepts ($12 * a$) and t-statistics for the intercepts ($t(Coef)$) for the CAPM, three-factor, and four-factor versions of regression (1) estimated on equal-weight (EW) and value-weight (VW) net and gross returns on the portfolios of actively managed mutual funds in our sample. The table also shows the regression slopes (b, s, h, and m, for $R_M - R_f$, SMB, HML, and MOM, respectively), t-statistics for the slopes, and the regression R^2, all of which are the same to two decimals for gross and net returns. For the market slope, $t(Coef)$ tests whether b is different from 1.0. Net returns are those received by investors. Gross returns are net returns plus $1/12^{\text{th}}$ of a fund's expense ratio for the year. When a fund's expense ratio for a year is missing, we assume it is the same as other actively managed funds with similar assets under management (AUM). The period is January 1984 through September 2006. On average there are 1,308 funds and their average AUM is $648.0 million.

	$12 * a$						
	Net	Gross	b	s	h	m	R^2
EW Returns							
Coef	−1.11	0.18	1.01				0.96
$t(Coef)$	−1.80	0.31	1.12				
Coef	−0.93	0.36	0.98	0.18	−0.00		0.98
$t(Coef)$	−2.13	0.85	−1.78	16.09	−0.24		
Coef	−0.92	0.39	0.98	0.18	−0.00	−0.00	0.98
$t(Coef)$	−2.05	0.90	−1.78	16.01	−0.25	−0.14	
VW Returns							
Coef	−1.13	−0.18	0.99				0.99
$t(Coef)$	−3.03	−0.49	−2.10				
Coef	−0.81	0.13	0.96	0.07	−0.03		0.99
$t(Coef)$	−2.50	0.40	−5.42	7.96	−3.22		
Coef	−1.00	−0.05	0.97	0.07	−0.03	0.02	0.99
$t(Coef)$	−3.02	−0.15	−5.03	7.78	−3.03	2.60	

each month. The EW portfolio weights funds equally each month. The intercepts in (1) for EW fund returns tell us whether funds on average produce returns different from those implied by their exposures to common factors in returns, whereas VW returns tell us about the fate of aggregate wealth invested in funds. Table II shows estimates of (1) for fund returns measured gross and net of fund expenses. Net returns are those received by investors. Monthly gross returns are net returns plus $1/12^{\text{th}}$ of a fund's expense ratio for the year.

The market slopes in Table II are close to 1.0, which is not surprising since our sample is funds that invest primarily in U.S. stocks. The HML_t and MOM_t slopes are close to zero. Thus, in aggregate, active funds show little exposure to the value-growth and momentum factors. The EW portfolio of funds produces a larger SMB_t slope (0.18) than the VW portfolio (0.07). We infer that smaller funds are more likely to invest in small stocks, but total dollars invested in active funds (captured by VW returns) show little tilt toward small stocks.

The intercepts in the estimates of (1) summarize the average performance of funds (EW returns) and the performance of aggregate wealth invested in funds (VW returns) relative to passive benchmarks. In terms of net returns to investors, performance is poor. The three-factor and four-factor (annualized) intercepts for EW and VW net returns are negative, ranging from −0.81% to −1.00% per year, with t-statistics from −2.05 to −3.02. These results are in line with previous work (e.g., Jensen (1968), Malkiel (1995), Gruber (1996)).

The intercepts in (1) for EW and VW net fund returns tell us whether on average active managers have sufficient skill to generate returns that cover the costs funds impose on investors. Gross returns come closer to testing whether managers have any skill. For EW gross fund returns, the three-factor and four-factor intercepts for 1984 to 2006 are positive, 0.36% and 0.39% per year, but they are only 0.85 and 0.90 standard errors from zero. The intercepts in (1) for VW gross returns are quite close to zero, 0.13% per year ($t = 0.40$) for the three-factor version of (1), and −0.05% per year ($t = -0.15$) for the four-factor model.

Table II also shows estimates of the CAPM version of (1), in which $R_{Mt} - R_{ft}$ is the only explanatory return. The annualized CAPM intercept for VW gross fund returns for 1984 to 2006, −0.18% per year ($t = -0.49$), is again close to zero and similar to the estimates for the three-factor and four-factor models. It is not surprising that the intercepts of the three models are so similar (−0.18%, 0.13%, and −0.05% per year) since VW fund returns produce slopes close to zero for the non-market explanatory returns in (1).

We can offer an equilibrium accounting perspective on the results in Table II. When we add back the costs in expense ratios, α estimates for VW gross fund returns are close to zero. Thus, before expenses, there is no evidence that total wealth invested in active funds gets any benefits or suffers any losses from active management. VW fund returns also show little exposure to the size, value, and momentum returns, and the market return alone explains 99% of the variance of the monthly VW fund return. Together these facts say that during 1984 to 2006, active mutual funds in aggregate hold a portfolio that, before expenses, mimics market portfolio returns. The return to investors, however, is reduced by the high expense ratios of active funds. These results echo equilibrium accounting, but for a subset of investment managers where the implications of equilibrium accounting for aggregate investor returns need not hold.

C. Measurement Issues in the Tests on Gross Returns

The benchmark explanatory returns in (1) are before all costs. This is appropriate in tests on net fund returns where the issue addressed is whether managers have sufficient skill to produce expected returns that cover their costs. Gross returns pose more difficult measurement issues.

The issue in the tests on gross fund returns is whether managers have skill that causes expected returns to differ from those of comparable passive benchmarks. For this purpose, one would like fund returns measured before all

costs and non-return revenues. This would put funds on the same pure return basis as the benchmark explanatory returns, so the regressions could focus on manager skill. Our gross fund returns are before the costs in expense ratios (including management fees), but they are net of other costs, primarily trading costs, and they include the typically small revenues from securities lending.

We could attempt to add trading costs to our estimates of gross fund returns. Funds do not report trading costs, however, and estimates are subject to large errors. For example, trading costs are likely to vary across funds because of differences in style tilts, trading skill, and the extent to which a fund demands immediacy in trade execution. Trading costs also vary through time. Our view is that estimates of trading costs for individual funds, especially actively managed funds, are fraught with error and potential bias, and are likely to be misleading. We prefer to stay with our simple definition of gross returns (net returns plus the costs in expense ratios), with periodic qualifications to our inferences.

An alternative approach (suggested by a referee) is to put the passive benchmarks produced by combining the explanatory returns in (1) in the same units as the gross fund returns on the left of (1). This involves taking account of the costs not covered in expense ratios that would be borne by an efficiently managed passive benchmark with the same style tilts as the fund whose gross returns are to be explained. Appendix A discusses this approach in detail. The bottom line is that for efficiently managed passive funds, the costs missed in expense ratios are close to zero. Thus, adjusting the benchmarks produced by (1) for estimates of these costs is unnecessary.

This does not mean our tests on gross fund returns capture the pure effects of skill. Though it appears that all substantial costs incurred by efficiently managed passive funds are in their expense ratios, this is less likely to be true for actively managed funds. The typical active fund trades more than the typical passive fund, and active funds are likely to demand immediacy in trading that pushes up costs. Our tests on gross returns thus produce α estimates that capture skill, less whatever net costs (costs minus non-return revenues) are missed by expense ratios. Equivalently, the tests say that a fund's management has skill only if it is sufficient to cover the missing costs (primarily trading costs). This seems like a reasonable definition of skill since an efficiently managed passive fund can apparently avoid these costs. More important, this is the definition of skill we can accurately test, given the unavoidable absence of accurate trading cost estimates for active funds.

The fact that our gross fund returns are net of the costs missed in expense ratios, however, does affect the inferences about equilibrium accounting we can draw from the aggregate results in Table II. Since the α estimates for VW gross fund returns in Table II are close to zero, they suggest that in aggregate funds show sufficient skill to produce expected returns that cover some or all of the costs missed in expense ratios. If this is the correct inference (precision is an issue), equilibrium accounting then says that the costs recovered by funds are matched by equivalent losses on investments outside mutual funds.

II. Berk and Green (2004)

Readers contend that our results (Table II and below) are consistent with Berk and Green (2004). Their model is attractive theory, but our results reject most of its predictions about mutual fund returns.

In their world, a fund is endowed with a permanent α, before costs, but it faces costs that are an increasing convex function of AUM. Investors use returns to update estimates of α. A fund with a positive expected α before costs attracts inflows until AUM reaches the point where expected α, net of costs, is zero. Outflows drive out funds with negative expected α. In equilibrium, all active funds (and thus funds in aggregate) have positive expected α before costs and zero expected α net of costs.

Our evidence that the aggregate portfolio of mutual funds has negative α net of costs contradicts the predictions of Berk and Green (2004). The results below on the net returns of individual funds also reject their prediction that all active managers have zero α net of costs. In fact, our results say that for most if not all funds, true α in net returns is negative.

Finally, equilibrium accounting poses a theoretical problem for Berk and Green (2004). Their model focuses on rational investors who optimally choose among passive and active alternatives. In aggregate, their investors have positive α before costs and zero α after costs. Equilibrium accounting, however, says that in aggregate investors have zero α before costs and negative α after costs.

III. Bootstrap Simulations

Table II says that, on average, active mutual funds do not produce gross returns above (or below) those of passive benchmarks. This may just mean that managers with skill that allows them to outperform the benchmarks are balanced by inferior managers who underperform. We turn now to simulations that use individual fund returns to infer the existence of superior and inferior managers.

A. Setup

To lessen the effects of "incubation bias" (see below), we limit the tests to funds that reach 5 million 2006 dollars in AUM. Since the AUM minimum is in 2006 dollars, we include a fund in 1984, for example, if it has more than about \$2.5 million in AUM in 1984. Once a fund passes the AUM minimum, it is included in all subsequent tests, so this requirement does not create selection bias. We also show results for funds after they pass \$250 million and \$1 billion. Since we estimate benchmark regressions for each fund, we limit the tests to funds that have at least 8 months of returns after they pass an AUM bound, so there is a bit of survival bias. To avoid having lots of new funds with short return histories, we only use funds that appear on CRSP at least 5 years before the end of our sample period.

Fund management companies commonly provide seed money to new funds to develop a return history. Incubation bias arises because funds typically

open to the public—and their pre-release returns are included in mutual fund databases—only if the returns turn out to be attractive. The $5 million AUM bound for admission to the tests alleviates this bias since AUM is likely to be low during the pre-release period.

Evans (2010) suggests that incubation bias can be minimized by using returns only after funds receive a ticker symbol from NASDAQ, which typically means they are available to the public. Systematic data on ticker symbol start dates are available only after 1998. We have replicated our tests for 1999 to 2006 using CRSP start dates for new funds (as in our reported results) and then using NASDAQ ticker dates (from Evans). Switching to ticker dates has almost no effect on aggregate fund returns (as in Table II), and has only trivial effects on the cross-section of $t(\alpha)$ estimates for funds (as in Table III below). We conclude that incubation bias is probably unimportant in our results for 1984 to 2006.

Our goal is to draw inferences about the cross-section of true α for active funds, specifically, whether the cross-section of α estimates suggests a world where true α is zero for all funds or whether there is nonzero true α, especially in the tails of the cross-section of α estimates. We are interested in answering this question for 12 different cross-sections of α estimates—for gross and net returns, for the three-factor and four-factor benchmarks, and for the three AUM samples. Thus, we use regression (1) to estimate each fund's three-factor or four-factor α for gross or net returns for the part of 1984 to 2006 after the fund passes each AUM bound.

The tests for nonzero true α in actual fund returns use bootstrap simulations on returns that have the properties of fund returns, except that true α is set to zero for every fund. To set α to zero, we subtract a fund's α estimate from its monthly returns. For example, to compute three-factor benchmark-adjusted gross returns for a fund in the $5 million group, we subtract its three-factor α estimated from monthly gross returns for the part of 1984 to 2006 that the fund is in the $5 million group from the fund's monthly gross returns for that period. We calculate benchmark-adjusted returns for the three-factor and four-factor models, for gross and net returns, and for the three AUM bounds. The result is 12 populations of benchmark-adjusted (zero-α) returns. (CAPM simulation results are in Appendix B.)

A simulation run is a random sample (with replacement) of 273 months, drawn from the 273 calendar months of January 1984 to September 2006. For each of the 12 sets of benchmark-adjusted returns, we estimate, fund by fund, the relevant benchmark model on the simulation draw of months of adjusted returns, dropping funds that are in the simulation run for less than 8 months. Each run thus produces 12 cross-sections of α estimates using the same random sample of months from 12 populations of adjusted (zero-α) fund returns.

We do 10,000 simulation runs to produce 12 distributions of t-statistics, $t(\alpha)$, for a world in which true α is zero. We focus on $t(\alpha)$, rather than estimates of α, to control for differences in precision due to differences in residual variance and in the number of months funds are in a simulation run.

Note that setting true α equal to zero builds different assumptions about skill into the tests on gross and net fund returns. For net returns, setting true α to zero leads to a world where every manager has sufficient skill to generate expected returns that cover all costs. In contrast, setting true α to zero in gross returns implies a world where every fund manager has just enough skill to produce expected returns that cover the costs missed in expense ratios.

Our simulation approach has an important advantage. Because a simulation run is the same random sample of months for all funds, the simulations capture the cross-correlation of fund returns and its effects on the distribution of $t(\alpha)$ estimates. Since we jointly sample fund and explanatory returns, we also capture any correlated heteroskedasticity of the explanatory returns and disturbances of a benchmark model. We shall see that these details of our approach are important for inferences about true α in actual fund returns.

Defining a simulation run as the same random sample of months for all funds also has a cost. If a fund is not in the tests for the entire 1984 to 2006 period, it is likely to show up in a simulation run for more or less than the number of months it is in our sample. This is not serious. We focus on $t(\alpha)$, and the distribution of $t(\alpha)$ estimates depends on the number of months funds are in a simulation run through a degrees of freedom effect. The distributions of $t(\alpha)$ estimates for funds that are oversampled in a simulation run have more degrees of freedom (and thinner extreme tails) than the distributions of $t(\alpha)$ for the actual returns of the funds. Within a simulation run, however, oversampling of some funds should roughly offset undersampling of others, so a simulation run should produce a representative sample of $t(\alpha)$ estimates for simulated returns that have the properties of actual fund returns, except that true α is zero for every fund. Oversampling and undersampling of fund returns in a simulation run should also about balance out in the 10,000 runs used in our inferences.

A qualification of this conclusion is in order. In a simulation run, as in the tests on actual returns, we discard funds that have less than 8 months of returns. This means we end up with a bit more oversampling of fund returns. As a result, the distributions of $t(\alpha)$ estimates in the simulations tend to have more degrees of freedom (and thinner tails) than the estimates for actual fund returns. This means our tests are a bit biased toward finding false evidence of performance in the tails of $t(\alpha)$ estimates for actual fund returns.

There are two additional caveats. (i) Random sampling of months in a simulation run preserves the cross-correlation of fund returns, but we lose any effects of autocorrelation. The literature on autocorrelation of stock returns (e.g., Fama (1965)) suggests that this is a minor problem. (ii) Because we randomly sample months, we also lose any effects of variation through time in the regression slopes in (1). (The issues posed by time-varying slopes are discussed by Ferson and Schadt (1996).) Capturing time variation in the regression slopes poses thorny problems, and we leave this potentially important issue for future research.

To develop perspective on the simulations, we first compare, in qualitative terms, the percentiles of the cross-section of $t(\alpha)$ estimates from actual fund returns and the average values of the percentiles from the simulations. We then

turn to likelihood statements about whether the cross-section of $t(\alpha)$ estimates for actual fund returns points to the existence of skill.

B. First Impressions

When we estimate a benchmark model on the returns of each fund in an AUM group, we get a cross-section of $t(\alpha)$ estimates that can be ordered into a cumulative distribution function (CDF) of $t(\alpha)$ estimates for actual fund returns. A simulation run for the same combination of benchmark model and AUM group also produces a cross-section of $t(\alpha)$ estimates and its CDF for a world in which true α is zero. In our initial examination of the simulations we compare (i) the values of $t(\alpha)$ at selected percentiles of the CDF of the $t(\alpha)$ estimates from actual fund returns and (ii) the averages across the 10,000 simulation runs of the $t(\alpha)$ estimates at the same percentiles. For example, the first percentile of three-factor $t(\alpha)$ estimates for the net returns of funds in the \$5 million AUM group is -3.87, versus an average first percentile of -2.50 from the 10,000 three-factor simulation runs for the net returns of funds in this group (Table III).

For each combination of gross or net returns, AUM group, and benchmark model, Table III shows the CDF of $t(\alpha)$ estimates for actual returns and the average of the 10,000 simulation CDFs. The average simulation CDFs are similar for gross and net returns and for the two benchmark models. This is not surprising since true α is always zero in the simulations. The dispersion of the average simulation CDFs decreases from lower to higher AUM groups. This is at least in part a degrees of freedom effect; on average, funds in lower AUM groups have shorter sample periods.

B.1. Net Returns

The Berk and Green (2004) prediction that most fund managers have sufficient skill to cover their costs fares poorly in Table III. The left tail percentiles of the $t(\alpha)$ estimates from actual net fund returns are far below the corresponding average values from the simulations. For example, the 10^{th} percentiles of the actual $t(\alpha)$ estimates, -2.34, -2.37, and -2.53 for the \$5 million, \$250 million, and \$1 billion groups, are much more extreme than the average estimates from the simulation, -1.32, -1.31, and -1.30. The right tails of the $t(\alpha)$ estimates also do not suggest widespread skill sufficient to cover costs. In the tests that use the three-factor model, the $t(\alpha)$ estimates from the actual net returns of funds in the \$5 million group are below the average values from the simulations for all percentiles below the 98^{th}. For the \$1 billion group, only the 99^{th} percentile of three-factor $t(\alpha)$ for actual net fund returns is above the average simulation 99^{th} percentile, and then only slightly. For the \$250 million group, the percentiles of three-factor $t(\alpha)$ for actual net fund returns are all below the averages from the simulations. Figure 1 shows the actual and average simulated CDFs for the \$5 million AUM group.

Table III

Percentiles of $t(\alpha)$ Estimates for Actual and Simulated Fund Returns: January 1984 to September 2006

The table shows values of $t(\alpha)$ at selected percentiles (Pct) of the distribution of $t(\alpha)$ estimates for actual (Act) net and gross fund returns. The table also shows the percent of the 10,000 simulation runs that produce lower values of $t(\alpha)$ at the selected percentiles than those observed for actual fund returns (% < Act). Sim is the average value of $t(\alpha)$ at the selected percentiles from the simulations. The period is January 1984 to September 2006 and results are shown for the three- and four-factor models for the \$5 million, \$250 million, and \$1 billion AUM fund groups. There are 3,156 funds in the \$5 million group, 1,422 in the \$250 million group, and 660 in the \$1 billion group.

	5 Million			250 Million			1 Billion		
Pct	Sim	Act	%<Act	Sim	Act	%<Act	Sim	Act	%<Act
3-Factor Net Returns									
1	−2.50	−3.87	0.08	−2.45	−3.87	0.10	−2.39	−4.39	0.01
2	−2.17	−3.42	0.06	−2.13	−3.38	0.13	−2.09	−3.55	0.09
3	−1.97	−3.15	0.07	−1.94	−3.15	0.12	−1.91	−3.36	0.07
4	−1.83	−2.99	0.06	−1.80	−3.04	0.10	−1.78	−3.16	0.07
5	−1.71	−2.84	0.08	−1.69	−2.91	0.10	−1.67	−2.99	0.10
10	−1.32	−2.34	0.05	−1.31	−2.37	0.10	−1.30	−2.53	0.08
20	−0.87	−1.74	0.03	−0.86	−1.87	0.04	−0.86	−1.98	0.03
30	−0.54	−1.27	0.06	−0.54	−1.41	0.06	−0.54	−1.59	0.02
40	−0.26	−0.92	0.05	−0.27	−1.03	0.07	−0.27	−1.19	0.02
50	−0.01	−0.62	0.04	−0.01	−0.71	0.06	−0.01	−0.82	0.03
60	0.25	−0.29	0.11	0.25	−0.39	0.19	0.24	−0.51	0.05
70	0.52	0.08	0.51	0.52	−0.08	0.25	0.52	−0.20	0.08
80	0.85	0.50	3.20	0.84	0.37	1.68	0.84	0.25	0.85
90	1.30	1.01	8.17	1.29	0.89	5.19	1.28	0.82	4.81
95	1.68	1.54	30.55	1.66	1.36	14.17	1.64	1.34	17.73
96	1.80	1.71	40.06	1.76	1.49	17.24	1.74	1.52	26.33
97	1.94	1.91	49.35	1.90	1.69	25.92	1.87	1.79	42.86
98	2.13	2.17	58.70	2.08	1.90	30.43	2.04	2.02	50.07
99	2.45	2.47	57.42	2.36	2.29	43.92	2.31	2.40	63.11
4-Factor Net Returns									
1	−2.55	−3.94	0.04	−2.47	−3.94	0.08	−2.40	−4.22	0.01
2	−2.20	−3.43	0.04	−2.14	−3.43	0.09	−2.09	−3.48	0.08
3	−2.00	−3.08	0.13	−1.95	−3.07	0.25	−1.91	−3.11	0.23
4	−1.85	−2.88	0.13	−1.80	−2.88	0.22	−1.77	−2.95	0.21
5	−1.73	−2.74	0.12	−1.69	−2.78	0.18	−1.66	−2.86	0.14
10	−1.33	−2.23	0.14	−1.30	−2.34	0.14	−1.29	−2.48	0.07
20	−0.86	−1.67	0.10	−0.85	−1.80	0.11	−0.84	−1.96	0.05
30	−0.53	−1.25	0.12	−0.52	−1.39	0.10	−0.52	−1.54	0.04
40	−0.25	−0.88	0.21	−0.25	−1.04	0.14	−0.25	−1.23	0.05
50	0.01	−0.60	0.18	0.01	−0.76	0.11	0.01	−0.87	0.07
60	0.26	−0.29	0.25	0.27	−0.42	0.29	0.26	−0.49	0.19
70	0.54	0.02	0.37	0.54	−0.13	0.24	0.54	−0.18	0.24
80	0.87	0.44	1.76	0.86	0.27	0.72	0.86	0.17	0.45
90	1.33	1.04	10.62	1.31	0.86	4.40	1.30	0.86	7.07
95	1.72	1.53	23.82	1.69	1.37	14.35	1.67	1.31	14.13
96	1.84	1.67	28.21	1.80	1.51	18.23	1.78	1.45	17.16
97	1.99	1.84	31.30	1.94	1.65	18.62	1.91	1.57	17.05
98	2.19	2.09	39.12	2.12	1.79	15.57	2.08	1.76	18.86
99	2.52	2.40	36.96	2.42	2.22	29.88	2.36	2.26	42.00

(*continued*)

Table III—*Continued*

	5 Million			250 Million			1 Billion		
Pct	Sim	Act	%<Act	Sim	Act	%<Act	Sim	Act	%<Act
3-Factor Gross Returns									
1	−2.49	−3.07	4.11	−2.45	−3.16	3.16	−2.39	−3.29	1.88
2	−2.17	−2.68	4.79	−2.13	−2.67	6.01	−2.09	−2.70	5.64
3	−1.97	−2.48	4.20	−1.94	−2.51	4.47	−1.91	−2.51	5.12
4	−1.83	−2.31	4.41	−1.80	−2.35	4.68	−1.78	−2.33	5.77
5	−1.71	−2.19	4.15	−1.69	−2.18	5.99	−1.67	−2.18	6.52
10	−1.32	−1.72	5.75	−1.31	−1.77	5.94	−1.30	−1.86	4.15
20	−0.87	−1.10	13.61	−0.86	−1.24	7.18	−0.86	−1.43	2.52
30	−0.54	−0.71	20.03	−0.54	−0.79	15.10	−0.54	−1.00	4.28
40	−0.26	−0.36	29.74	−0.27	−0.43	23.84	−0.27	−0.59	10.25
50	−0.01	−0.06	38.87	−0.01	−0.15	26.28	−0.01	−0.28	13.48
60	0.25	0.28	56.05	0.25	0.14	31.47	0.24	0.05	21.21
70	0.52	0.63	71.81	0.52	0.48	43.62	0.52	0.35	26.70
80	0.85	1.06	85.21	0.84	0.88	58.14	0.84	0.79	44.31
90	1.30	1.59	90.01	1.29	1.41	69.39	1.28	1.34	60.63
95	1.68	2.04	92.10	1.66	1.81	72.89	1.64	1.78	70.37
96	1.80	2.20	93.73	1.76	1.93	73.44	1.74	1.96	77.00
97	1.94	2.44	95.97	1.90	2.19	84.36	1.87	2.22	85.47
98	2.13	2.72	97.29	2.08	2.47	89.30	2.04	2.37	83.72
99	2.45	3.03	96.66	2.36	2.83	90.95	2.31	2.97	94.63
4-Factor Gross Returns									
1	−2.55	−3.06	5.49	−2.47	−3.02	6.72	−2.40	−3.34	1.67
2	−2.20	−2.71	4.99	−2.14	−2.63	7.84	−2.09	−2.48	14.14
3	−2.00	−2.46	5.46	−1.95	−2.43	7.33	−1.91	−2.40	8.43
4	−1.85	−2.27	6.39	−1.80	−2.33	5.73	−1.77	−2.25	8.66
5	−1.73	−2.11	7.71	−1.69	−2.12	8.62	−1.66	−2.11	9.52
10	−1.33	−1.62	12.27	−1.30	−1.71	8.63	−1.29	−1.85	4.69
20	−0.86	−1.09	16.23	−0.85	−1.19	11.13	−0.84	−1.34	5.29
30	−0.53	−0.65	28.46	−0.52	−0.75	19.76	−0.52	−0.92	8.75
40	−0.25	−0.33	35.43	−0.25	−0.45	22.31	−0.25	−0.57	12.54
50	0.01	−0.02	44.53	0.01	−0.16	26.29	0.01	−0.29	14.40
60	0.26	0.28	53.17	0.27	0.09	25.86	0.26	0.05	22.48
70	0.54	0.62	64.90	0.54	0.48	43.11	0.54	0.36	27.78
80	0.87	0.98	70.19	0.86	0.85	50.07	0.86	0.82	47.07
90	1.33	1.58	84.76	1.31	1.36	58.66	1.30	1.41	65.72
95	1.72	2.05	88.77	1.69	1.87	73.81	1.67	1.83	70.55
96	1.84	2.21	91.03	1.80	2.01	76.27	1.78	1.95	70.91
97	1.99	2.39	92.01	1.94	2.21	81.22	1.91	2.04	66.61
98	2.19	2.58	91.20	2.12	2.43	83.35	2.08	2.30	74.26
99	2.52	3.01	93.44	2.42	2.72	81.41	2.36	2.57	71.98

Evidence of skill sufficient to cover costs is even weaker with an adjustment for momentum exposure. In the tests that use the four-factor model, the percentiles of the $t(\alpha)$ estimates for actual net fund returns are always below the average values from the simulations. In other words, the averages of the

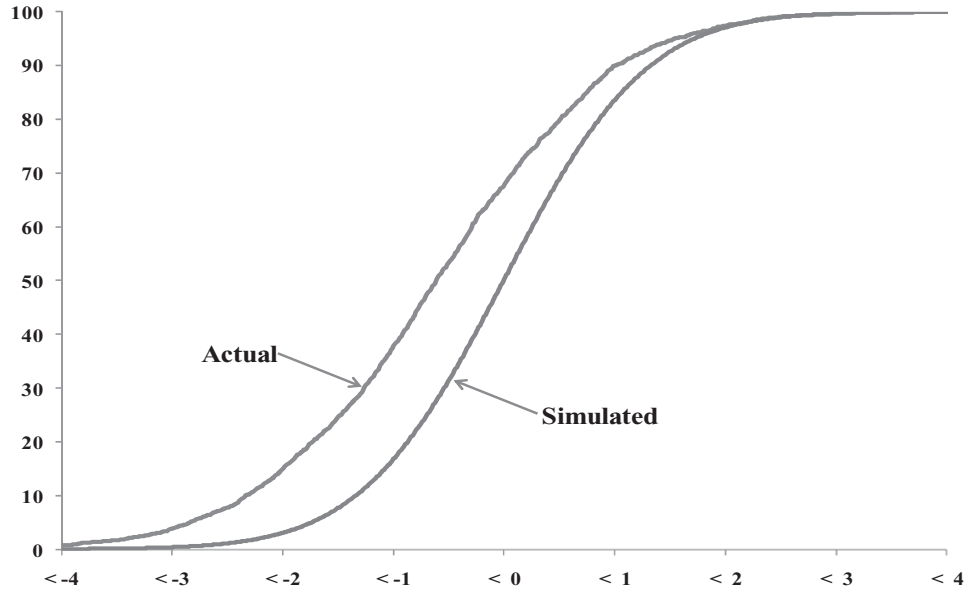

Figure 1. Simulated and actual cumulative density function of three-factor $t(\alpha)$ for net returns, 1984–2006.

percentile values of four-factor $t(\alpha)$ from the simulations of net returns (where by construction skill suffices to cover costs) always beat the corresponding percentiles of $t(\alpha)$ for actual net fund returns.

There is a glimmer of hope for investors in the tests on net returns. Even in the four-factor tests, the 99th and, for the \$5 million group, the 98th percentiles of the $t(\alpha)$ estimates for actual fund returns are close to the average values from the simulations. This suggests that some fund managers have enough skill to produce expected benchmark-adjusted net returns that cover costs. This is, however, a far cry from the prediction of Berk and Green (2004) that most if not all fund managers can cover their costs.

B.2. Gross Returns

It is possible that the fruits of skill do not show up more generally in net fund returns because they are absorbed by expenses. The tests on gross returns in Table III show that adding back the costs in expense ratios pushes up $t(\alpha)$ for actual fund returns. For all AUM groups, however, the left tail of three-factor $t(\alpha)$ estimates for actual gross fund returns is still to the left of the average from the simulations. For example, in the simulations the average value of the fifth percentile of $t(\alpha)$ for gross returns for the \$5 million group is -1.71, but the actual fifth percentile from actual fund returns is much lower, -2.19.

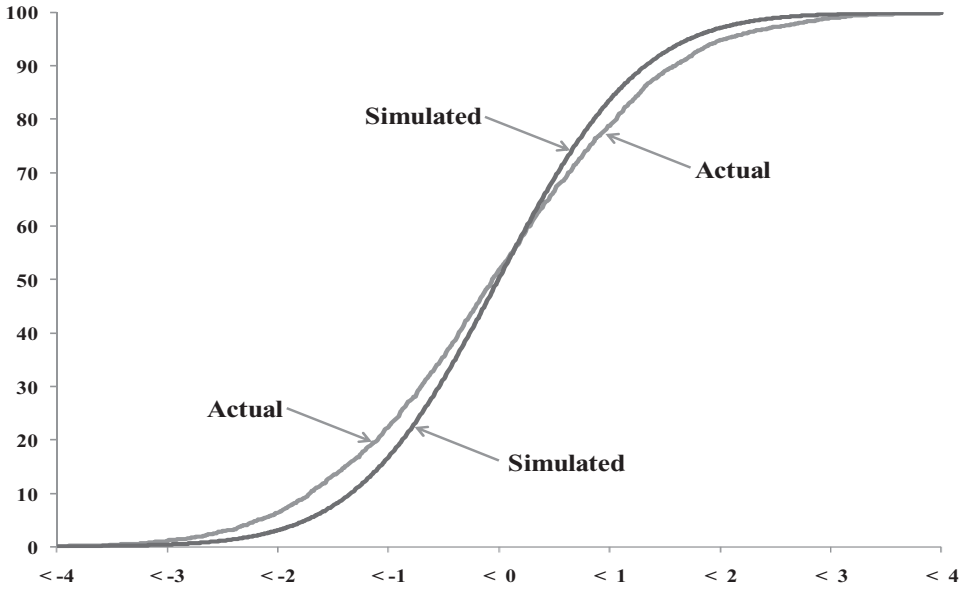

Figure 2. Simulated and actual cumulative density function of three-factor $t(\alpha)$ for gross returns, 1984–2006.

Thus, the left tails of the CDFs of three-factor $t(\alpha)$ suggest that when returns are measured before expenses, there are inferior fund managers whose actions result in negative true α relative to passive benchmarks.

Conversely, the right tails of three-factor $t(\alpha)$ suggest that there are superior managers who enhance expected returns relative to passive benchmarks. For the \$5 million AUM group, the CDF of $t(\alpha)$ estimates for actual gross fund returns moves to the right of the average from the simulations at about the 60^{th} percentile. For example, the 95^{th} percentile of $t(\alpha)$ for funds in the \$5 million group averages 1.68 in the simulations, but the actual 95^{th} percentile is higher, 2.04. For the two larger AUM groups the crossovers occur at higher percentiles, around the 80^{th} percentile for the \$250 million group and the 90^{th} percentile for the \$1 billion group. Figure 2 graphs the results for the three-factor benchmark and the \$5 million AUM group.

The four-factor results for gross returns in Table III are similar to the three-factor results, with a minor nuance. Adding a momentum control tends to shrink slightly the left and right tails of the cross-sections of $t(\alpha)$ estimates for actual fund returns. This suggests that funds with negative three-factor α estimates tend to have slight negative MOM_t exposure and funds with positive three-factor α tend to have slight positive exposure. Controlling for momentum pulls the α estimates toward zero, but only a bit.

Finally, the average simulation distribution of $t(\alpha)$ for the \$5 million fund group is like a t distribution with about 24 degrees of freedom. The average sample life of these funds is 112 months, so we can probably conclude that the

simulation distributions of $t(\alpha)$ are more fat-tailed than can be explained by degrees of freedom. This may be due in part to fat-tailed distributions of stock returns (Fama (1965)). A referee suggests that active trading may also fatten the tails of fund returns. And properties of the joint distribution of fund returns may have important effects on the cross-section of $t(\alpha)$ estimates—a comment of some import in our later discussion of Kosowski et al. (2006).

C. Likelihoods

Comparing the percentiles of $t(\alpha)$ estimates for actual fund returns with the simulation averages gives hints about whether manager skill affects expected returns. Table III also provides likelihoods, in particular, the fractions of the 10,000 simulation runs that produce lower values of $t(\alpha)$ at selected percentiles than actual fund returns. These likelihoods allow us to judge more formally whether the tails of the cross-section of $t(\alpha)$ estimates for actual fund returns are extreme relative to what we observe when true α is zero.

Specifically, we infer that some managers lack skill sufficient to cover costs if low fractions of the simulation runs produce left tail percentiles of $t(\alpha)$ below those from actual net fund returns, or equivalently, if large fractions of the simulation runs beat the left tail $t(\alpha)$ estimates from actual net fund returns. Likewise, we infer that some managers produce benchmark-adjusted expected returns that more than cover costs if large fractions of the simulation runs produce right tail percentiles of $t(\alpha)$ below those from actual net fund returns. The logic is similar for gross returns, but the question is whether there are managers with skill sufficient to cover the costs (primarily trading costs) missing from expense ratios.

There are two problems in drawing inferences from the likelihoods in Table III. (i) Results are shown for many percentiles so there is a multiple comparisons issue. (ii) The likelihoods for different percentiles are correlated. One way to address these problems is to focus on a given percentile of each tail of $t(\alpha)$, for example, the 5^{th} and the 95^{th} percentiles, and draw inferences entirely from them. But this approach discards lots of information. We prefer to examine all the likelihoods, with emphasis on the extreme tails, where performance is most likely to be identified. As a result, our inferences from the formal likelihoods are somewhat informal.

C.1. Net Returns

The likelihoods in Table III confirm that skill sufficient to cover costs is rare. Below the 80^{th} percentile, the three-factor $t(\alpha)$ estimates for actual net fund returns beat those from the simulations in less than 1.0% of the net return simulation runs. For example, the 70^{th} percentile of the cross-section of three-factor $t(\alpha)$ estimates from the net returns of \$5 million funds (our full sample) is 0.08, and only 0.51% (about half of one percent) of the 10,000 simulation runs for this group produce 70^{th} percentile $t(\alpha)$ estimates below 0.08. It seems safe to conclude that most fund managers do not have enough skill to produce

benchmark-adjusted net returns that cover costs. This again is bad news for Berk and Green (2004) since their model predicts that skill sufficient to cover costs is the general rule.

The likelihoods for the most extreme right tail percentiles of the three-factor $t(\alpha)$ estimates in Table III also confirm our earlier conclusion that some managers have sufficient skill to cover costs. For the \$5 million group, the 97th, 98th, and 99th percentiles of the cross-section of three-factor $t(\alpha)$ estimates from actual net fund returns are close to the average values from the simulations, and 49.35% to 58.70% of the $t(\alpha)$ estimates from the 10,000 simulation runs are below those from actual net returns. The likelihoods that the highest percentiles of the $t(\alpha)$ estimates from the net returns of funds in the \$5 million group beat those from the simulations drop below 40% when we use the four-factor model to measure α, but the likelihoods nevertheless suggest that some fund managers have enough skill to cover costs.

Some perspective is helpful. For the \$5 million group, about 30% of funds produce positive net return α estimates. The likelihoods in Table III tell us, however, that most of these funds are just lucky; their managers are not able to produce benchmark-adjusted expected returns that cover costs. For example, the 90th percentile of the $t(\alpha)$ estimates for actual net fund returns is near 1.00. The average standard error of the α estimates is 0.28% (monthly), which suggests that funds around the 90th percentile of $t(\alpha)$ beat our benchmarks by more than 3.3% per year for the entire period they are in the sample. These managers are sure to be anointed as highly skilled active investors. But about 90% of the net return simulation runs produce 90th percentiles of $t(\alpha)$ that beat those from actual fund returns. It thus seems that, like funds below the 90th percentile, most funds around the 90th percentile do not have managers with sufficient skill to cover costs; that is, true net return α is negative.

The odds that managers have enough skill to cover costs are better for funds at or above the 97th percentile of the $t(\alpha)$ estimates. In the \$5 million group, funds at the 97th, 98th, and 99th percentiles of three-factor $t(\alpha)$ estimates do about as well as would be expected if all fund managers were able to produce benchmark-adjusted expected returns that cover costs. But this just means that our estimate of true net return three-factor α for these funds is close to zero. If we switch to the four-factor model, the estimate of true α is negative for all percentiles of the $t(\alpha)$ estimates since the percentiles from actual net fund returns beat those from the simulations in less than 40% of the simulation runs.

What mix of active funds might generate the net return results in Table III? Suppose there are two groups of funds. Managers of good funds have just enough skill to produce zero α in net returns; bad funds have negative α. When the two groups are mixed, the expected cross-section of $t(\alpha)$ estimates is entirely to the left of the average of the cross-sections from the net return simulation runs (in which all managers have sufficient skill to cover costs). Even the extreme right tail of the $t(\alpha)$ estimates for actual net fund returns will be weighed down by bad managers who are extremely lucky but have smaller $t(\alpha)$ estimates than if they were extremely lucky good managers. In our tests,

most of the cross-section of $t(\alpha)$ estimates for actual net fund returns is way left of what we expect if all managers have zero true α. Thus, most funds are probably in the negative true α group. At least for the \$5 million AUM sample, the 97th, 98th, and 99th percentiles of the three-factor $t(\alpha)$ estimates for actual net fund returns are similar to the simulation averages. This suggests that buried in the results are fund managers with more than enough skill to cover costs, and the lucky among them pull up the extreme right tail of the net return $t(\alpha)$ estimates. Unfortunately, these good funds are indistinguishable from the lucky bad funds that land in the top percentiles of the $t(\alpha)$ estimates but have negative true α. As a result, our estimate of the three-factor net return α for a portfolio of the top three percentiles of the \$5 million group is near zero; the positive α of the lucky (but hidden) good funds is offset by the negative α of the lucky bad funds. And when we switch to the four-factor model, our estimate of true α turns negative even for the top three percentiles of the $t(\alpha)$ estimates.

Finally, our tests exclude index funds, but we can report that for 1984 to 2006 the net return three-factor α estimate for the VW portfolio of index funds (in which large, low cost funds get heavy weight) is -0.16% per year (-0.01% per month, $t = -0.61$), and four-factor α is 0.01% per year ($t = 0.02$). Since large, low cost index funds are not subject to the vagaries of active management, it seems reasonable to infer that the net return true α for a portfolio of these funds is close to zero. In other words, going forward we expect that a portfolio of low cost index funds will perform about as well as a portfolio of the top three percentiles of past active winners, and better than the rest of the active fund universe.

C.2. Gross Returns

The simulation tests for net returns ask whether active managers have sufficient skill to cover all their costs. In the tests on gross returns, the bar is lower. Specifically, the issue is whether managers have enough skill to at least cover the costs (primarily trading costs) missing from expense ratios.

The three-factor gross return simulations for the \$5 million AUM group suggest that most funds in the left tail of three-factor $t(\alpha)$ estimates do not have enough skill to produce benchmark-adjusted expected returns that cover trading costs, but many managers in the right tail have such skill. For the 40th and lower percentiles, the three-factor $t(\alpha)$ estimates for the actual gross returns of funds in the \$5 million group beat those from the simulations in less than 30% of the simulation runs, falling to less than 6% for the 10th and lower percentiles. Conversely, above the 60th percentile, the three-factor $t(\alpha)$ estimates for actual gross fund returns beat those from the simulations in at least 56% of the simulation runs, rising to more than 90% for the 96th and higher percentiles. As usual, the results are weaker when we switch from three-factor to four-factor benchmarks, but the general conclusions are the same.

For many readers, the important insight of Berk and Green (2004) is their assumption that there are diseconomies of scale in active management, not their detailed predictions about net fund returns (which are rejected in our tests).

The right tails of the $t(\alpha)$ estimates for gross returns suggest diseconomies. The extreme right tail percentiles of $t(\alpha)$ are typically lower for the \$250 million and \$1 billion groups than for the \$5 million group, and more of the simulation runs beat the extreme right tail percentiles of the $t(\alpha)$ estimates for the larger AUM funds. In the world of Berk and Green (2004), however, the weeding out of unskilled managers should also lead to left tails for $t(\alpha)$ estimates that are less extreme for larger funds. This prediction is not confirmed in our results. The left tails of the $t(\alpha)$ estimates for the \$250 million and \$1 billion groups are at least as extreme as the left tail for the \$5 million group. This contradiction in the left tails of the $t(\alpha)$ estimates makes us reluctant to interpret the right tails as evidence of diseconomies of scale.

The tests on gross returns point to the presence of skill (positive and negative). We next estimate the size of the skill effects. A side benefit is evidence on the power of the simulation tests.

IV. Estimating the Distribution of True α in Gross Fund Returns

To examine the likely size of the skill effects in gross fund returns we repeat the simulations but with α injected into fund returns. We then examine (i) how much α is necessary to reproduce the cross-section of $t(\alpha)$ estimates for actual gross fund returns, and (ii) levels of α too extreme to be consistent with the $t(\alpha)$ estimates for actual fund returns.

Given the evidence that, at least for the \$5 million group (our full sample), the distribution of $t(\alpha)$ estimates in gross fund returns is roughly symmetric about zero (Table III), it is reasonable to assume that true α is distributed around zero. It is also reasonable to assume that extreme levels of skill (good or bad) are rare. Concretely, we assume that each fund is endowed with a gross return α drawn from a normal distribution with a mean of zero and a standard deviation of σ per year.

The new simulations are much like the old. The first step again is to adjust the gross returns of each fund, setting α to zero for the three-factor and four-factor benchmarks and each of the three AUM groups. But now, before drawing the random sample of months for a simulation run, we draw a true α from a normal distribution with mean zero and standard deviation σ per year—the same α for every combination of benchmark model and AUM group for a given fund, but an independent drawing of α for each fund.

It seems reasonable that more diversified funds have less leeway to generate true α. To capture this idea, we scale the α drawn for a fund by the ratio of the fund's (three-factor or four-factor) residual standard error to the average standard error for all funds. We add the scaled α to the fund's benchmark-adjusted returns. We then draw a random sample (with replacement) of 273 months, and for each fund we estimate three-factor and four-factor regressions on the adjusted gross returns of the fund's three AUM samples. The simulations thus use returns that have the properties of actual fund returns, except we know true α has a normal distribution with mean zero and (for the "average" fund) standard deviation σ per year. We do 10,000 simulation runs, and a fund

gets a new drawing of α in each run. To examine power, we vary σ, the standard deviation of true α, from 0.0% to 2.0% per year, in steps of 0.25%.

Table IV shows percentiles of the cross-section of $t(\alpha)$ estimates for actual gross fund returns (from Table III) and the average $t(\alpha)$ estimates at the same percentiles from the 10,000 simulation runs, for each value of σ. These are useful for judging how much dispersion in true α is consistent with the actual cross-section of $t(\alpha)$ estimates. For each σ, the table also shows the fraction of the simulation runs that produce percentiles of $t(\alpha)$ estimates below those from actual fund returns. We use these for inferences about the amount of dispersion in true α we might rule out as too extreme.

A. Likely Levels of Performance

If true α comes from a normal distribution with mean zero and standard deviation σ, Table IV provides two slightly different ways to infer the value of σ. We can look for the value of σ that produces average simulation percentile values of $t(\alpha)$ most like those from actual fund returns. Or we can look for the σ that produces simulation $t(\alpha)$ estimates below those for actual returns in about 50% of the simulation runs. If α has a normal distribution with mean zero and standard deviation σ, we expect the effects of the level of σ to become stronger as we look further into the tails of the cross-section of $t(\alpha)$. Thus, we are most interested in values of σ that match the extreme tails of the $t(\alpha)$ estimates for actual gross fund returns.

The normality assumption for true α is an approximation. We do not expect that a single value of σ (the standard deviation of true α) completely captures the tails of the $t(\alpha)$ estimates for actual fund returns, even if we allow a different σ for each tail. With this caveat, the three-factor and four-factor simulations for the \$5 million group suggest that σ around 1.25% to 1.50% per year captures the extreme left tail of the $t(\alpha)$ estimates for actual gross fund returns, and 1.25% works for the right tail. For the \$250 million and \$1 billion groups, the three-factor simulations again suggest σ around 1.25% to 1.50% per year for the left tail of the $t(\alpha)$ estimates for gross fund returns, but for the right tail σ is lower, 0.75% to 1.00% per year. In the four-factor simulations for the \$250 and \$1 billion groups $\sigma = 1.25\%$ per year seems to capture the extreme left tail of the $t(\alpha)$ estimates for gross fund returns, but the estimate of σ for the right tail is again lower, 0.75% per year. (To save space, Table IV shows results only for the \$5 million and \$1 billion AUM groups.)

The estimates do not suggest much performance, especially for larger funds. Thus, $\sigma = 1.25\%$ says that about one-sixth of funds have true gross return α greater than 1.25% per year (about 0.10% per month) and only about 2.4% have true α greater than 2.50% per year (0.21% per month). For perspective, the average of the OLS standard errors of individual fund α estimates—the average imprecision of α estimates—is 0.28% per month (3.4% per year). Moreover, much lower right tail σ estimates for the \$250 million and \$1 billion funds say that a lot of the right tail performance observed in the full (\$5 million) sample is due to tiny funds.

Table IV

Percentiles of $t(\alpha)$ Estimates for Actual and Simulated Gross Fund Returns with Injected α

The table shows values of $t(\alpha)$ at selected percentiles (Pct) of the distribution of $t(\alpha)$ estimates for Actual gross fund returns (repeated from Table III). The table also shows the average values of the $t(\alpha)$ estimates at the same percentiles from the 10,000 simulations, for seven values of σ (the annual standard deviation of injected α). The final seven columns of the table show, for each value of σ, the percent of the 10,000 simulation runs that produce lower $t(\alpha)$ estimates at the selected percentiles than actual fund returns. The period is January 1984 to September 2006 and results are shown for the three- and four-factor models for the \$5 million and \$1 billion AUM fund groups.

		Average $t(\alpha)$ from Simulations							Percent of Simulations below Actual						
Pct	Actual $t(\alpha)$	0.50	0.75	1.00	1.25	1.50	1.75	2.00	0.50	0.75	1.00	1.25	1.50	1.75	2.00
3-Factor α, AUM > 5 Million															
1	−3.07	−2.63	−2.78	−2.99	−3.24	−3.54	−3.87	−4.23	7.46	15.74	36.37	69.29	92.13	99.00	99.94
2	−2.68	−2.27	−2.38	−2.52	−2.69	−2.89	−3.10	−3.34	8.03	13.67	26.30	49.41	76.25	93.77	99.04
3	−2.48	−2.06	−2.15	−2.27	−2.40	−2.55	−2.72	−2.91	6.55	10.94	19.32	35.16	57.73	80.85	94.54
4	−2.31	−1.91	−1.99	−2.08	−2.20	−2.33	−2.47	−2.62	6.85	10.63	17.71	30.94	49.77	71.63	88.96
5	−2.19	−1.78	−1.85	−1.94	−2.04	−2.16	−2.28	−2.41	6.36	9.68	15.54	26.25	41.95	61.44	80.67
10	−1.72	−1.37	−1.42	−1.48	−1.55	−1.62	−1.70	−1.78	7.86	10.82	15.39	22.37	32.27	44.75	59.63
90	1.59	1.35	1.40	1.46	1.53	1.60	1.68	1.76	86.35	81.64	74.23	64.06	51.23	36.75	22.61
95	2.04	1.75	1.83	1.92	2.02	2.13	2.26	2.39	88.27	82.46	72.10	56.14	36.57	18.02	5.63
96	2.20	1.87	1.96	2.06	2.17	2.30	2.45	2.60	90.76	85.40	74.75	57.87	36.04	16.06	4.08
97	2.44	2.03	2.12	2.23	2.37	2.53	2.70	2.88	93.73	89.76	80.72	63.35	38.59	15.13	3.39
98	2.72	2.23	2.35	2.49	2.66	2.85	3.07	3.31	95.29	91.75	82.56	61.96	32.18	8.75	1.24
99	3.03	2.58	2.74	2.95	3.20	3.49	3.82	4.18	93.48	85.84	63.90	29.57	5.82	0.41	0.02

		Average $t(\alpha)$ from Simulations							Percent of Simulations below Actual						
Pct	Actual $t(\alpha)$	0.25	0.50	0.75	1.00	1.25	1.50	1.75	0.25	0.50	0.75	1.00	1.25	1.50	1.75
3-Factor α, AUM > 1 Billion															
1	−3.29	−2.42	−2.54	−2.73	−2.99	−3.31	−3.68	−4.09	2.20	3.63	8.89	22.71	48.69	76.60	92.24
2	−2.70	−2.12	−2.21	−2.34	−2.52	−2.73	−2.98	−3.25	6.27	9.11	15.83	28.78	51.16	75.26	90.50
3	−2.51	−1.94	−2.01	−2.12	−2.27	−2.44	−2.63	−2.84	5.82	8.10	13.21	22.97	39.66	61.21	80.93
4	−2.33	−1.80	−1.87	−1.97	−2.09	−2.23	−2.40	−2.57	6.43	8.75	13.73	22.41	36.16	55.28	74.51
5	−2.18	−1.69	−1.75	−1.84	−1.95	−2.08	−2.22	−2.37	7.42	9.81	14.45	22.65	35.12	51.86	70.08
10	−1.86	−1.32	−1.36	−1.42	−1.49	−1.58	−1.67	−1.77	4.46	5.54	7.88	11.48	17.22	25.15	36.41
90	1.34	1.29	1.34	1.40	1.47	1.55	1.64	1.74	58.48	52.67	43.86	34.00	23.16	13.78	6.93
95	1.78	1.66	1.72	1.81	1.92	2.04	2.18	2.33	67.79	60.84	49.40	35.18	20.70	9.71	3.14
96	1.96	1.77	1.83	1.93	2.05	2.19	2.35	2.53	74.69	68.02	56.45	40.71	24.10	11.07	3.42
97	2.22	1.90	1.97	2.08	2.23	2.39	2.58	2.78	84.12	79.12	68.67	52.24	32.34	14.98	4.56
98	2.37	2.07	2.16	2.29	2.46	2.67	2.90	3.16	82.02	75.12	61.59	41.62	21.00	6.82	1.56
99	2.97	2.35	2.46	2.64	2.88	3.18	3.52	3.90	93.92	90.90	81.49	61.18	32.94	11.99	2.84

(*continued*)

Table IV—*Continued*

Pct	Actual $t(\alpha)$	Average $t(\alpha)$ from Simulations							Percent of Simulations below Actual						
		0.50	0.75	1.00	1.25	1.50	1.75	2.00	0.50	0.75	1.00	1.25	1.50	1.75	2.00
4-Factor α, AUM > 5 Million															
1	−3.06	−2.69	−2.85	−3.06	−3.33	−3.63	−3.97	−4.34	10.99	22.26	47.07	78.71	95.82	99.64	99.97
2	−2.71	−2.31	−2.42	−2.57	−2.74	−2.94	−3.16	−3.41	8.36	14.61	28.65	52.04	78.54	94.44	99.25
3	−2.46	−2.09	−2.18	−2.30	−2.44	−2.60	−2.77	−2.96	8.82	14.08	25.32	42.58	66.01	86.35	96.53
4	−2.27	−1.93	−2.01	−2.11	−2.23	−2.36	−2.51	−2.66	9.85	15.06	24.90	39.73	60.10	80.07	93.38
5	−2.11	−1.80	−1.87	−1.96	−2.07	−2.18	−2.31	−2.45	11.46	16.87	26.22	39.83	57.77	76.75	90.48
10	−1.62	−1.38	−1.43	−1.49	−1.56	−1.64	−1.72	−1.80	16.05	21.02	27.97	37.70	49.46	63.12	76.66
90	1.58	1.38	1.43	1.50	1.56	1.64	1.72	1.80	79.81	74.21	66.46	55.96	43.49	30.25	18.17
95	2.05	1.80	1.87	1.96	2.07	2.18	2.31	2.44	83.71	76.60	66.13	50.01	31.99	15.33	4.86
96	2.21	1.92	2.00	2.11	2.22	2.36	2.50	2.66	86.46	79.25	68.28	50.91	30.64	12.88	3.36
97	2.39	2.08	2.17	2.29	2.43	2.59	2.76	2.95	87.10	79.52	66.70	46.44	24.15	7.38	1.23
98	2.58	2.29	2.41	2.55	2.72	2.92	3.14	3.38	85.21	75.29	57.57	32.34	10.71	1.74	0.15
99	3.01	2.66	2.81	3.02	3.28	3.58	3.91	4.27	88.10	75.85	49.98	19.30	3.25	0.19	0.01

Pct	Actual $t(\alpha)$	Average $t(\alpha)$ from Simulations							Percent of Simulations below Actual						
		0.25	0.50	0.75	1.00	1.25	1.50	1.75	0.25	0.50	0.75	1.00	1.25	1.50	1.75
4-Factor α, AUM > 1 Billion															
1	−3.34	−2.44	−2.56	−2.76	−3.03	−3.36	−3.74	−4.16	2.00	3.35	8.55	22.91	48.31	75.80	92.22
2	−2.48	−2.12	−2.22	−2.36	−2.54	−2.77	−3.02	−3.30	16.25	22.21	35.17	54.99	75.98	91.84	97.93
3	−2.40	−1.93	−2.01	−2.13	−2.28	−2.46	−2.66	−2.88	9.63	13.33	20.63	34.25	54.13	74.16	89.62
4	−2.25	−1.80	−1.87	−1.97	−2.10	−2.25	−2.42	−2.60	9.86	13.41	19.68	30.72	47.91	66.77	83.79
5	−2.11	−1.68	−1.75	−1.84	−1.96	−2.09	−2.24	−2.40	10.76	13.64	19.81	29.76	44.52	62.13	78.56
10	−1.85	−1.30	−1.35	−1.41	−1.49	−1.58	−1.67	−1.78	5.10	6.48	8.67	12.86	19.05	27.52	38.83
90	1.41	1.32	1.36	1.43	1.50	1.59	1.69	1.79	63.74	58.42	50.10	40.62	29.70	19.17	10.28
95	1.83	1.69	1.76	1.85	1.96	2.09	2.24	2.40	68.10	61.80	50.88	37.12	22.46	10.67	3.71
96	1.95	1.80	1.87	1.97	2.10	2.25	2.41	2.59	68.50	61.33	49.38	34.72	19.04	7.98	2.30
97	2.04	1.94	2.02	2.13	2.28	2.45	2.64	2.86	64.06	55.68	41.85	26.04	11.89	3.83	0.74
98	2.30	2.12	2.21	2.35	2.53	2.74	2.98	3.25	71.76	62.55	47.39	28.62	11.89	3.45	0.70
99	2.57	2.40	2.52	2.71	2.96	3.26	3.61	4.00	68.67	57.31	38.54	18.10	5.17	0.98	0.11

Our gross fund returns are net of trading costs. Returning trading costs to funds (if that is deemed appropriate) would increase the $t(\alpha)$ estimates in both the left and the right tails, which, depending on the (unknown) magnitudes, may move them toward more similar estimates of σ.

B. Unlikely Levels of Performance

What levels of σ can we reject? The answer depends on how confident we wish to be about our inferences. Suppose we are willing to accept a 20% chance of setting a lower bound for σ that is too high and a 20% chance of setting an upper bound that is too low. These bounds imply a narrower range than we would have with standard significance levels, but they are reasonable if our goal is to provide perspective on likely values of σ.

Under the 20% rule, the lower bound for the left tail estimate of σ is the value that produces left tail percentile $t(\alpha)$ estimates below those from actual fund returns in about 20% of the simulation runs. The upper bound for the left tail σ is the value that produces left tail percentiles of $t(\alpha)$ below those from actual fund returns in about 80% of the simulation runs. Conversely, under the 20% rule, the lower bound for the right tail σ estimate produces right tail percentile $t(\alpha)$ estimates below those from actual fund returns in about 80% of the simulation runs. And the upper bound for the right tail σ produces right tail percentiles of $t(\alpha)$ below those from actual fund returns in about 20% of the simulation runs.

In brief, applying the 20% rule leads to intervals for σ that are equal to the point estimates of the preceding section plus and minus 0.5%. For example, 1.25% per year works fairly well as the left tail estimate of σ for all AUM groups and for the three-factor and four-factor models, and the interval for the left tail σ estimates is 0.75% to 1.75%. For the \$5 million group, $\sigma = 1.25\%$ also works for the right tail, and the interval is again 0.75% to 1.75%. For the \$250 million and \$1 billion groups, the right tail estimate of σ drops to about 0.75% per year, and the 20% rule leads to an interval for σ from 0.25% to 1.25% per year.

What do these results say about the power of the simulation approach? The upper bound on σ for the \$5 million group, 1.75% per year, translates to a monthly σ for the cross-section of true α of about 0.146%. Suppose the standard error of each fund's α estimate is 0.28% per month (the sample average). With a monthly σ of 0.146%, the standard deviation of the cross-section of α estimates—caused by measurement error and dispersion in true α—is $(0.146^2 + 0.28^2)^{1/2} = 0.316\%$. This is only a bit bigger than 0.299%, the standard deviation implied by our estimate of σ for the \$5 million group, 1.25% per year. The fact that the simulations assign a relatively low probability to $\sigma \geq 1.75\%$ despite the small difference between the implied standard deviations of the α estimates for $\sigma = 1.25\%$ (the point estimate) and $\sigma = 1.75\%$ suggests that the simulations have power. The source of the power is our large sample of funds (3,156 in the \$5 million group). With so many funds, the percentiles of $t(\alpha)$

are estimated precisely, which produces power to draw inferences about σ. (We thank a referee for this insight.)

V. Kosowski et al. (2006)

The paper closest to ours is Kosowski et al. (2006). They use bootstrap simulations to draw inferences about performance in the cross-section of four-factor $t(\alpha)$ estimates for net fund returns. Their main inference is more positive than ours. They find that the 95$^{\text{th}}$ and higher percentiles of four-factor $t(\alpha)$ estimates for net fund returns are above the same simulation percentiles in more than 99% of simulation runs. This seems like strong evidence that among the best funds, many have more than sufficient skill to cover costs. Our simulations on net returns uncover much less evidence of skill. Two features of their tests account for their stronger results—simulation approach and time period.

We jointly sample fund (and explanatory) returns, whereas Kosowski et al. (2006) do independent simulations for each fund. The benefit of their approach is that the number of months a fund is in a simulation run always matches the fund's actual number of months of returns. The cost is that their simulations do not take account of the correlation of α estimates for different funds that arises because a benchmark model does not capture all common variation in fund returns. They summarize but do not show simulations that jointly sample the four-factor residuals of funds. But they never jointly sample fund returns and explanatory returns, which means (for example) they miss any effects of correlated movement in the volatilities of four-factor explanatory returns and residuals. In fact, in the results they show, the explanatory returns do not vary across simulation runs; the historical sequence of explanatory returns is used in every run.

Their rules for including funds in the simulation tests are also different. They include the complete return histories of all funds that survive more than 60 months (so there is survival bias). We include funds after they pass \$5 million in AUM if they have at least 8 months of returns thereafter (less survival bias).

Table V shows simulation results for their 1975 to 2002 period using (i) their rules for including funds and (ii) our rules. Note that both sets of simulations use our approach to drawing simulation samples, that is, a simulation run uses the same random sample of months for all funds, which allows for all effects implied by the joint distribution of fund returns, and of fund and explanatory returns.

The rules used to include funds affect the cross-section of $t(\alpha)$ estimates for actual fund returns. Specifically, the right tail $t(\alpha)$ estimates for actual fund returns are less extreme for our sample. This suggests that their rule that a fund must have at least 60 months of returns produces more survival bias than our 8-month rule. Another possibility is that some funds have high returns when they are tiny but do not do as well after they pass \$5 million. This may be due in part to an incubation bias in the fund sample of Kosowski et al. (2006),

Table V
Percentiles of Four-Factor $t(\alpha)$ for Actual and Simulated Fund Returns: 1975 to 2002

The table shows values of four-factor $t(\alpha)$ at selected percentiles (Pct) of the distribution of $t(\alpha)$ for actual (Act) net and gross fund returns for funds selected using the exclusion rules of Kosowski et al. (2006) and for funds in our \$5 million AUM group selected using our exclusion rules. The period is 1975 to 2002 (as in Kosowski et al. (2006)). The table also shows the fraction (%<Act) of the 10,000 simulation runs that produce lower values of $t(\alpha)$ at the selected percentiles than those observed for actual fund returns. Sim is the average value of $t(\alpha)$ at the selected percentiles from the simulations.

	Kosowski et al. Exclusion Rules			Our Exclusion Rules		
Pct	Sim	Act	%<Act	Sim	Act	%<Act
1	−2.48	−3.69	0.18	−2.46	−3.70	0.16
2	−2.16	−3.25	0.19	−2.14	−3.17	0.30
3	−1.96	−2.87	0.53	−1.95	−2.80	0.70
4	−1.82	−2.55	1.34	−1.80	−2.63	0.69
5	−1.70	−2.36	1.90	−1.69	−2.41	1.36
10	−1.31	−1.92	2.17	−1.30	−1.95	1.66
20	−0.85	−1.41	2.15	−0.85	−1.41	2.17
30	−0.52	−1.01	3.18	−0.52	−1.00	3.54
40	−0.25	−0.65	5.75	−0.24	−0.66	5.35
50	0.01	−0.33	9.19	0.01	−0.34	8.50
60	0.27	−0.02	12.20	0.27	−0.03	11.92
70	0.55	0.29	16.51	0.55	0.27	14.86
80	0.87	0.73	32.80	0.87	0.69	28.11
90	1.32	1.44	68.19	1.32	1.34	56.29
95	1.69	1.97	82.42	1.69	1.81	68.32
96	1.80	2.18	88.38	1.80	2.00	75.70
97	1.94	2.38	90.73	1.94	2.25	83.74
98	2.12	2.59	91.38	2.12	2.51	87.57
99	2.40	3.07	95.79	2.42	2.83	88.37

since they include a fund's entire return history if the fund survives for 60 months.

For either sample of funds, joint sampling of fund returns (our approach) affects the simulation results. Kosowski et al. (2006) report that more than 99% of their simulation runs produce 95th percentile four-factor $t(\alpha)$ estimates below the 95th percentile from actual net fund returns. In Table V, the number drops to 82.42% for the fund sample selected using their rules and 68.32% using our rules. Skipping the details, we can report that the stronger performance results from the fund sample chosen using their rules is due to the 60-month survival rule. If the survival rule is reduced to 8 months, their rules for including funds produce simulation results close to ours. The important point, however, is that whatever inclusion rules are used, failure to account for the joint distribution of fund returns, and of fund and explanatory returns, biases the inferences of Kosowski et al. (2006) toward positive performance. (Cuthbertson, Nitzche, and O'Sullivan (2008) apply the simulation approach of Kosowski

et al. to U.K. mutual funds, with similar results and, we guess, similar problems.)

Time period is also an important source of differences in results. Our simulations for 1984 to 2006 produce much less evidence of funds with sufficient skill to cover costs. In Table III, the CDFs of four-factor $t(\alpha)$ estimates for the net fund returns of 1984 to 2006 are always to the left of the average CDFs from the net return simulations (in which funds have sufficient skill to cover costs). Even in the extreme right tail of four-factor $t(\alpha)$ for net returns, more than 60% of the simulation runs beat the $t(\alpha)$ estimates for actual fund returns. But when our approach is applied to the 1975 to 2002 period of Kosowski et al. (2006), the 90th and higher percentiles of $t(\alpha)$ for net fund returns are above the average values from the simulations (Table V). And for the 97th and higher percentiles, less than 20% of the simulation runs beat the $t(\alpha)$ estimates for actual fund returns.

What do we make of the stronger results for 1975 to 2002 versus 1984 to 2006? One story is that in olden times there were fewer funds and a larger percentage of managers with sufficient skill to cover costs. Over time the skilled managers lost their edge or went on to more lucrative pursuits (e.g., hedge funds). Or perhaps, the entry of hordes of mediocre managers posing as skilled (Cremers and Petajisto (2009)) buries the tracks of true skill. Stronger results for 1975 to 2002 may also be due to biases in the CRSP data that are more prevalent in earlier years (Elton et al. (2001)). Whatever the explanation, the stronger evidence for performance during 1975 to 2002 is interesting, but irrelevant for today's investors.

VI. Conclusions

For 1984 to 2006, when the CRSP database is relatively free of biases, mutual fund investors in aggregate realize net returns that underperform CAPM, three-factor, and four-factor benchmarks by about the costs in expense ratios. Thus, if there are fund managers with enough skill to produce benchmark-adjusted expected returns that cover costs, their tracks are hidden in the aggregate results by the performance of managers with insufficient skill.

When we turn to individual funds, the challenge is to distinguish skill from luck. With 3,156 funds in our full ($5 million AUM) sample, some do extraordinarily well and some do extraordinarily poorly just by chance. To distinguish between luck and skill, we compare the distribution of $t(\alpha)$ estimates from actual fund returns with the distribution from bootstrap simulations in which all funds have zero true α. The tests on net returns say that few funds have enough skill to cover costs. The distribution of three-factor $t(\alpha)$ estimates from net fund returns is almost always to the left of the zero α distribution. The extreme right tail of the three-factor $t(\alpha)$ estimates for net fund returns, however, is roughly in line with the simulated distribution. This suggests that some managers do have sufficient skill to cover costs. But the estimate of net return three-factor true α is about zero even for the portfolio of funds in the top percentiles of historical three-factor $t(\alpha)$ estimates, and the estimate of four-factor true α is

negative. Moreover, the estimate of true α for funds in the top percentiles is no better than the estimated α (also near zero) for large, efficiently managed passive funds.

The simulation results for gross fund returns say that when returns are measured before the costs in expense ratios, there is stronger evidence of manager skill, negative as well as positive. For our \$5 million AUM sample, true three-factor or four-factor gross return α seems to be symmetric about zero with a cross-section standard deviation of about 1.25% per year (about 10 basis points per month). For larger (\$250 million and \$1 billion AUM) funds, the standard deviation for the left tail is again about 1.25% per year, but the right tail standard deviation of true α falls to about 0.75%.

Appendix A: Measurement Issues in Gross Returns

The question in the tests on gross fund returns is whether managers have skill that causes expected returns to differ from those of comparable passive benchmarks. For this purpose, we would like to have fund returns measured before all costs but net of non-return income like revenues from securities lending. This would put funds on the same pure return basis as the benchmark explanatory returns, so the tests could focus on the effects of skill. Our gross fund returns are before the costs in expense ratios, but they are net of other costs, primarily trading costs, and they include income from securities lending.

We could attempt to add trading costs to our estimates of gross fund returns. Funds do not report trading costs, however, and even when turnover is available, estimates of trading costs are subject to large errors (Carhart (1997)). For example, trading costs are likely to vary across funds because of differences in style tilts, trading skill, and the extent to which a fund is actively managed and demands immediacy in trade execution. Trading costs can also vary through time because of changes in a fund's management and general changes in the costs of trading. All this leads us to conclude that estimates of trading costs for individual funds, especially actively managed funds, are fraught with error and potential bias, and so can be misleading. As a result, we do not take that route in our tests on gross returns.

An alternative approach (suggested by a referee) is to put the passive benchmarks produced by combining the explanatory returns in (1) in the same units as the gross fund returns on the left of (1). This involves taking account of the costs (primarily trading costs) not covered in expense ratios that would be borne by an efficiently managed passive benchmark with the same style tilts as the fund whose gross returns are to be explained.

Vanguard's index funds are good candidates for this exercise since, except for momentum, Vanguard provides index funds (Total Stock Market Index Fund, Growth Index Fund, Value Index Fund, Small-Cap Index Fund, Small-Cap Growth Index Fund, and Small-Cap Value Index Fund) that track well-defined target passive portfolios much like the market portfolio and the components of SMB_t and HML_t in (1). (We thank an Associate Editor for this insight.) Because the Vanguard index funds closely track their targets and stock picking skill is

not an issue, we can estimate the average annual costs not included in a fund's expense ratio. Specifically, we add a fund's expense ratio to its reported average annual return for the 10 years through 2008 and then subtract the result from the average annual return of the fund's target for the same period. (The same calculation for an actively managed fund would include the effects of skill, as well as the costs not in expense ratios.) For every Vanguard index fund, this estimate of the costs missed in expense ratios is negative; that is, the fund's target return, which is before all costs, beats the fund's actual net return by less than the fund's expense ratio. If anything, Vanguard's small cap index funds do better on this score than its large cap funds—a clear warning that presumptions about trading costs can be misleading.

The Vanguard results are probably not unusual. We can report that the CAPM, three-factor, and four-factor α estimates for 1984 to 2006 for the net returns on a VW portfolio of index funds (which is dominated by large funds with low expense ratios) are close to zero, 0.08%, −0.16%, and 0.01% per year ($t = 0.18$, -0.61, and 0.02). In other words, in aggregate, wealth invested in index funds seems to earn average returns that cover costs, including trading costs.

Passive mutual funds that focus on momentum do not as yet exist, so we do not have estimates of trading costs for such funds. Existing work (Grundy and Martin (2001), Korajczyk and Sadka (2004)) suggests that the costs are significant. In our tests, however, the cross-sections of four-factor α estimates for funds are similar to the cross-sections of three-factor estimates, and the three-factor and four-factor tests produce much the same inferences. Given the large average MOM_t return, these results suggest that nontrivial long-term exposure to MOM_t is rare, so ignoring MOM_t trading costs is inconsequential. Moreover, the discussion of results in the text centers primarily on the three-factor model. The four-factor results are primarily a robustness check.

The Vanguard evidence and the results for a VW portfolio of index funds suggest that for the market and the components of SMB_t and HML_t, comparable efficiently managed passive mutual funds can enhance returns through trading, securities lending, and perhaps in other ways, so that their total costs are close to their expense ratios. Thus, our three-factor α estimates for the gross returns of funds would hardly change if we adjusted their passive benchmarks for the costs missed in expense ratios.

This does not mean our tests on gross returns capture the pure effects of skill. Though expense ratios seem to capture the total costs of efficiently managed passive funds, this is less likely to be true for actively managed funds. The typical active fund trades more than the typical passive fund, and active funds are likely to demand immediacy in trading that produces positive costs. Because of their high turnover, active funds also have fewer opportunities to generate revenues via securities lending (which are also trivial for the Vanguard funds). In short, it seems more likely that for active funds the costs not included in expense ratios are positive. Thus, our tests on the gross returns of funds produce α estimates that capture the effects of skill, less any costs missed by the expense ratios of the funds.

Equivalently, our tests on gross returns say that a fund's management has skill only if the fund's expected gross returns are sufficient to cover the costs (primarily trading costs) not included in its expense ratio. This is a reasonable definition of skill since a comparable efficiently managed passive fund would apparently avoid these costs. More important, this definition of skill is the only one we can accurately test in the absence of accurate estimates of the trading costs of active funds (impossible with available data).

It is fortuitous that efficiently managed passive benchmarks do not seem to have substantial costs missed in their expense ratios since accurate adjustment for such costs is nontrivial, perhaps impossible. For example, consider an actively managed small value fund. The passive benchmark for the fund produced by the three-factor version of (1) is likely to imply positive weights on the market, *SMB*, and *HML*, which implies positive weights on the market (M), small stocks (S), and value stocks (H) and negative weights on big stocks (B) and growth stocks (L). Suppose that (contrary to our estimates) efficiently managed passive funds have nontrivial trading costs. We might then increase the three-factor gross return α estimate for an active fund for the trading costs of the long positions in M, S, and H and the short positions in B and L that passively replicate the small value style of the active fund. But this is overkill. The three-factor model produces a passive clone for an actively managed fund by inefficiently combining five passive portfolios. A small value fund simply buys a diversified portfolio of small value stocks and only bears the trading costs of these stocks. As a result, even a passive small value fund evaluated with the three-factor model is likely to produce a positive α estimate if we enhance the estimate with positive trading costs for the five components of its three-factor clone.

If we wish to adjust the tests on gross returns for the trading costs of an efficiently managed passive fund with the same style tilts as the active fund to be evaluated, the correct procedure is to add an estimate of the trading costs of a comparable efficiently managed passive fund to the active fund's gross return α estimate. For example, a small value active fund would be reimbursed for the trading costs (more precisely, for all the costs missed in the expense ratio) of an efficiently managed passive fund with the same style tilts. This is nontrivial since a style group includes active funds with widely different style tilts, and we need an efficiently managed passive clone for every active fund. Fortunately, the costs missed in expense ratios are apparently close to zero for efficiently managed passive funds, and ignoring them (as we do in our tests) is inconsequential for inferences.

Appendix B: CAPM Bootstrap Simulations

Table AI replicates the bootstrap simulations in Table III for a CAPM benchmark, that is, regression (1) with the excess market return as the only explanatory variable. The CAPM results are different. The CAPM tests on net returns produce what seems like strong evidence that some fund managers have sufficient skill to cover costs. Thus, for percentiles above the 90^{th}, the CAPM $t(\alpha)$

Table AI

Percentiles of CAPM $t(\alpha)$ Estimates for Actual and Simulated Fund Returns

The table shows values of $t(\alpha)$ at selected percentiles (Pct) of the distribution of CAPM $t(\alpha)$ estimates for actual (Act) net and gross fund returns. The table also shows the percent of the 10,000 simulation runs that produce lower values of $t(\alpha)$ at the selected percentiles than those observed for actual fund returns (%<Act). Sim is the average value of $t(\alpha)$ at the selected percentiles from the simulations. The period is January 1984 to September 2006 and results are shown for the \$5 million, \$250 million, and \$1 billion AUM fund groups.

	5 Million			250 Million			1 Billion		
Pct	Sim	Act	%<Act	Sim	Act	%<Act	Sim	Act	%<Act
Net Returns									
1	−2.36	−3.72	0.25	−2.30	−3.70	0.40	−2.27	−4.10	0.08
2	−2.06	−3.28	0.45	−2.02	−3.29	0.58	−2.00	−3.50	0.24
3	−1.88	−3.00	0.64	−1.85	−3.02	0.79	−1.84	−3.29	0.23
4	−1.75	−2.84	0.62	−1.72	−2.92	0.65	−1.71	−3.18	0.19
5	−1.65	−2.69	0.74	−1.62	−2.76	0.77	−1.62	−3.00	0.27
10	−1.29	−2.16	1.08	−1.28	−2.18	1.64	−1.28	−2.47	0.46
20	−0.86	−1.48	1.93	−0.86	−1.58	1.98	−0.87	−1.79	0.70
30	−0.54	−1.05	2.09	−0.55	−1.11	2.30	−0.56	−1.35	0.44
40	−0.26	−0.65	3.84	−0.27	−0.75	2.50	−0.28	−0.88	0.48
50	0.00	−0.29	8.05	0.00	−0.36	5.23	−0.01	−0.46	1.29
60	0.26	0.08	20.79	0.26	0.06	19.86	0.26	−0.10	4.02
70	0.53	0.49	46.40	0.53	0.47	43.16	0.54	0.31	18.52
80	0.84	0.95	71.01	0.84	0.89	61.89	0.84	0.72	36.21
90	1.26	1.66	91.09	1.25	1.49	79.61	1.24	1.42	73.88
95	1.61	2.31	97.29	1.58	2.09	92.39	1.56	1.91	84.74
96	1.71	2.45	97.55	1.67	2.23	93.43	1.66	2.03	85.94
97	1.84	2.68	98.46	1.79	2.43	95.05	1.77	2.22	89.01
98	2.01	2.89	98.69	1.95	2.60	95.07	1.92	2.47	92.06
99	2.29	3.21	98.88	2.21	2.96	96.51	2.16	2.76	92.96
Gross Returns									
1	−2.36	−3.04	4.09	−2.30	−3.01	5.35	−2.27	−3.29	2.00
2	−2.06	−2.66	5.29	−2.02	−2.67	6.32	−2.00	−2.93	2.57
3	−1.88	−2.45	5.88	−1.85	−2.45	7.17	−1.84	−2.76	2.37
4	−1.75	−2.26	7.41	−1.72	−2.31	7.54	−1.71	−2.49	3.99
5	−1.65	−2.13	7.82	−1.62	−2.16	8.80	−1.62	−2.34	4.91
10	−1.29	−1.65	11.87	−1.28	−1.66	13.56	−1.28	−1.95	4.93
20	−0.86	−0.95	33.12	−0.86	−1.04	25.14	−0.87	−1.35	7.59
30	−0.54	−0.55	44.63	−0.55	−0.63	35.49	−0.56	−0.88	12.00
40	−0.26	−0.19	62.18	−0.27	−0.26	50.41	−0.28	−0.43	24.27
50	0.00	0.16	77.76	0.00	0.10	67.74	−0.01	−0.05	41.74
60	0.26	0.53	89.27	0.26	0.46	81.45	0.26	0.36	67.57
70	0.53	0.98	96.44	0.53	0.91	91.87	0.54	0.77	82.64
80	0.84	1.44	97.60	0.84	1.37	95.08	0.84	1.18	87.03
90	1.26	2.12	98.96	1.25	1.98	97.29	1.24	1.82	94.23
95	1.61	2.76	99.65	1.58	2.47	98.14	1.56	2.33	96.87
96	1.71	2.89	99.69	1.67	2.72	98.98	1.66	2.46	97.14
97	1.84	3.12	99.77	1.79	2.85	99.01	1.77	2.59	97.16
98	2.01	3.35	99.84	1.95	3.05	99.18	1.92	2.84	98.03
99	2.29	3.72	99.89	2.21	3.37	99.35	2.16	3.34	99.14

estimates for actual net fund returns are always above the averages from the net return simulations (in which all managers have sufficient skill to cover costs), and the $t(\alpha)$ estimates for actual fund returns typically beat those from the simulations in more than 80% of simulation runs. Relative to the three-factor and four-factor tests in Table III, the CAPM tests on gross returns in Table AI also produce what seems like stronger evidence that some managers have skill that leads to positive true α, while others have negative true α.

In fact, the CAPM results just illustrate well-known patterns in average returns that cause problems for the CAPM during our sample period. Actual mutual fund returns contain the effects of size, value-growth, and momentum tilts in fund portfolios that are missed by the CAPM. Thus, even passive funds that tilt toward small stocks, value stocks, or positive momentum stocks are likely to produce positive α estimates in CAPM tests, despite the fact that their managers make no effort to pick individual stocks. The CAPM simulations allow for the relation between average return and market exposure, but they wash out all other patterns in average returns when they subtract each fund's CAPM α estimate from its returns. As a result, the CAPM simulations say that actual fund returns have nonzero true α.

Which patterns in average returns left unexplained by the CAPM are most responsible for the differences between the CAPM simulation results and the results for the three-factor and four-factor models? Table III says that adding the momentum factor to the three-factor model has minor effects on estimates of $t(\alpha)$. Since the momentum return MOM_t has the highest average premium during our sample period, we infer that long-term exposure to momentum is probably rare among mutual funds. The average size (SMB_t) premium is trivial during our 1984 to 2006 sample period (0.03% per month, Table I), so size tilts probably are not driving the different results for the CAPM. That leaves the value (HML_t) premium as the focus of the story. Funds in the right tail of the CAPM $t(\alpha)$ estimates are more likely to have positive HML_t exposure that makes them look good in CAPM tests, and funds in the left tail are likely to have negative HML_t exposure.

In short, the CAPM tests are a lesson about how failure to account for common patterns in returns and average returns can affect inferences about the skill of fund managers.

REFERENCES

Berk, Jonathan B., and Richard C. Green, 2004, Mutual fund flows in rational markets, *Journal of Political Economy* 112, 1269–1295.

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

Cremers, Martijn, and Antti Petajisto, 2009, How active is your fund manager? A new measure that predicts performance, *Review of Financial Studies* 22, 3329–3365.

Dybvig, Philip H., and Stephen A. Ross, 1985, The analytics of performance measurement using a security market line, *Journal of Finance* 40, 401–416.

Elton, Edwin J., Martin J. Gruber, and Christopher R. Blake, 2001, A first look at the accuracy of the CRSP mutual fund database and a comparison of the CRSP and Morningstar mutual fund databases, *Journal of Finance* 56, 2415–2430.

Evans, Richard, 2010, Mutual fund incubatier, *Journal of Finance* 65, Forthcoming.

Fama, Eugene F., 1965, The behavior of stock market prices, *Journal of Business* 38, 34–105.

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

Ferson, Wayne E., and Rudi W. Schadt, 1996, Measuring fund strategy and performance in changing economic conditions, *Journal of Finance* 51, 425–462.

French, Kenneth R., 2008, The cost of active investing, *Journal of Finance* 63, 1537–1573.

Grinblatt, Mark, and Sheridan Titman, 1992, Performance persistence in mutual funds, *Journal of Finance* 47, 1977–1984.

Gruber, Martin J., 1996, Another puzzle: The growth of actively managed mutual funds, *Journal of Finance* 51, 783–810.

Grundy, Bruce D., and J. Spencer Martin, 2001, Understanding the nature of the risks and the sources of the rewards to momentum investing, *Journal of Financial Studies* 14, 29–78.

Jensen, Michael C., 1968, The performance of mutual funds in the period 1945–1964, *Journal of Finance* 23, 2033–2058.

Korajczyk, Robert A., and Ronnie Sadka, 2004, Are momentum profits robust to trading costs? *Journal of Finance* 59, 1039–1082.

Kosowski, Robert, Allan Timmermann, Russ Wermers, and Hal White, 2006, Can mutual fund "stars" really pick stocks? New evidence from a bootstrap analysis, *Journal of Finance* 61, 2551–2595.

Malkiel, Burton G., 1995, Returns from investing in equity mutual funds: 1971–1991, *Journal of Finance* 50, 549–572.

O'Sullivan, N., 2008, UK mutual fund performance: Skill or luck? *Journal of Empirical Finance* 15, 613–634 [with K. Cuthbertson and D. Nitzsche].

Sharpe, William F., 1991, The arithmetic of active management, *Financial Analysts Journal* 47, 7–9.