Passive versus Active Fund Performance: Do Index Funds Have Skill?

Alan D. Crane and Kevin Crotty*

Abstract

We apply methods designed to measure mutual fund skill to a cross section of funds that is unlikely to exhibit managerial portfolio selection skill: index funds. Surprisingly, these tests imply index fund skill exists, is persistent, and is in similar proportion as in active funds. We use the distribution of passive fund performance to gauge the incremental ability of active managers. Outperformance by top active funds is lower when benchmarked to the index fund distribution and disappears when we account for residual risk. Stochastic dominance tests suggest no risk-averse investor should choose a random active fund over a random index fund.

I.   Introduction

The performance evaluation literature continues to debate the extent of skill in actively managed mutual funds. Recent work on disentangling skill and luck focuses on tests using the cross-sectional distribution of active fund performance (e.g., Fama and French (2010), Kosowski, Timmermann, Wermers, and White (2006), and Barras, Scaillet, and Wermers (2010)). Most of this recent evidence implies that some active mutual fund managers are, at least before fees, skilled. However, Cremers, Petajisto, and Zitzewitz (CPZ) (2013) document that even benchmark indices such as the Standard & Poor's (S&P) 500 exhibit abnormal performance under standard benchmark models. As a result, tests may identify skill in the right tail of the actively managed fund performance distribution because of heterogeneity in the underlying benchmark choice, not as a result of stock-picking or market-timing ability.[1]

---

[1]CPZ (2013) provide a simple method for reducing the alpha of common benchmark indices by incorporating some indices into the benchmark model. However, indices not explicitly included in their model may still exhibit significant alphas. Some of these excluded indices serve as benchmarks for actively managed mutual funds.

To better understand the extent of skill in actively managed funds given performance dispersion in the underlying benchmarks, we turn to an idea that is intuitive and dates back at least to Malkiel (1995): the use of index funds as the opportunity cost of active management. We exploit the massive growth in the number of index funds over the last 2 decades to extend the intuitive comparison of passive and active funds to distributional tests.[2] We apply these tests, designed to disentangle skill from luck in active management, to the cross section of these competing assets that are unlikely to exhibit managerial portfolio selection skill. Our results are surprising. They imply that index fund skill exists, is persistent, and is found in similar proportion as in active funds. The results suggest that measurement of active management skill can be informed by the distribution of passive fund performance.

The first contribution of the article is the application of distributional tests of skill versus luck to passive funds. Kosowski et al. (2006) and Fama and French (2010) test for skill in active management by benchmarking performance to a "zero-alpha" bootstrap distribution. Surprisingly, a large fraction of index funds outperform this simulated distribution. Additionally, the Barras et al. (2010) false discovery rate methodology classifies over 20% of index funds as skilled under a Fama–French–Carhart 4-factor model. Index funds also exhibit gross performance persistence, which has been used as a measure of skill (e.g., Carhart (1997)). On average, the likelihood of an index fund remaining in the same performance quintile is about 30% from one 5-year period to the next, more than 10% higher than what one would expect by chance. Index funds thus display significant dispersion in performance, which, unlike in active funds, should not be due to the fund manager's portfolio-selection or market-timing ability.[3]

Our index fund findings inform the debate on the performance of active management. Average index fund returns are commonly used as a passive benchmark for active fund performance (e.g., Del Guercio and Reuter (2014), Berk and van Binsbergen (2015)). Outperformance is generally viewed as an active manager's investment skill, which makes sense if index funds are homogeneous assets with no meaningful differences in performance. We document that the proliferation of both indices and the passive funds tracking them results in dispersion in index fund performance. For instance, some of the index funds we study are "smart-beta" funds whose underlying index criteria exploit past outperformance associated with observable stock characteristics such as dividend yield. Previous inference about the extent of skilled management may be overstated if a fund's outperformance is driven not by investment skill but by the same factors that drive index fund dispersion.

The second contribution of the article is our investigation of the implications of dispersion in index fund performance for the evaluation of active management. We first test whether the best (and worst) active funds are skilled (or unskilled) by

---

[2]While the Vanguard 500 Index Fund was the only passive mutual fund for many years, the number of index funds has grown to over 350, and index funds now manage 20% of equity mutual fund assets (Investment Company Institute (2015)).

[3]An index fund manager's primary objective is to track the underlying index rather than dynamically pick stocks. However, passive managers could skillfully manage changes in index constitutions or provide other operational efficiencies. We discuss this operational skill in Section III.E.

comparing their before-fee performance to before-fee index fund performance using quantile regressions. The estimates of active fund performance in the far-right tail are lower when compared with the distribution of index funds. For example, the market model alpha of the 95th percentile active fund is 48 basis points (bps) per month. However, the 95th percentile index fund earns 42 bps per month. The incremental performance of the 95th percentile active fund is thus only 6 bps per month when using the index fund distribution as a benchmark. Moreover, t-statistics, which account for residual risk, suggest that active funds perform no better than index funds, indicating that active funds' incremental outperformance in alpha is risky.

We next test whether the aggregate amount of skill in active funds warrants investing in active funds versus index funds using stochastic dominance tests. This is, ex ante, not obvious. Although the average active fund underperforms the average index fund, some investors may find it desirable to gamble on active funds, trading off the possibility of picking an active fund with a very high alpha against the cost of potentially ending up with an active fund with a large negative alpha. This is consistent with the empirical fact that investors do invest with active funds. However, stochastic dominance tests show that the upside of the best active funds is insufficient to warrant investment in active funds over index funds. That is, the cross section of index fund performance second-order stochastically dominates that of active funds. This is true for both alphas and t-statistics under a variety of benchmark models. This deepens the active management puzzle discussed by Gruber (1996).

How does the market perceive performance differences among passive funds? Berk and Green (2004) argue that the observed positive relationship between past fund performance and future fund flows is due to rational learning about the skill of fund managers. We find that the flow–performance relationship also exists for index funds based on gross performance. For example, an increase in Fama–French–Carhart abnormal performance of 10 bps per month is associated with increased flows of 3.8 bps of assets under management for index funds. These results suggest that investors view past performance differences across passive funds as informative about underlying differences between these funds. The flow–performance relationship in passive funds thus responds to actual performance differences before fees in addition to the response to fees (Elton, Gruber, and Busse (2004)) and the behavioral response to perceived performance differences due to the framing of performance information (Choi, Laibson, and Madrian (2010)).

Our study is the first to use the index fund *distribution* to better understand the performance ability of active managers. Prior work studying active versus passive performance (e.g., Malkiel (1995), Elton, Gruber, and Blake (1996), and Gruber (1996)) has generally focused on average net returns to investors, which reflect both potential manager skill and the rent-sharing agreement between the investors and the fund. Del Guercio and Reuter (2014) compare average active and passive net performance to study incentives induced by the fund's distribution channel for active managers to exert effort. Berk and van Binsbergen (2015) use Vanguard index funds in a benchmark model to estimate gross dollar performance and conclude that skill is widespread in mutual fund managers. However, index

funds exhibit significant dispersion under gross dollar returns as well, consistent with our findings using other performance measures. Unlike the prior literature, we focus on the entire distribution of performance rather than average effects. Our distributional tests were previously not possible simply due to the limited number of passive funds in the cross section. These tests are now feasible because of the growth in the number of index funds.

Our article is most closely related to that of CPZ (2013), who document that underlying stock indices have alpha under standard performance models such as the Fama–French–Carhart model. They propose improvements to standard models to account for the alpha exhibited by a set of standard benchmarks. Our article builds on their insights to make several contributions to the literature. First, we show that investors respond to index fund performance differences due to the benchmark heterogeneity identified by CPZ (2013). Second, we show that even under their index-based benchmark model, there is substantial variation in the performance of index funds. This is due to the fact that a number of index funds track benchmark indices that are not included in the CPZ (2013) benchmark models. Distributional tests using the full distribution of index funds imply skill in some passive funds, even under the improved benchmark models of CPZ (2013) and Berk and van Binsbergen (2015). This changes inferences about the extent of skill in the cross section of active management.

Third, even adjusting index fund performance for the actual benchmark's return, where the issues identified in CPZ (2013) do not apply, we still find small variation in performance within the index fund sample due to operational skill. Although any abnormal performance is economically quite small when measured in returns, so is the accompanying tracking error. As a result, we find substantial cross-sectional variation in performance per unit of risk (i.e., t-statistics of excess returns) that is unrelated to heterogeneity in underlying benchmark performance, the subject of CPZ (2013). The distributional tests of Fama and French (2010) and Barras et al. (2010) both identify skilled management using variation in t-statistics. We show that variation in t-statistics arises both due to operational skill and to nonzero alphas arising from benchmark models not perfectly pricing passive indices.

Our analysis of the distribution of gross performance in a broad sample of index funds complements prior work on net index fund performance. Elton et al. (2004) find that net-of-fee performance is persistent within S&P 500 funds due primarily to fee differences, to which investor flows respond. We show persistence within the broader cross section of index funds, even before fees, and that flows respond to pre-expense return differences. Hortacsu and Syverson (2004) develop a theory to explain the variation in S&P 500 index fund fees, assuming that these products are homogeneous. Our results show not only performance differences across a wider set of index funds but also performance differences in terms of tracking error among funds with the same benchmark (e.g., S&P 500), suggesting some heterogeneity even within benchmarks.

More broadly, our results contribute to the literature on the skill of mutual fund managers. Some articles conclude that active managers are skilled, whereas

other articles conclude the opposite.[4] Our results, using a new economic hurdle to assess skill, show that although some active funds are more skilled than index funds in terms of gross alphas, the incremental outperformance is reduced when accounting for the distribution of passive performance. The differences disappear when using a performance measure that adjusts for the amount of residual risk (i.e., t-statistics). Additionally, this study is the first to document that index funds second-order stochastically dominate active funds.

The rest of the article is organized as follows: In Section II, we describe our sample and benchmark models. Section III shows that index funds appear skilled using methodologies designed to identify skill in the cross section. In Section IV, we use the distribution of index fund performance to evaluate the extent of incremental skill in active funds. Section V concludes.

II.   Data and Benchmark Models

A.   Sample Construction

We use fund characteristics and monthly returns from the Center for Research in Security Prices (CRSP) Survivor-Bias-Free U.S. Mutual Fund Database. Although the Vanguard 500 Index Fund was introduced in the mid-1970s, the number of index funds was small for the next 2 decades. Thus, we start our sample in 1995 with 29 index funds. Our sample contains 237 index funds in total. We merge these data with S12 holdings data from Thomson Reuters using the Wharton Research Data Services MF Links file, requiring a match to be included in the sample. To avoid double-counting observations for multiple share classes, we aggregate information across share classes, weighting by total net assets in each class and summing total net assets across classes.[5] We employ two screens to avoid the incubation bias documented by Evans (2010). First, funds must be at least 3 years old to be included in our sample. Second, we exclude funds whose average net fund assets are below $5 million in the sample. We focus on equity funds, requiring that on average over the sample, at least 90% and at most 105% of the fund's assets be invested in common stocks for a fund to be included in the sample.

Many studies identify index funds as funds containing "index" in the fund's name. We use a stricter definition of index funds, utilizing the CRSP index fund flag. This flag is only populated later in the sample, so we carry the earliest value back. Under our definition of index funds, we identify funds with a value of "D" as index funds. This corresponds to "Pure Index Funds" in the CRSP manual.[6]

[4]Examples of articles concluding at least some active skill include those by Grinblatt and Titman (1989), (1992), (1993), Daniel, Grinblatt, Titman, and Wermers (1997), Chen, Jegadeesh, and Wermers (2000), Wermers (2000), Bollen and Busse (2001), Kosowski et al. (2006), Jiang, Yao, and Yu (2007), Kacperczyk, Sialm, and Zheng (2008), Cremers and Petajisto (2009), Fama and French (2010), Barras et al. (2010), Glode (2011), Berk and van Binsbergen (2015), CPZ (2013), Pastor, Stambaugh, and Taylor (2015), Jiang, Verbeek, and Wang (2014), Hunter, Kandel, Kandel, and Wermers (2014), and Kacperczyk, van Nieuwerburgh, and Veldkamp (2014). Papers concluding no skill include those by Jensen (1968), Elton, Gruber, Das, and Hlavka (1993), Malkiel (1995), Gruber (1996), and Carhart (1997).
[5]We exclude several fund-months with obvious reporting errors in returns.
[6]Our conclusions are unchanged when using a broader, name-based definition of index funds.

We hand-check each fund included in the index fund sample to verify our classification. Exchange-traded funds (ETFs) are included in the sample. For our purposes, the differences between index mutual funds and ETFs are minor because both represent traded, passive portfolios. We discuss the performance of open-ended and exchange-traded funds separately in Section III.F.

It is worth noting that our index fund sample is composed of funds tracking many underlying indices. Each of these underlying indices makes portfolio choice decisions with varying degrees of activity and discretion levels. For instance, the S&P 500 constituents are selected using both quantitative and qualitative criteria by Standard & Poor's. Although it is generally a low-turnover benchmark, there is an underlying qualitative portfolio choice decision in its construction. On the other hand, indices such as the Russell 1000 are more clearly rules-based. Our sample includes well-known indices (e.g., Wilshire 5000) as well as lesser-known benchmarks that, although still rules-based indices, have different selection criteria than simple size or value screens (e.g., Morningstar Dividend Leaders Index or Nasdaq Capital Strength Index). Some of these indices represent the trend toward so-called "smart beta."

In Section IV, we test whether active funds exhibit performance superior to index funds. To avoid index fund performance dispersion due to concentrated holdings, we restrict the sample to exclude sector funds as well as international and emerging market funds (our results are robust to including these funds). To identify these, we parse the fund names from CRSP and manually identify keywords associated with these funds. We exclude any fund whose name contains these keywords in any month of the sample. We also exclude sector funds based on Lipper codes provided by CRSP.[7] Finally, we manually look at all remaining index funds to ensure that the fund is not a sector fund.

Table 1 reports summary statistics for our sample of index and active funds. The sample includes 2,060 distinct funds, 237 of which are passive index mutual funds or ETFs. On average, the index funds in our sample are over twice as large as active funds, but they are also younger. This is consistent with the rapid increase in index funds over the last 2 decades. As expected, expenses are much lower for index funds. The average expense ratio is 46 bps for index funds and 124 bps for active funds.

Our index fund sample consists of funds CRSP identifies as passively tracking an underlying benchmark index. Holdings and return data reflect the more passive investments of the index funds. In Table 1, turnover, as reported by CRSP, is much lower for index funds; the median index fund has a turnover of 24% compared to 74% for active funds. Table 1 shows that the cross-sectional average and median return gap measures of Kacperczyk et al. (2008), which capture unobserved actions of funds, are close to 0 for both active and passive funds, but the dispersion is almost twice as large for active funds relative to passive funds, indicating that index funds are much less active funds.

---

[7]A list of the keywords and Lipper codes is available from the authors. Some keywords appear as part of the fund name (e.g., due to the fund family) or in ways that are clearly not related to a sector fund. We manually checked words where this is the case and did not flag funds that are clearly not sector funds.

TABLE 1

Summary Statistics

Table 1 presents summary statistics for our sample. The sample contains monthly fund observations from 1995 to 2013. Risk loadings are estimated using the Fama–French–Carhart 4-factor model (see Table 2). Fund characteristics are winsorized at the 1/99% level and averaged over the time series of each fund. The cross-sectional mean, median, and standard deviation of these time-series averages are reported. Tracking error and active share are for the subset of funds benchmarked to Standard & Poor's (S&P) and Russell indices.

	Mean	Median	Std. Dev.
No. of distinct mutual funds	2,060		
No. of active funds	1,823		
No. of passive funds	237		
No. of fund-month observations	262,162		
No. of funds per month	1,148	1,322	346
No. of index funds per month	111	125	65
Index Funds			
TNA (total net assets) (millions)	1,913	290	4,070
Age (years)	7.8	7.0	3.8
Expense ratio (in %)	0.46	0.36	0.33
Turnover ratio (in %)	42.02	24.44	61.99
Return gap (in %)	−0.03	−0.01	0.08
Active share	0.14	0.02	0.25
SD (tracking error) (annualized %)	0.52	0.17	1.18
MKT loading	1.00	1.00	0.08
SMB loading	0.23	0.15	0.41
HML loading	0.05	0.03	0.26
UMD loading	−0.02	−0.01	0.10
Active Funds			
TNA (total net assets) (millions)	890	220	2,316
Age (years)	12.0	8.4	11.1
Expense ratio (in %)	1.24	1.22	0.39
Turnover ratio (in %)	84.95	73.59	55.81
Return gap (in %)	−0.02	−0.02	0.14
Active share	0.78	0.80	0.16
SD (tracking error) (annualized %)	6.72	6.13	3.23
MKT loading	1.02	1.01	0.12
SMB loading	0.20	0.09	0.34
HML loading	0.01	0.01	0.29
UMD loading	0.01	0.00	0.11

Cremers and Petajisto (2009) develop a holdings-based measure, active share, to study how active a manager is. If index funds deviate substantially from benchmark weights, the active share should deviate from 0. Positions orthogonal to the index would have an active share of 1.[8] Table 1 shows large differences in active shares across fund type. The median index fund has an active share of 0.02, indicating that these funds hold assets in proportions very close to those of the benchmark. On the other hand, the median active fund deviates widely from its benchmark, as evidenced by the median active share of 0.8.

B.   Benchmark Models and Performance Measures

Our primary measures of performance are alphas from benchmark models and their t-statistics. The appropriate benchmark model is a matter of extensive debate in the mutual fund literature. For completeness, we present results using a number of benchmark models to account for different levels of systematic risk taking (detailed in Table 2). We start by using the single-market capital asset pricing model (CAPM) of Jensen (1968) as well as the Fama–French–Carhart (FFC)

---

[8]We obtain active share data from Antti Petajisto's Web site (http://www.petajisto.net/data.html), constructed using the methodology outlined by Petajisto (2013). These data are available through 2009. We consider the subsample of funds in the Petajisto (2013) data set matching our sample.

TABLE 2

Benchmark Models

Table 2 presents the benchmark model specifications used in our analysis.

Model	Specification	Benchmark Returns
1. Excess of S&P 500	$r_{it} - r_t^{SP500} = \alpha + \epsilon_{it}$	r_t^{SP500}: S&P 500 return
2. Market model/CAPM (Jensen (1968))	$r_{it} - r_t^f = \alpha + \beta MKT_t + \epsilon_{it}$	MKT_t: CRSP value-weighted market excess return
3. Fama–French–Carhart (Fama and French (1993), Carhart (1997))	$r_{it} - r_t^f = \alpha + \beta_1 MKT_t + \beta_2 SMB_t$ $+ \beta_3 HML_t + \beta_4 UMD_t + \epsilon_{it}$	SMB_t: small-cap minus large-cap HML_t: value minus growth UMD_t: past winners minus past losers (lagged returns)
4. Vanguard basis (Berk and van Binsbergen (2015))	$r_{it} - r_t^f = \alpha + \sum_{j=1}^{11} \beta^j V_t^j + \epsilon_{it}$	V_t^j: excess return on Vanguard index j, orthogonalized to $V_t^{n<j}$. The Vanguard index funds are S&P 500 Index, Extended Market Index, Small-Cap Index, European Stock Index, Pacific Stock Index, Value Index, Balanced Index, Emerging Markets Stock Index, Mid-Cap Index, Small-Cap Growth Index, and Small-Cap Value Index.
5. Cremers–Petajisto–Zitzewitz 7-factor model (CPZ (2013))	$r_{it} - r_t^f = \alpha + \beta_1 S5RF_t + \beta_2 RMS5_t$ $+ \beta_3 R2RM_t + \beta_4 S5VS5G_t$ $+ \beta_5 RMVRMG_t$ $+ \beta_6 R2VR2G_t + \beta_7 UMD_t + \epsilon_{it}$	$S5RF_t$: S&P 500 index excess return $RMS5_t$: Russell Midcap minus S&P 500 $R2RM_t$: Russell 2000 minus Russell Midcap $S5VS5G_t$: S&P 500 Value minus S&P 500 Growth $RMVRMG_t$: Russell Midcap Value minus Russell Midcap Growth $R2VR2G_t$: Russell 2000 Value minus Russell 200 Growth
6. Conditional 4-factor model (Ferson and Schadt (1996))	$r_{it} - r_t^f = \alpha + \beta_1 MKT_t + \beta_2 SMB_t$ $+ \beta_3 HML_t + \beta_4 UMD_t$ $+ \sum_{j=1}^{K} B_j [z_{j,t-1} \times MKT_t]$ $+ \epsilon_{it}$	$z_{j,t-1}$: public information variable j's deviation from its time-series mean We use $K = 4$ conditioning variables: 1. The 1-month Treasury bill yield 2. The dividend yield of New York Stock Exchange/American Stock Exchange firms over the previous 12 months 3. The term spread (10-year Treasury minus 3-month Treasury yield) 4. The default spread (Baa minus Aaa corporate yields)

4-factor model of Carhart (1997).[9] Recently, Berk and van Binsbergen (2015) use an orthogonal basis of 11 Vanguard index funds as benchmark funds, which we include. We also use the 7-factor model (CPZ7) proposed by CPZ (2013). Finally, we adjust for time-varying risk taking using the conditional 4-factor (FS) model of Ferson and Schadt (1996).

For each benchmark model, we estimate loadings and time-series alphas for each fund according to the following model:

$$(1) \qquad r_{it} - r_t^f \quad = \quad \alpha_i + \sum_{j=1}^{n} \beta_i^j F_t^j + \epsilon_{it},$$

where r_{it} is fund i's return in month t, r_t^f is the risk-free rate, and F_t^j is the excess return on benchmark return or factor j in period t. We require at least 36 months of data for a given fund in order to estimate the models. Standard errors are adjusted for heteroskedasticity.

We report summary statistics of the full-sample gross Fama–French–Carhart factor loadings in Table 1. The loadings are similar across the index and actively managed funds. Both groups have average market betas of approximately 1 and a slight tilt toward small firms. Neither group loads heavily on value or momentum strategies on average.

The estimated α is our first measure of performance. An advantage of this measure is that it provides the economic magnitude of any abnormal performance,

[9]Factor returns are obtained from Ken French's Web site (http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html).

allowing us to gauge the economic value added by a fund.[10] Due to different sample lengths or heterogeneous risk taking by funds, an estimated α may not have attractive sampling properties. For these reasons, Kosowski et al. (2006) and Fama and French (2010) analyze the distribution of t-statistics associated with estimated alphas. Consequently, we use $t(\alpha)$ as our second measure of performance. Economically, t-statistics provide information about performance per unit of residual risk taken by a fund. This is important for comparison of funds with very different idiosyncratic risk profiles.

We examine these measures using before-fee (gross) returns.[11] Gross alphas and t-statistics allow us to ask the question of whether or not a fund exhibits sufficient skill to outperform the passive benchmark implied by the model (an alpha greater than 0). A study of net returns, conversely, addresses a different question: whether active managers have sufficient skill to cover the fees they charge to investors. However, this will capture, in part, the bargaining process between fund investors and managers (Berk and Green (2004)). Because we are interested in the cross-sectional distribution of passive portfolio performance, we abstract from this confounding economic mechanism by analyzing gross performance.

III.   Index Funds and Distributional Tests of Skill

In this section, we assess the extent of heterogeneity in index fund performance using tests designed to detect skill in performance evaluation. We find evidence of dispersion in performance using several tests from the literature. We proceed to test whether investors perceive these differences as meaningful by examining whether index fund flows respond to past performance in the same way as active funds. Finally, we explore the source of index fund performance.

A.   Bootstrapping the Cross Section of $t(\alpha)$

Recent work by Fama and French (2010) and Kosowski et al. (2006) uses bootstrap analysis to simulate distributions of skill measures under the null of no skill. These studies both recognize that the underlying cross-sectional distribution of fund returns is likely to be nonnormal, and therefore inferences based on standard critical values can be confounded. The idea is simply that there is a spread in alpha estimates due to noise in estimation and the statistical properties of the individual fund returns, even in the absence of true alpha (i.e., skill). Both studies simulate the null distribution of fund returns by sampling from actual fund returns net of estimated alphas. The studies differ in sample construction and bootstrap methodology, but both conclude that a small set of active funds possesses skill (at least before fees).

How do index funds, a set of funds generally considered unskilled, fare relative to the bootstrap distribution? To assess this, we use the approach of Fama and French (2010). For each fund-month, we subtract a fund's estimated gross alpha from the fund's monthly gross return. This leaves us with a panel of monthly

---

[10]Another measure of economic value is the dollar return described by Berk and van Binsbergen (2015), which we analyze in the Internet Appendix (available at www.jfqa.org).

[11]For some funds, CRSP does not report expenses monthly. For these funds, we carry forward the annually reported fees to subsequent monthly observations.

zero-alpha fund returns. From these data, we draw a bootstrap sample of months (with replacement) from the set of all months in our sample. If we draw a given month, we use all fund returns from that month to retain any cross-sectional correlation in monthly returns. For each bootstrap sample, we then calculate the time-series alpha and $t(\alpha)$ for each fund. This provides us with a cross-sectional distribution of $t(\alpha)$ estimated from returns that, by construction, should have a true alpha of 0.[12] We repeat this 1,000 times and average across the bootstrap samples at each point in the distribution of estimated $t(\alpha)$. We perform this analysis for zero-alpha distributions using each of the benchmark models described in Section II.B.

We tabulate the percentiles of the bootstrapped and actual distributions of gross $t(\alpha)$ performance for index funds in Table 3. The empirical distribution of $t(\alpha)$ for index funds outperforms the bootstrap distribution above the 50th percentile for all models. That is, the index fund distribution exhibits higher $t(\alpha)$ values at a given percentile than the bootstrap distribution. Given that this outperformance has been interpreted as skilled investing by the fund, it is surprising that the index funds perform better than the bootstrapped sample over a large part

TABLE 3

Percentiles of $t(\alpha)$ for Index and Active Fund Actual and Bootstrapped Returns

Table 3 shows $t(\alpha)$ values at various percentiles ("Pct" column) for both the bootstrapped zero-alpha distribution and the actual fund performance distribution for index (Panel A) and active funds (Panel B). The bootstrap methodology follows Fama and French (2010). The "Sim" columns contain the average bootstrap value of $t(\alpha)$ at the various percentiles (averaged over 1,000 draws). The "Act" columns contain the empirical distribution. Percentiles in which the actual value exceeds the average bootstrap value are in bold font. The "Lik" column reports the fraction of bootstrap runs in which the bootstrapped percentile falls below the actual percentile. A value of 0.5 indicates the actual distribution is indistinguishable from the bootstrap distribution. We display the percentiles for five benchmark models.

Pct	CAPM			Fama–French–Carhart			Vanguard			Cremers–Petajisto–Zitzewitz			Ferson–Schadt		
	Sim	Act	Lik	Sim	Act	Lik	Sim	Act	Lik	Sim	Act	Lik	Sim	Act	Lik
Panel A. Index Funds															
5	−1.61	−1.58	0.47	−1.61	−1.91	0.22	−1.79	−2.76	0.00	−1.75	**−0.95**	1.00	−1.62	−1.60	0.47
10	−1.29	**−1.19**	0.58	−1.28	−1.37	0.36	−1.36	−2.03	0.01	−1.35	**−0.73**	1.00	−1.25	**−1.17**	0.57
20	−0.87	**−0.59**	0.80	−0.86	−0.87	0.46	−0.88	−1.21	0.04	−0.87	**−0.28**	1.00	−0.81	**−0.78**	0.52
30	−0.55	**−0.22**	0.89	−0.54	**−0.44**	0.62	−0.54	−0.68	0.20	−0.53	**0.15**	1.00	−0.48	**−0.36**	0.66
40	−0.28	**0.01**	0.87	−0.27	**0.01**	0.81	−0.26	−0.22	0.59	−0.27	**0.44**	1.00	−0.21	**−0.01**	0.75
50	−0.01	**0.20**	0.81	−0.01	**0.30**	0.85	0.01	**0.05**	0.61	0.00	**0.81**	1.00	0.05	**0.33**	0.84
60	0.26	**0.50**	0.84	0.24	**0.56**	0.84	0.27	**0.43**	0.84	0.26	**1.11**	1.00	0.31	**0.56**	0.80
70	0.52	**0.79**	0.83	0.52	**0.88**	0.86	0.55	**0.72**	0.83	0.53	**1.49**	1.00	0.58	**0.76**	0.74
80	0.83	**1.16**	0.86	0.83	**1.13**	0.80	0.89	**1.10**	0.84	0.85	**1.81**	1.00	0.90	**1.06**	0.70
90	1.25	**1.71**	0.89	1.26	**1.34**	0.62	1.38	**1.63**	0.85	1.32	**2.50**	1.00	1.34	**1.37**	0.58
95	1.58	**2.09**	0.89	1.60	**1.64**	0.58	1.81	**2.06**	0.81	1.72	**3.41**	1.00	1.71	1.65	0.48
Panel B. Active Funds															
5	−1.63	−1.75	0.31	−1.66	−1.97	0.14	−1.75	−2.25	0.01	−1.71	−2.08	0.06	−1.67	−2.01	0.11
10	−1.28	−1.24	0.50	−1.29	−1.50	0.21	−1.35	−1.74	0.02	−1.32	−1.65	0.07	−1.29	−1.54	0.17
20	−0.85	**−0.68**	0.71	−0.85	−0.96	0.32	−0.89	−1.11	0.09	−0.87	−1.05	0.15	−0.85	−0.98	0.28
30	−0.53	**−0.31**	0.82	−0.53	−0.56	0.44	−0.55	−0.74	0.10	−0.54	−0.66	0.24	−0.53	−0.63	0.33
40	−0.26	**−0.03**	0.84	−0.26	−0.25	0.52	−0.27	−0.41	0.13	−0.26	−0.33	0.35	−0.25	−0.31	0.40
50	−0.01	**0.29**	0.89	−0.01	**0.09**	0.65	−0.01	−0.07	0.31	−0.01	−0.03	0.46	0.00	−0.02	0.47
60	0.25	**0.55**	0.89	0.24	**0.38**	0.71	0.25	**0.25**	0.48	0.25	**0.30**	0.63	0.25	**0.31**	0.61
70	0.52	**0.88**	0.92	0.52	**0.69**	0.76	0.53	**0.61**	0.73	0.53	**0.71**	0.85	0.53	**0.61**	0.65
80	0.83	**1.23**	0.92	0.83	**1.05**	0.80	0.86	**1.02**	0.87	0.86	**1.08**	0.89	0.85	**0.97**	0.72
90	1.25	**1.78**	0.95	1.28	**1.58**	0.87	1.33	**1.57**	0.92	1.31	**1.73**	0.97	1.29	**1.52**	0.81
95	1.59	**2.28**	0.97	1.65	**2.01**	0.89	1.73	**2.11**	0.96	1.70	**2.25**	0.99	1.67	**1.90**	0.80

---

[12]As pointed out by Fama and French (2010), there is the potential for unbalanced samples across funds. This will be true for index as well as active funds.

of the distribution. As in Fama and French (2010), we also report the fraction of bootstrap runs in which a given percentile of the simulated distribution falls below the empirical percentile value. Fama and French (2010) discuss how this measure can be used for informal inference concerning the likelihood of observing the difference in performance between the simulated and actual data. In particular, if the bootstrapped and actual distributions are equal at a given percentile (i.e., no skill), the likelihood value should be 0.5.[13]

Panel B of Table 3 reports the same tests for our actively managed sample. The results look quite similar to both those in Fama and French (2010) and those of the index funds. The actual percentiles are generally greater than the bootstrapped percentiles above the median. Kosowski et al. (2006) and Fama and French (2010) interpret such results as evidence of skill, although the latter emphasizes that any excess returns disappear when considered net-of-fees. A similar interpretation of the index fund results implies that some index funds are skilled as well.

B.   The Proportion of Skilled, Zero-Alpha, and Unskilled Funds

In a similar vein to the bootstrap tests, Barras et al. (2010) use a variant of the estimation of the false discovery rate (FDR) developed by Storey (2002) to estimate the fractions of funds in the cross section that are skilled, unskilled, and zero-alpha. The technique controls for false discoveries of mutual fund skill (i.e., mutual funds exhibiting significant alphas by luck alone). If one assumes that funds are drawn from one of three populations (skilled, unskilled, and zero-alpha), the cross-sectional distribution of t-statistics for risk-adjusted alphas will be a mixed distribution. The right tail of this mixed distribution will contain both skilled funds and lucky zero-alpha funds. Using a critical value for the t-statistics alone will falsely attribute skill to these lucky zero-alpha funds.

We estimate the proportions of unskilled, zero-alpha, and skilled funds $(\hat{\pi}^-, \hat{\pi}^0, \hat{\pi}^+)$ in the index fund population using the Barras et al. (2010) methodology.[14] The results using gross returns are shown in Table 4. For reference, we also report estimates for the active fund sample. The estimated proportion of skilled funds $(\hat{\pi}^+)$ in the index fund population is statistically significant, and the point estimate is at least as large as the estimated proportion of skilled funds in the active fund population in all benchmark models except the market model. Under a standard Fama–French–Carhart 4-factor model, 21% of index funds are classified as skilled, compared with 9% of active funds. Interestingly, under 3 of the 5

---

[13]Interestingly, the likelihood statistic can be biased away from 0.5 for extreme percentiles even for distributions with no skill (i.e., alpha of 0). This bias is decreasing in the number of funds in the cross section (not the number of bootstrap samples), so the small size of the index fund cross section is a possible concern. The estimated likelihoods for percentiles where the actual estimate exceeds the average bootstrap value (bold entries in Table 3) are generally well in excess of critical values determined using Monte Carlo simulations of the likelihood statistic under the null of no skill.

[14]$\lambda^* \in [0,1]$ denotes the threshold above which p-values are assumed to be generated from zero-alpha funds only (i.e., funds with alpha p-values greater than λ^* are comprised solely of zero-alpha funds). γ^* denotes the significance level used for determining the critical t-value used to estimate the fraction of lucky zero-alpha funds incorrectly identified as possessing skill (or lack of skill). We fix λ^* at 0.5 and γ^* at 0.35 to put active and index funds on equal footing, but our results are qualitatively unchanged if we follow the selection algorithms for λ^* and γ^* used by Barras et al. (2010).

TABLE 4

Proportion of Skilled, Unskilled, and Zero-Alpha Funds: Gross Alpha

Table 4 presents estimates of the proportions of unskilled, zero-alpha, and skilled funds ($\hat{\pi}^-$, $\hat{\pi}^0$, $\hat{\pi}^+$) in the population of index funds using the false discovery rate (FDR) methodology of Barras et al. (2010). For reference, we also report estimates for active funds. Results are presented for five benchmark models. The estimated proportions are truncated at 0 if negative. Standard errors are calculated as in Barras et al. (2010) and are shown in parentheses. $\lambda^* \in [0, 1]$ denotes the threshold above which p-values are assumed to be generated from zero-alpha funds only (i.e., funds with alpha p-values greater than λ^* are comprised solely of zero-alpha funds). To ensure comparable estimates across index and active fund distributions, the threshold λ^* is fixed at 0.5, and the significance level (γ) is fixed at 0.35 (see Barras et al. (2010) for details).

Fund Type	Unskilled ($\hat{\pi}^-$)	Zero-Alpha ($\hat{\pi}^0$)	Skilled ($\hat{\pi}^+$)
Panel A. CAPM			
Index	0.000	0.979	0.021
	(0.030)	(0.065)	(0.049)
Active	0.000	0.883	0.117
	(0.011)	(0.023)	(0.018)
Panel B. Fama–French–Carhart			
Index	0.053	0.734	0.213
	(0.031)	(0.063)	(0.049)
Active	0.049	0.865	0.086
	(0.012)	(0.023)	(0.017)
Panel C. Vanguard			
Index	0.095	0.759	0.146
	(0.033)	(0.063)	(0.047)
Active	0.110	0.815	0.075
	(0.017)	(0.023)	(0.012)
Panel D. Cremers–Petajisto–Zitzewitz 7-Factor			
Index	0.000	0.709	0.291
	(0.019)	(0.062)	(0.051)
Active	0.090	0.803	0.107
	(0.012)	(0.023)	(0.017)
Panel E. Ferson–Schadt			
Index	0.000	0.852	0.148
	(0.028)	(0.064)	(0.050)
Active	0.066	0.843	0.091
	(0.012)	(0.023)	(0.017)

models, we estimate no unskilled index funds. This stands in contrast to the active funds, where we see a substantial fraction of unskilled funds under all models except the CAPM. For models with unskilled passive funds (FFC and Vanguard), the proportion is similar to that found in the active distribution. These tests can speak to the fraction of skilled (or unskilled) funds, but they do not address the magnitude of that skill.[15] It is possible that the skilled funds in the active space have much larger α. We examine this explicitly in Section IV.

C.   Persistence of α

Empirical studies of mutual fund performance often point to the (lack of) persistence of risk-adjusted performance as evidence of the (lack of) skill for managers (e.g., Carhart (1997)).[16] In this section, we evaluate the persistence of

---

[15]Ferson and Chen (2015) propose a methodology to estimate both the proportions of skilled/unskilled funds and the magnitudes of the performance. We find similar estimates of $\hat{\pi}^+$ and $\hat{\pi}^-$ for index funds using their method under the Fama–French–Carhart model (untabulated).

[16]Berk and Green (2004) show that if flows respond to performance and there are decreasing returns to scale in managers' ability, then persistence should not arise as a result of skill. However, if

performance for index funds and compare it to that of active funds. Specifically, we estimate Fama–French–Carhart alphas for half-decade subsamples (i.e., 2000–2004, 2005–2009, etc.) and sort funds into quintiles based on the alphas in each period. If relative performance persists, then the transition matrices of the alpha quintiles should be disproportionately populated along the diagonals. We should expect persistence in the top quintile in particular if funds are truly skilled. If there is no persistence and performance rankings are random from one period to the next, we should see uniform transition probabilities of 20% across the entire matrix.

We present gross alpha transition matrices for index and active funds in Table 5. Panels A and C show transitions from 2000–2004 to 2005–2009, and Panels B and D show transitions from 2005–2009 to 2010–2013. For either transition, there is no evidence of persistence in skill for the top active funds. In contrast, for those index funds in the top quintile, about 30% remain in that quintile for each transition period.

For index funds, there is also evidence of persistently bad funds. In the early period, 26% of the worst funds remain in that quintile over the next 5-year period. However, in the later period, over 30% of the worst-quintile funds remain in that quintile in the second half of the period.

Overall, we see significant persistence along the entire diagonal for index funds. Across the 2 periods, the average diagonal transition probability for index funds is over 30% compared with about 20% for active funds. Moreover, the persistence in the performance of index funds is statistically significant. The coefficient obtained by regressing alpha on lagged alpha is positive with a p-value of 0.08 for index funds (untabulated).

TABLE 5

Persistence of Gross α

Table 5 presents transition matrices for alpha estimates under the Fama–French–Carhart model. Funds are sorted into quintiles based on their estimated gross alphas from each half-decade subsample. Each row shows the transition of a fund from the lagged quintile into quintiles in the current period. Panel headers indicate the current period. The top quintile (High) contains the top-performing funds; the worst-performing funds are in the Low quintile. No shading represents transition probabilities close to random (15%–25%). Bold entries represent above-random persistence (>25%), and italicized entries represent less-than-random persistence (<15%).

$\hat{\alpha}_{t-1}$	$\hat{\alpha}_t$ Low	2	3	4	High	$\hat{\alpha}_{t-1}$	$\hat{\alpha}_t$ Low	2	3	4	High
Panel A. Index Funds (2005–2009)						*Panel C. Active Funds (2005–2009)*					
Low	**25.9**	18.5	*14.8*	**29.6**	*11.1*	Low	22.4	*11.8*	16.5	19.8	**29.5**
2	*7.7*	**34.6**	**42.3**	*11.5*	*3.8*	2	*14.3*	21.5	21.1	21.1	21.9
3	*3.7*	**48.2**	**37.0**	*11.1*	*0.0*	3	19.1	**26.8**	24.5	17.9	*11.7*
4	*8.3*	**50.0**	**29.2**	*12.5*	*0.0*	4	18.4	23.4	24.9	17.6	15.7
High	*11.5*	*0.0*	15.4	**42.3**	**30.8**	High	16.5	18.1	19.3	**26.8**	19.3
Panel B. Index Funds (2010–2013)						*Panel D. Active Funds (2010–2013)*					
Low	**32.5**	*10.0*	*12.5*	20.0	**25.0**	Low	19.8	17.6	18.2	23.5	20.9
2	15.1	**60.6**	18.1	*3.0*	*3.0*	2	19.9	*13.3*	**25.0**	22.7	19.1
3	*0.0*	**28.6**	**40.0**	*11.4*	20.0	3	17.0	23.3	19.3	21.8	18.5
4	**25.6**	*0.0*	**33.3**	**28.2**	*12.8*	4	20.3	**25.4**	19.6	15.9	18.8
High	21.0	*10.5*	*5.3*	**34.2**	**29.0**	High	23.8	18.7	17.6	17.6	22.3

---

there are frictions in how quickly the scale of the fund increases, then we would still expect to observe some persistence.

Elton et al. (2004) show significant persistence in S&P 500 index fund net returns. They find that much of this persistence is driven by fees. Our results indicate a significant amount of persistence in the broader index fund space, even for gross returns.

D.   The Flow–Performance Relationship

A number of articles have documented a positive relationship between net fund flows and lagged performance in active mutual funds (e.g., Sirri and Tufano (1998), Chevalier and Ellison (1997)). A leading explanation for this relationship is that investors rationally update their beliefs about manager skill based on past performance (Berk and Green (2004)). We assess the flow–performance relationship in the context of our group of index funds. If flow–performance is due to rational learning by investors about managers' stock-picking abilities, index fund flows should not be responsive to flows. However, if investors learn about other aspects of performance, there may be a relationship for index funds as well.

We examine the relationship between net fund flows and lagged performance for both active and index funds. We measure new money growth as follows:

$$(2) \qquad \text{FLOW}_{it} \quad = \quad \frac{\text{TNA}_{it} - \text{TNA}_{it-1}(1+r_{it})}{\text{TNA}_{it-1}},$$

where TNA_{it} is the total net assets under management by fund i in month t.[17] Flows are winsorized at the 1% level.

Table 6 presents panel regressions of net fund flows on lagged returns and an interaction of lagged returns with an index fund indicator variable. We use gross excess returns and benchmark-adjusted returns, controlling for the natural log of total net assets and a fund's expense ratio.[18] Each regression contains year and fund fixed effects. As has been widely documented in the literature for active funds, new money growth is positively correlated with lagged fund performance measured using any performance measure. We show that this effect exists for a broad set of index funds as well. The index fund interaction term estimates are positive and statistically significant using all benchmark-adjusted returns except the Ferson–Schadt alpha. An increase in Fama–French–Carhart abnormal performance of 10 bps per month is associated with increased flows of 3.8 bps of assets under management for index funds. This is significantly greater than the 1.8-bps increase in assets for active funds, so new money growth in index funds is more sensitive to past performance than it is in active funds.[19]

This result is inconsistent with investors rationally updating about fund skill if index funds have no skill. It is puzzling that index funds would be more responsive to past gross returns than active funds. Investors may be rationally updating

---

[17]Alternatively, we can scale flows by $\text{TNA}_{it-1}(1+r_{it})$ so that complete liquidation results in a flow of -1. Our results are unchanged using this formulation.

[18]Using preferences revealed by mutual fund flows, Berk and van Binsbergen (2016) argue that CAPM is closest to the true model. To be consistent with our previous tests, we use abnormal returns from each benchmark model.

[19]Although index mutual funds and ETFs are used for sector exposure, we note that we have excluded sector-specific funds from our sample, so this does not drive the result. Importantly, we also include time fixed effects, which should absorb any source of index flows due to time-varying sentiment.

TABLE 6

The Fund Flow–Performance Relationship

Table 6 presents estimates of net fund flows on lagged gross performance. The flow for fund i in month t is given by $\text{FLOW}_{it} = (\text{TNA}_{it} - \text{TNA}_{it-1}(1 + r_{it}))/(\text{TNA}_{it-1})$, where r_{it} is the fund return, and TNA_{it-1} is the total net asset value of the fund. We estimate the following regression:

$$\text{FLOW}_{it} = \delta_i + \delta_t + \beta_1 \alpha_{it-1} + \beta_2 \alpha_{it-1} \times \text{INDEX}_i + \gamma_1 \ln(\text{TNA})_{it} + \gamma_2 \text{EXPENSE\_RATIO}_{it} + \epsilon_{it},$$

where α_{it-1} represents the lagged return of fund i, INDEX_i is an indicator equal to 1 for index funds, $\ln(\text{TNA})_{it}$ is the natural log of total net assets under management, and $\text{EXPENSE\_RATIO}_{it}$ is the expense ratio for the fund. The regression includes fund and year fixed effects. Each column presents the estimates for different returns, including excess returns and various benchmark-adjusted returns. Standard errors are clustered at the fund level, t-statistics are in square brackets, and * and ** indicate significance at the 5% and 1% levels, respectively.

Variable	1	2	3	4	5	6
EXCESS_RETURN	0.055** [26.18]					
EXCESS_RETURN × INDEX	0.014 [1.48]					
CAPM_ALPHA		0.150** [26.07]				
CAPM_ALPHA × INDEX		0.158** [4.84]				
FFC_ALPHA			0.177** [21.77]			
FFC_ALPHA × INDEX			0.201** [4.39]			
VANGUARD_ALPHA				0.180** [19.28]		
VANGUARD_ALPHA × INDEX				0.291** [3.66]		
CPZ7_ALPHA					0.177** [18.01]	
CPZ7_ALPHA × INDEX					0.375** [5.50]	
FS_ALPHA						0.099** [15.95]
FS_ALPHA × INDEX						−0.005 [−0.16]
ln(TNA)	0.003** [8.35]	0.003** [9.12]	0.003** [9.14]	0.003** [9.04]	0.003** [9.14]	0.003** [8.96]
EXPENSE_RATIO	0.813** [6.72]	0.816** [6.69]	0.818** [6.76]	0.823** [6.80]	0.820** [6.71]	0.829** [6.76]
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Fund fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
No. of obs.	256,823	256,823	256,823	256,823	256,823	256,823

about which investment strategies outperform and which funds have the best execution in addition to responding to variation in fees (Elton et al. (2004)). On the other hand, in experimental settings, investors chase past performance even within S&P 500 funds if presented with differently framed information (Choi et al. (2010)), which could explain part of our findings.

E. Why Do Some Index Funds Appear Skilled?

In this subsection, we investigate why the distributional tests identify some index funds as skilled. One possible explanation is that some underlying indices may have nonzero alpha (CPZ (2013)). This dispersion could result either from benchmark model misspecification or portfolio-selection skill in the underlying index. Dispersion in passive funds exists even under the CPZ benchmark model, so benchmark performance heterogeneity may not fully explain the dispersion in

passive fund performance. A second possibility is that differences in operational skill (e.g., trading execution, management of securities lending programs, etc.) result in dispersion of alphas and/or t-statistics.

To assess the relative importance of these two explanations, we decompose index fund performance into i) the average return of the fund in excess of its stated benchmark and ii) the alpha of the underlying benchmark relative to a particular risk model. Some of the benchmark returns and holdings are proprietary, so we cannot decompose returns for all index funds. Therefore, we begin the analysis in this subsection by restricting the sample to 107 funds benchmarked to a standard set of S&P and Russell benchmarks.[20]

In Figure 1, we plot the cross-sectional distribution of both average index fund returns in excess of their stated benchmark (Graph A) and the CAPM alphas of the underlying benchmark indices (Graph B). We also plot their respective t-statistics. Dispersion in average returns in excess of the stated benchmark is due to heterogeneity in operational skills. Graph A shows that the magnitude of any index fund skill due to fund management, but *not* the choice of benchmark, is quite small. Nonetheless, there is still heterogeneity across index funds, and many funds beat their underlying index.[21] In Graph E, we plot the distribution of t-statistics of average returns in excess of the stated benchmark, which exhibits substantial dispersion. Most investors would gauge skill in index funds by how closely the fund tracks its benchmark; better funds have lower tracking error. The t-statistic distribution shows that the tracking error is small enough for many funds to statistically outperform their benchmarks. In other words, even though the magnitude of index fund operational skill is small in terms of returns, some earn these positive returns in excess of the stated benchmark consistently. Such funds have large measures of skill on a t-statistic basis.

In order to gauge the extent to which the risk model alpha of the underlying index explains index fund performance dispersion, we report the CAPM alphas of the underlying benchmarks, estimated over the full sample and weighted by the number of funds tracking each benchmark (Graph B of Figure 1). If the benchmark indices are unskilled, these alphas should all be 0. However, as documented by CPZ (2013), there is some dispersion in benchmark alphas, indicating that the CAPM does not perfectly price all indices. There is also substantial estimation error when estimating the model alphas of the underlying indices. As a result, the dispersion in the $t(\alpha)$ distribution of the CAPM alphas of the underlying index (Graph F) is substantially narrower than the $t(\alpha)$ distribution of the average fund returns in excess of their benchmark (Graph E).

---

[20]The 17 indices we consider are the S&P 500, S&P 500 Growth, S&P 500 Value, S&P 400, S&P 600, Russell 1000, Russell 1000 Growth, Russell 1000 Value, Russell 2000, Russell 2000 Growth, Russell 2000 Value, Russell 3000, Russell 3000 Growth, Russell 3000 Value, Russell Midcap, Russell Midcap Growth, and Russell Midcap Value. As for all index funds in our sample, we hand-check the benchmarks for these index funds. We obtain benchmark returns from Bloomberg.

[21]One potential explanation for outperformance is securities lending programs. Blocher and Whaley (2016) document that passive funds can earn substantial revenue from these programs. Managing costs is another potential explanation; for example, Keim (1999) documents the performance implications of transaction cost management by Dimensional Fund Advisors.

FIGURE 1

Decomposition of Index Fund Performance: CAPM Benchmark Model

Figure 1 plots the histograms of average excess returns and associated *t*-statistics for index funds benchmarked to 17 Standard & Poor's (S&P) and Russell benchmarks listed in footnote 20 (Graphs A–C and E–G) and all index funds in the sample (Graphs D and H). Graphs A and E plot average return in excess of the stated benchmark and associated *t*-statistic for each fund. Graphs B and F plot the capital asset pricing model (CAPM) alpha and $t(\alpha)$ for the underlying index. For each underlying index, we estimate the model over the full sample and weight the estimate by the number of funds tracking the index. Graphs C and G plot the CAPM alpha and $t(\alpha)$ for each index fund benchmarked to the 17 S&P and Russell benchmarks. Graphs D and H use all index funds in the sample and plot the CAPM alpha and $t(\alpha)$ for each index fund.
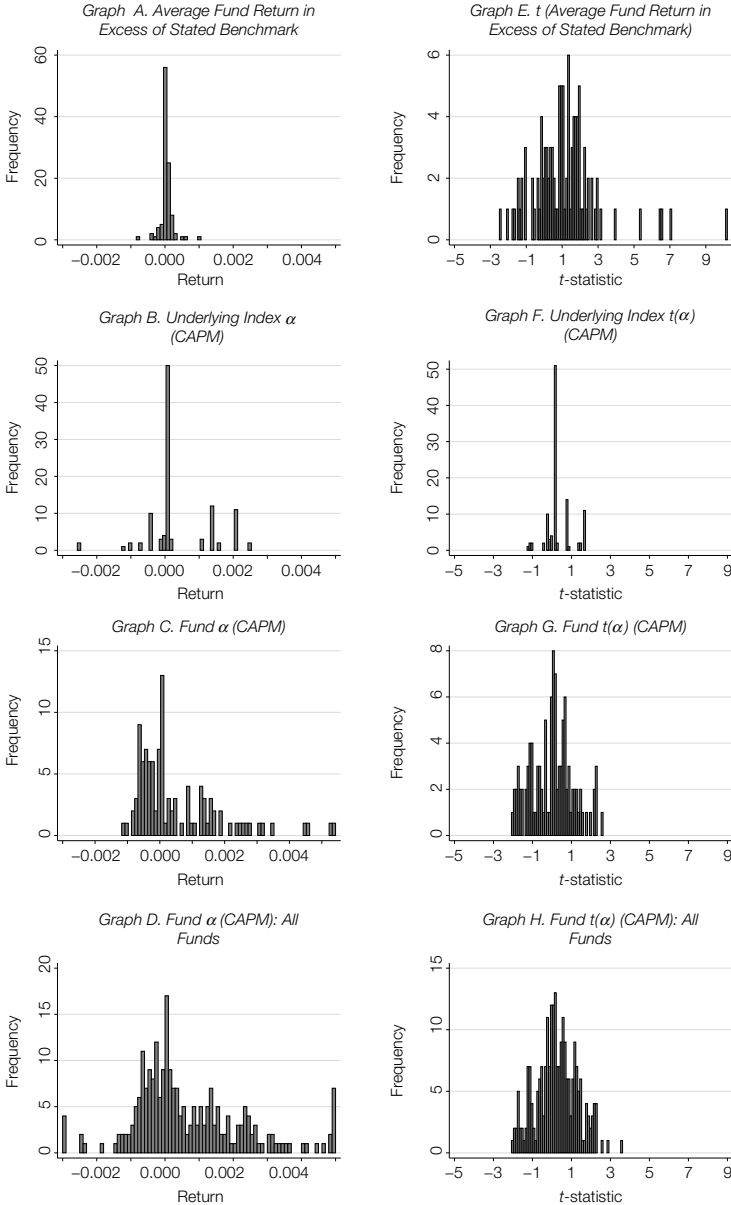
Graphs C and G of Figure 1 report the cross-sectional distributions of the model alphas and $t(\alpha)$ under CAPM for each fund.[22] As to the question of what drives dispersion in index fund performance, the answer depends on which measure of performance one uses: model alphas or $t(\alpha)$. In the first three rows of each column of Figure 1, we see that benchmark model misspecification or underlying benchmark skill accounts for dispersion in alpha for the sample of standard S&P and Russell indices; there is little dispersion in average returns in excess of the stated benchmark. On the other hand, differences in operational skill (i.e., tracking error) are the predominant factor in explaining dispersion in index fund performance as measured by $t(\alpha)$. If we include funds benchmarked to indices outside of the S&P and Russell subsample (final row of Figure 1), the distributions are wider, particularly for CAPM alpha.

It is interesting to assess the performance of benchmark models that improve pricing of indices in particular. For instance, the benchmark models of CPZ (2013) are designed to eliminate alphas for benchmark indices like the S&P 500. In Figure 2, we report estimates of funds' average returns in excess of the stated benchmark and CPZ7 model alphas for the underlying indices. Graph B shows that the CPZ7 model substantially reduces the dispersion in alphas for the underlying S&P and Russell benchmarks. Indeed, only two benchmarks, the S&P 400 and the S&P 600, now have alpha estimates that deviate much from 0. As a result, the distribution of index fund alpha estimates in Graph C is much tighter under CPZ7 than under the CAPM. However, the t-statistic distribution of the fund-specific alphas in Graph G is actually wider under CPZ7 than under CAPM due to operational skill (which is invariant to the benchmark risk model) and smaller idiosyncratic risk for underlying indices relative to the CPZ7 model. The reduction in idiosyncratic risk moving from the CAPM to the CPZ7 model is precisely due to the fact that the CPZ7 is constructed to eliminate the alphas in the underlying benchmarks. Thus, slight operational performance differences in terms of tracking error can result in large t-statistic dispersion.

It is also important to note that although the CPZ7 model corrects for much of the alpha dispersion in the subset of S&P and Russell index funds, it does not correct for this dispersion in the full sample. This is evident in the final row of Figure 2, which plots the performance distributions for all index funds in our sample (no longer restricting to the set of S&P and Russell benchmarks). Even under the CPZ7 model, there is substantial dispersion in the index fund alphas when the index fund sample includes funds benchmarked to indices not included in the CPZ7 model. Therefore, although part of the performance dispersion in index funds is driven by the benchmark-alpha issue identified by CPZ (2013), their solution does not eliminate all variation in the benchmark index alpha.

F.   Which Index Funds Appear Skilled?

Table 7 presents a league table of the top- and bottom-performing index funds based on the fund's CPZ7 alpha or $t(\alpha)$. The table also reports a fund's

---

[22]These estimates are for the period each fund is in the sample, so the underlying benchmark alpha estimate may differ across two funds with the same underlying benchmark but differing sample periods.

FIGURE 2

Decomposition of Index Fund Performance: CPZ7 Benchmark Model

Figure 2 plots the histograms of average excess returns and associated t-statistics for index funds benchmarked to 17 Standard & Poor's (S&P) and Russell benchmarks listed in footnote 20 (Graphs A–C and E–G) and all index funds in the sample (Graphs D and H). Graphs A and E plot average return in excess of the stated benchmark and associated t-statistic for each fund. Graphs B and F plot the CPZ7 alpha and $t(\alpha)$ for the underlying index. For each underlying index, we estimate the model over the full sample and weight the estimate by the number of funds tracking the index. Graphs C and G plot the CPZ7 alpha and $t(\alpha)$ for each index fund benchmarked to the 17 S&P and Russell benchmarks. Graphs D and H use all index funds in the sample and plot the CPZ7 alpha and $t(\alpha)$ for each index fund.
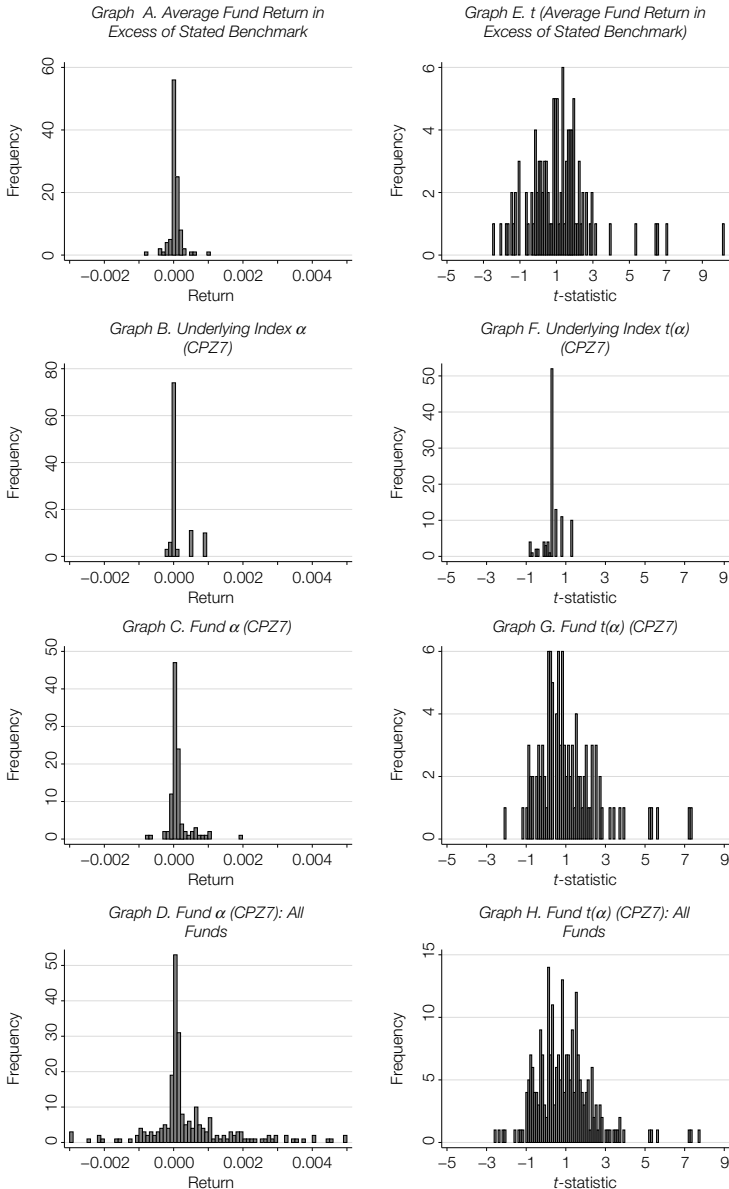
TABLE 7

League Table: Cremers–Petajisto–Zitzewitz 7-Factor Benchmark

Table 7 reports the top and bottom 10 index funds based on alpha and $t(\alpha)$ as well as the average rank of each measure across 6 benchmarks: excess of Standard & Poor's (S&P) 500, capital asset pricing model (CAPM), Fama–French–Carhart, Vanguard Basis, Cremers–Petajisto–Zitzewitz 7-factor, and Ferson–Schadt. The ranks are out of the 237 funds. A rank of 1 corresponds to the largest (smallest) alpha or t-statistic in Panels A and C (Panels B and D).

		Average Rank	
Rank	Fund	α	$t(\alpha)$
Panel A. Top 10 Alpha (CPZ7)			
1	Claymore ETF Trust: Claymore/Beacon Spin-Off ETF	2	8
2	First Trust ETF: First Trust Morningstar Dividend Leaders Index Fund	19	53
3	Rydex Series Funds: NASDAQ-100 Fund	24	51
4	PowerShares ETF Trust: Buyback Achievers Portfolio	10	14
5	WisdomTree Trust: WisdomTree SmallCap Dividend Fund	15	31
6	First Trust ETF: First Trust US IPO Index Fund	6	11
7	Rydex ETF Trust: Rydex S&P 500 Pure Value ETF	5	19
8	Morgan Stanley Nasdaq-100 Index Fund	66	101
9	Rydex Series Funds: S&P MidCap 400 Pure Growth Fund	10	26
10	Rydex ETF Trust: Rydex S&P SmallCap 600 Pure Growth ETF	19	56
Panel B. Bottom 10 Alpha (CPZ7)			
1	First Trust ETF: First Trust Value Line 100 ETF	2	12
2	First Trust ETF: First Trust Value Line Equity Allocation Index Fund	6	21
3	PowerShares ETF Trust: PowerShares Zacks Micro Cap Portfolio	4	10
4	PowerShares ETF Trust: PowerShares Dynamic OTC Portfolio	4	11
5	Schwab Capital Trust: Schwab Institutional Select Small-Cap Value Index Fund	105	87
6	WisdomTree Trust: WisdomTree LargeCap Value Fund	21	32
7	PowerShares ETF Trust: PowerShares Fundamental Pure Mid Growth	11	30
8	Claymore ETF Trust: Guggenheim Insider Sentiment ETF	39	58
9	First Trust ETF: First Trust Capital Strength ETF	42	57
10	First Trust ETF: Large Cap Growth AlphaDEX Fund	21	39
Panel C. Top 10 $t(\alpha)$ (CPZ7)			
1	Vanguard Index Funds: Vanguard 500 Index Fund	126	60
2	iShares Trust: iShares Russell 2000 Index Fund	141	140
3	Vanguard Institutional Index Fund: Vanguard Institutional Index Fund	128	69
4	iShares Trust: iShares Russell 2000 Growth Index Fund	136	130
5	iShares Trust: iShares Russell Midcap Index Fund	74	37
6	T Rowe Price Index Trust; T Rowe Price Equity Index 500 Fund	123	47
7	Schwab Capital Trust: S&P 500 Index Fund	139	78
8	iShares Trust: iShares S&P 500 Index Fund	170	156
9	WisdomTree Trust: WisdomTree SmallCap Dividend Fund	15	31
10	Schwab Capital Trust: Schwab Total Stock Market Index Fund	102	9
Panel D. Bottom 10 $t(\alpha)$ (CPZ7)			
1	First Trust ETF: First Trust Value Line 100 ETF	2	12
2	PowerShares ETF Trust: PowerShares Zacks Micro Cap Portfolio	4	10
3	First Trust ETF: First Trust Value Line Equity Allocation Index Fund	6	21
4	Federated Index Trust: Federated Mini-Cap Index Fund	44	28
5	PowerShares ETF Trust: PowerShares Dynamic OTC Portfolio	4	11
6	WisdomTree Trust: WisdomTree LargeCap Value Fund	21	32
7	PowerShares ETF Trust: PowerShares Fundamental Pure Mid Growth	11	30
8	TIAA-CREF Funds: Mid-Cap Growth Index Fund	141	102
9	Schwab Capital Trust: Schwab Institutional Select Small-Cap Value Index Fund	105	87
10	iShares Trust: iShares Morningstar Large-Cap Growth ETF	72	76

average alpha and t-statistic ranks across benchmark models.[23] The ranking of funds depends on the performance measure used. Consistent with the discussion of Figure 2, the best and worst funds based on CPZ7 alpha are those that benchmark to indices not included in the CPZ7 model, whereas the best and worst CPZ7

---

[23]We thank the referee for suggesting the analysis in this section. We report league tables for the other benchmark models in the Internet Appendix.

$t(\alpha)$ funds contain funds benchmarked to common indices like the S&P 500 or Russell 2000.

The league table in Table 7 shows that both exchange-traded and open-ended funds end up as the top-performing funds. ETFs comprise just over half of our index fund sample. The cross section of alpha for exchange-traded index funds exhibits more variance than does the cross section of open-ended index fund alpha, and ETFs occupy a disproportionate share of the top 10 alpha funds. However, the distribution of $t(\alpha)$ is similar across exchange-traded and open-ended index funds, so there is greater representation by open-ended index funds in the top-performing index funds based on t-statistics.

We also assess the average characteristics of top- and bottom-performing index funds by sorting on alpha or t-statistics (untabulated). The average expense ratio is higher for the top and bottom alpha quintiles. The high- and low-alpha quintiles also track higher turnover benchmarks and tend to be smaller and younger funds. This is consistent with some of the less well-known indices seen in the league tables. These patterns are not evident when funds are sorted by $t(\alpha)$. There is no clear pattern related to t-statistic performance relative to expenses or turnover. Although not monotonic, there is some evidence that larger and older funds have higher $t(\alpha)$. These results are consistent with the league tables, which show that many large, well-established index funds following standard benchmarks appear skilled based on $t(\alpha)$.

Finally, we conduct an analysis of whether some fund management firms systematically perform better than others (before fees). We average the across-benchmark-model average ranks of alphas and t-statistics across all funds for each management firm (untabulated). There is variation in this average rank, and the same management firms do well/poorly regardless of alpha or t-statistic sorting. On average, index funds managed by WisdomTree, Guggenheim, State Street (SPDR), and Vanguard perform best among firms with at least 10 index funds in our sample.

IV.   Implications for Active Fund Performance Evaluation

The result that portions of the passive index fund distribution can appear skilled obviously raises questions concerning its implication for performance evaluation of active managers because factors that drive dispersion in index fund performance could also contribute to the variation in performance within the active funds. In this section, we reassess the extent of actively managed skill by using the cross section of traded passive portfolios (index funds) as a baseline performance distribution. These tests are similar in spirit to the bootstrap tests of Kosowski et al. (2006) and Fama and French (2010), but we use the index fund distribution rather than the bootstrapped distribution from active fund returns as our counterfactual distribution. We first assess differences between active and passive fund performance at various points of the distribution using quantile regressions. We then use (second-order) stochastic dominance tests to assess whether the extent of any active fund skill in aggregate is sufficient to induce a risk-averse investor to choose an active fund rather than an index fund, even before fees.

A.   Cumulative Distribution Functions

In Figure 3, we plot the cumulative distribution functions (CDFs) of gross alphas for index and active funds under the various benchmark models. We also display the returns in excess of the S&P 500 because this is a common benchmark used for equity fund performance in practice. If index funds have less dispersion in performance than active funds, then their CDF should be to the right (left) of the active funds' CDF below (above) the median.

Graph A of Figure 3 shows the surprising result that the CDF of fund returns in excess of the S&P 500 is remarkably similar for above-median funds. The largest differences in the distributions are in the left half, where index funds

FIGURE 3

Cumulative Distribution Functions of Gross α

Figure 3 plots the cumulative distribution function of the model-adjusted return, α, for index and active funds for the indicated benchmark models.
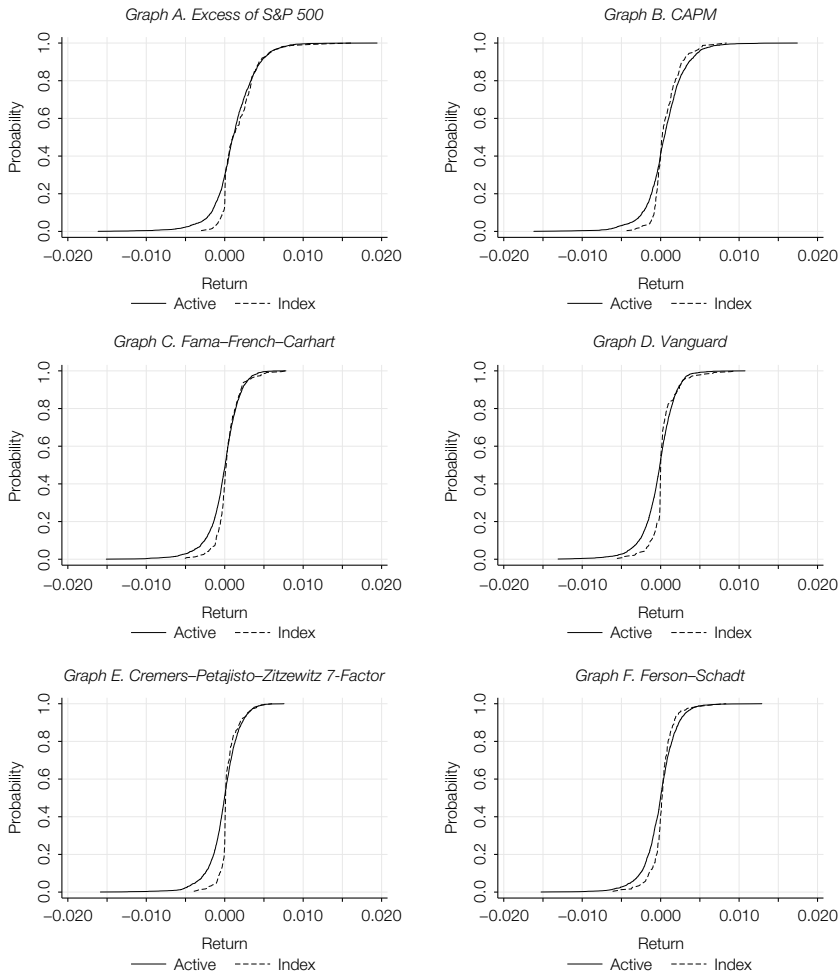
exhibit alphas much closer to 0 than those of the worst-performing active funds. When adjusted for systematic risk using benchmark models (Graphs B–F), the poorer performance of below-median active funds remains. There is substantial above-median dispersion for index funds, resulting in a smaller estimate of outperformance for the top-performing active funds relative to a comparison with the median active fund or an alpha of 0. The magnitudes are hard to address in figures; we explicitly test the statistical and economic significance of these differences in Section IV.B.

Figure 4 plots the CDFs of alpha t-statistics for index and active funds. There are some differences across benchmark models, but we see little evidence of outperformance by active funds even in the right tail. After adjusting for differences in the precision of alpha estimates due to residual variance differences, the top-performing funds appear quite similar in the portion of the distribution where one would expect to find the most skilled funds. The outperformance of index funds below the median is weaker under the t-statistic measure, but the results suggest they perform at least as well as the active funds. For the $t(\alpha)$ distribution under the S&P 500 or the CPZ7 model, the index fund distribution even lies to the right of the active fund distribution (first-order stochastic dominance).

B.   Quantile Regressions

We test for differences between index and active fund performance at various points of their respective distributions using quantile regressions (Angrist and Pischke (2008)). We analyze the 1st, 5th, 25th, 50th, 75th, 95th, and 99th percentiles of the benchmark-adjusted return distribution.[24] Statistical significance is determined by calculating bootstrapped standard errors.
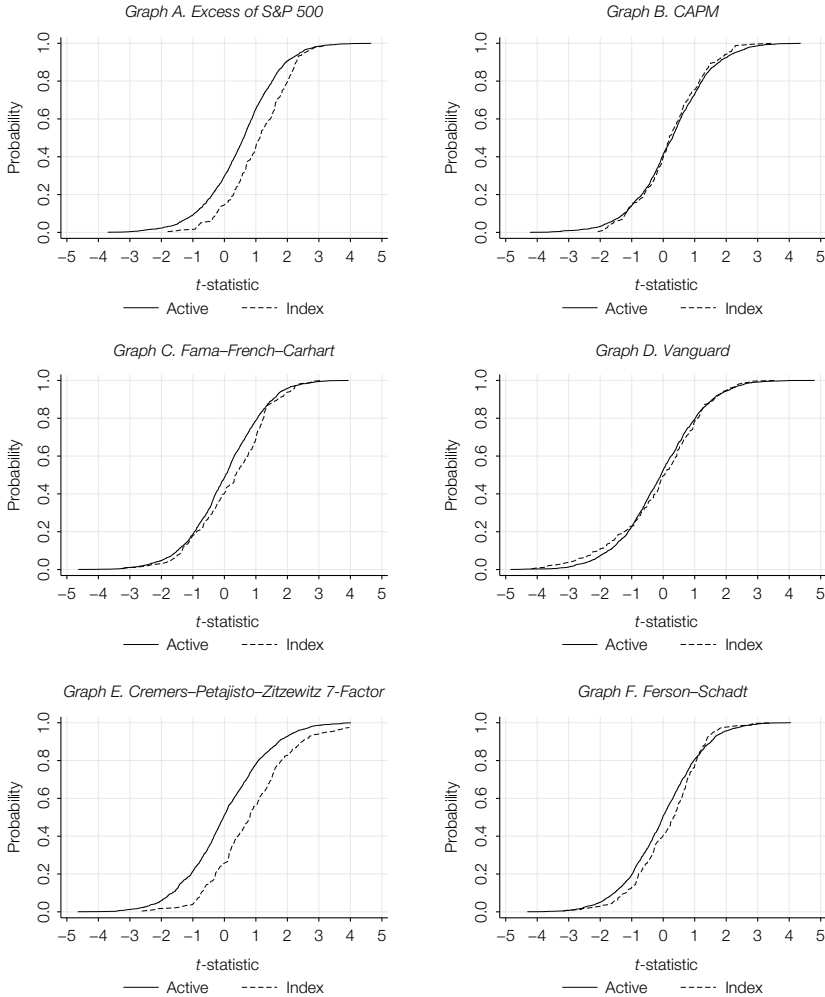
1.   Gross Alphas

Table 8 reports results from the quantile regressions of model-adjusted returns, before fees, on an index fund indicator variable. Each panel of the table presents the active fund performance (the constant) and the difference in alpha distributions between index funds and active funds (the coefficient on INDEX) under each of the different benchmark models. Each column represents a different quantile of the distribution in ascending order. For example, the middle column, Q50, represents the median active fund's benchmark-adjusted return (CONSTANT coefficient) and the difference in medians across the two distributions (INDEX coefficient). The results are consistent with the visual evidence provided by the CDFs.

For purposes of discussion, we focus mainly on the results from the distribution of 4-factor (Fama–French–Carhart) alphas. These results are presented in the third panel of Table 8. Median risk-adjusted performance for actively managed funds is approximately 1 bp per month. At the median, there is no economic or statistical difference between index funds and active funds. Even before fees, median risk-adjusted performance is similar across the two groups.

---

[24]Unlike ordinary least squares, quantile regressions can be biased when there is measurement error in the dependent variable (e.g., Hausman, Luo, and Palmer (2014)). Specifically, coefficients can be biased toward the median regression estimate. In untabulated simulations, we find that bias due to measurement error is unlikely to significantly bias the tests reported here. Simulation results are available from the authors.

FIGURE 4

Cumulative Distribution Functions of Gross $t(\alpha)$

Figure 4 plots the cumulative distribution function of t-statistics associated with the model-adjusted return, α, for index and active funds for the indicated benchmark models.

Graph A. Excess of S&P 500

Graph B. CAPM

Graph C. Fama–French–Carhart

Graph D. Vanguard

Graph E. Cremers–Petajisto–Zitzewitz 7-Factor

Graph F. Ferson–Schadt

The left tail of the distribution is where we observe the largest differences. Under the 4-factor model, we estimate that for the very worst funds (Q01), index funds outperform active funds. The estimated performance of active funds in this quantile is −73 bps per month, whereas the index funds lose only half that amount. This estimate is large economically and is statistically significant at the 1% level. The poor unconditional active fund performance in this region may represent mutual funds that produce countercyclical returns (Glode (2011)). We believe this is less likely to be true for index funds, so it is interesting that the worst index funds perform as poorly as they do. As we move along the distribution, the performance

TABLE 8

Quantile Regression Estimates: Gross Alphas

Table 8 presents quantile regression estimates from the cross section of model-adjusted mutual fund returns. Model-adjusted returns, α_i, are calculated using one of six benchmark models. Conditional quantile estimates are found by solving the following:

$$Q_\tau(\alpha_i|X_i) \quad = \quad \underset{q(X)}{\arg\min} E[\rho_\tau(\alpha_i - q(X_i))],$$

where $\rho_\tau(\mu) = (\tau - 1(\mu \le 0))\mu$ for quantile τ. To test the difference in quantiles between index funds and active funds, $q(X_i)$ is estimated as $q(X_i) = \beta_0 + \beta_1 \times INDEX_i$, where $INDEX_i$ takes a value of 1 if fund i is an index fund. The first column presents estimates for the 1st percentile of the distribution, with each subsequent column presenting estimates for the 5th, 25th, 50th, 75th, 95th, and 99th percentiles, respectively. Standard errors are bootstrapped, t-statistics are in square brackets, and * and ** indicate significance at the 5% and 1% levels, respectively.

Variable	Q01	Q05	Q25	Q50	Q75	Q95	Q99
Panel A. Excess of S&P 500							
INDEX	0.0050**	0.0022**	0.0003**	−0.0001	0.0004	0.0003	0.0021
	[5.18]	[6.30]	[5.89]	[−0.19]	[1.60]	[0.46]	[0.81]
CONSTANT	−0.0068**	−0.0030**	−0.0003**	0.0011**	0.0028**	0.0057**	0.0083**
	[−8.33]	[−11.00]	[−5.43]	[12.61]	[24.85]	[37.29]	[23.52]
Panel B. CAPM							
INDEX	0.0035**	0.0023**	0.0004**	−0.0003**	−0.0004**	−0.0006	−0.0009
	[5.09]	[5.53]	[4.27]	[−2.24]	[−2.39]	[−1.02]	[−0.70]
CONSTANT	−0.0068**	−0.0036**	−0.0008**	0.0005**	0.0019**	0.0048**	0.0076**
	[−15.28]	[−13.14]	[−11.10]	[6.64]	[29.72]	[32.03]	[14.92]
Panel C. Fama–French–Carhart							
INDEX	0.0036**	0.0020**	0.0007**	0.0001	−0.0001	−0.0000	0.0010
	[3.36]	[4.61]	[6.70]	[1.53]	[−0.61]	[−0.06]	[0.88]
CONSTANT	−0.0073**	−0.0039**	−0.0011**	0.0001	0.0011**	0.0030**	0.0043**
	[−10.37]	[−19.42]	[−17.64]	[1.54]	[19.77]	[30.12]	[21.60]
Panel D. Vanguard							
INDEX	0.0024**	0.0019**	0.0012**	0.0001**	−0.0004**	0.0002	0.0020
	[2.46]	[4.13]	[14.24]	[2.20]	[−2.99]	[0.29]	[1.21]
CONSTANT	−0.0072**	−0.0039**	−0.0013**	−0.0001**	0.0010**	0.0028**	0.0048**
	[−12.37]	[−21.59]	[−20.78]	[−2.39]	[18.70]	[29.08]	[11.72]
Panel E. Cremers–Petajisto–Zitzewitz 7-Factor							
INDEX	0.0028**	0.0027**	0.0011**	0.0001**	−0.0003**	−0.0001	0.0000
	[2.98]	[7.19]	[16.72]	[3.30]	[−2.80]	[−0.18]	[0.05]
CONSTANT	−0.0061**	−0.0037**	−0.0011**	−0.0000	0.0010**	0.0030**	0.0045**
	[−9.12]	[−18.25]	[−18.02]	[−0.80]	[21.85]	[27.36]	[22.88]
Panel F. Ferson–Schadt							
INDEX	0.0015	0.0016**	0.0008**	0.0002**	−0.0003**	−0.0008*	0.0007
	[1.47]	[3.64]	[7.23]	[2.74]	[−2.40]	[−1.68]	[0.55]
CONSTANT	−0.0067**	−0.0038**	−0.0012**	−0.0000	0.0011**	0.0032**	0.0053**
	[−15.16]	[−18.07]	[−17.57]	[−0.52]	[18.79]	[24.91]	[12.26]
No. of obs.	2,060	2,060	2,060	2,060	2,060	2,060	2,060

difference in favor of index funds gets smaller economically, but it is still large relative to the active fund performance at that point in the distribution. At the 25th percentile, active funds lose 11 bps per month, whereas index funds lose 4 bps.

Perhaps these differences in the left tail are not surprising. If index funds track passive portfolios whose benchmark performance falls in the center of the distribution, then we might expect active managers, either due to poor talent or bad luck, to do worse on the downside just by virtue of the fact that they are picking stocks. This would suggest that the active managers should then outperform by similar amounts in the right tail of the distribution. However, this is not the case.

Any performance differential (the INDEX coefficient) on the right-hand side is smaller than that on the left-hand side of the distributions.

For attractive funds, the largest performance advantages for active funds are found at the 95th and 75th percentiles. At the 75th percentile, with the exception of returns in excess of the S&P 500, there is consistent evidence across models in favor of the active funds with differences of up to 4 bps per month. The advantage is less than half the size of the underperformance of the active funds at the 25th percentile under most models. Overall, this suggests that the outperformance of active managers in the right shoulder of the distribution is more than offset by their poor performance in the left side of the distribution. We test this explicitly using stochastic dominance tests in Section IV.C. Moreover, the incremental outperformance by active funds relative to index funds substantially lowers the economic magnitude of an active fund's performance. For example, the 4-factor alpha of the 95th percentile active fund is 30 bps per month. The 95th percentile index fund also earns 30 bps per month, so any outperformance of the 95th percentile active fund is small when using the index fund distribution as a benchmark. Only under the Ferson–Schadt model is this incremental outperformance by active funds statistically significant.

2.   Gross $t(\alpha)$

Index funds and active funds have different objective functions with respect to performance. In particular, index funds are evaluated in part on their ability to tightly track their benchmark, that is, low volatility of tracking error. This likely results in low residual risk under a model-based benchmark adjustment as well. Active funds, however, try to beat their stated benchmark or passive factor portfolios, with less regard paid to residual risk. t-statistics incorporate both objectives simultaneously, allowing comparison along both dimensions of performance.

In Table 9, we report quantile regressions for the distribution of t-statistics. Incorporating residual risk into performance evaluation results in different conclusions relative to using alphas. Specifically, as discussed in Section III.E, some passive funds earn economically small positive (negative) alphas, but they do so with such precision that they have very large positive (negative) t-statistics. This leads to differences in the relative performance of active and passive funds. In particular, there is no significant outperformance by active funds, consistent with the graphical evidence in Figure 4. In the right half of the distribution, the magnitudes of any differences in $t(\alpha)$ are small, and the magnitude and signs of the estimates differ depending on the benchmark model choice. Overall, it would be difficult to conclude that the actively managed funds perform substantially better in the right tail of the distribution when considering performance per unit residual risk. On the other hand, the worst index funds also no longer strongly outperform the worst active funds, as they do when using alphas. This is primarily due to the fact that the poor performance of the worst index funds, although smaller in magnitude, is very precisely estimated. Therefore, adjusting for residual risk, the index funds' advantage in the left tail is reduced.

TABLE 9

Quantile Regression Estimates: Gross t-Statistics

Table 9 presents quantile regression estimates from the cross section of t-statistics of model-adjusted mutual fund returns. Model-adjusted returns, α_i, are calculated using one of six benchmark models. Conditional quantile estimates are found by solving the following:

$$Q_\tau(t(\alpha_i)|X_i) \quad = \quad \underset{q(X)}{\arg\min}\, \mathrm{E}[\rho_\tau(t(\alpha_i) - q(X_i))],$$

where $\rho_\tau(\mu) = (\tau - 1(\mu \le 0))\mu$ for quantile τ. To test the difference in quantiles between index funds and active funds, $q(X_i)$ is estimated as $q(X_i) = \beta_0 + \beta_1 \times \text{INDEX}_i$, where INDEX_i takes a value of 1 if fund i is an index fund. The first column presents estimates for the 1st percentile of the distribution, with each subsequent column presenting estimates for the 5th, 25th, 50th, 75th, 95th, and 99th percentiles, respectively. Standard errors are bootstrapped, t-statistics are in square brackets, and * and ** indicate significance at the 5% and 1% levels, respectively.

Variable	Q01	Q05	Q25	Q50	Q75	Q95	Q99
Panel A. Excess of S&P 500							
INDEX	1.21**	0.70**	0.62**	0.52**	0.51**	0.20	2.12
	[3.66]	[3.44]	[6.68]	[5.78]	[5.14]	[1.28]	[1.17]
CONSTANT	−2.59**	−1.45**	−0.17**	0.61**	1.33**	2.40**	3.26**
	[−23.29]	[−29.43]	[−3.72]	[18.75]	[31.04]	[46.67]	[23.51]
Panel B. CAPM							
INDEX	0.94**	0.17	0.08	−0.09	−0.10	−0.19	−0.58
	[4.25]	[0.99]	[0.69]	[−1.04]	[−0.78]	[−1.29]	[−1.46]
CONSTANT	−2.84**	−1.75**	−0.46**	0.29**	1.06**	2.28**	3.17**
	[−13.71]	[−24.07]	[−11.51]	[8.49]	[31.02]	[30.04]	[27.45]
Panel C. Fama–French–Carhart							
INDEX	0.31	0.28	0.16	0.33**	0.28**	0.27	−0.06
	[0.80]	[1.44]	[1.19]	[2.92]	[3.66]	[1.95]	[−0.23]
CONSTANT	−3.05**	−1.96**	−0.76**	0.05	0.85**	1.92**	2.83**
	[−23.58]	[−24.70]	[−18.62]	[1.18]	[22.04]	[31.12]	[31.24]
Panel D. Vanguard							
INDEX	−0.62*	−0.51	0.02	0.12	0.16	−0.05	−0.07
	[−2.16]	[−1.84]	[0.15]	[0.94]	[1.43]	[−0.29]	[−0.19]
CONSTANT	−3.10**	−2.25**	−0.92**	−0.07*	0.79**	2.11**	2.88**
	[−23.66]	[−32.28]	[−24.90]	[−2.10]	[15.59]	[30.39]	[19.52]
Panel E. Cremers–Petajisto–Zitzewitz 7-Factor							
INDEX	0.94*	1.13**	0.82**	0.84**	0.73**	1.16**	4.05**
	[2.40]	[8.00]	[5.52]	[8.55]	[7.24]	[2.72]	[3.57]
CONSTANT	−3.11**	−2.08**	−0.85**	−0.03	0.87**	2.25**	3.18**
	[−25.96]	[−32.77]	[−17.01]	[−0.77]	[22.48]	[34.13]	[19.81]
Panel F. Ferson–Schadt							
INDEX	0.24	0.41*	0.27*	0.35**	0.14	−0.25	−0.04
	[0.65]	[2.15]	[2.12]	[3.76]	[1.56]	[−1.63]	[−0.10]
CONSTANT	−2.84**	−2.01**	−0.83**	−0.02	0.79**	1.90**	2.83**
	[−27.49]	[−28.83]	[−21.76]	[−0.57]	[24.52]	[29.32]	[25.97]
No. of obs.	2,060	2,060	2,060	2,060	2,060	2,060	2,060

3. Benchmark-Adjusted Returns

We also examine the distributions of performance for the subsets of active and index funds for which we have the returns of the stated benchmark (S&P and Russell benchmarked funds listed in footnote 20). In Table 10, we present the distribution of returns in excess of the stated benchmark and the t-statistics of those excess returns. Active funds may take more risk than their stated benchmark and may even game their benchmark choice (Sensoy (2009)), so it is not surprising to see that the distribution of active performance relative to the stated benchmark is much wider than that of the index funds.

TABLE 10

Quantile Regression Estimates: S&P and Russell Funds

Table 10 presents quantile regression estimates from the cross section of gross average returns in excess of the stated benchmark and t-statistics of these excess returns for funds benchmarked to 17 Standard & Poor's (S&P) and Russell indices (listed in footnote 20). Conditional quantile estimates are found by solving the following:

$$Q_\tau(y_i|X_i) = \operatorname*{arg\,min}_{q(X)} E[\rho_\tau(y_i - q(X_i))],$$

where $\rho_\tau(\mu) = (\tau - 1(\mu \le 0))\mu$ for quantile τ of performance measure y_i. To test the difference in quantiles between index funds and active funds, $q(X_i)$ is estimated as $q(X_i) = \beta_0 + \beta_1 \times \text{INDEX}_i$, where INDEX_i takes a value of 1 if fund i is an index fund. The first column presents estimates for the 1st percentile of the distribution, with each subsequent column presenting estimates for the 5th, 25th, 50th, 75th, 95th, and 99th percentiles, respectively. Standard errors are bootstrapped, t-statistics are in square brackets, and * and ** indicate significance at the 5% and 1% levels, respectively.

Variable	Q01	Q05	Q25	Q50	Q75	Q95	Q99
Panel A. Gross Average Return in Excess of Stated S&P or Russell Benchmark							
INDEX	0.0061**	0.0031**	0.0007**	−0.0004**	−0.0014**	−0.0035**	−0.0050**
	[8.18]	[13.49]	[7.81]	[−7.51]	[−18.05]	[−19.59]	[−9.17]
CONSTANT	−0.0065**	−0.0033**	−0.0006**	0.0004**	0.0015**	0.0037**	0.0057**
	[−9.21]	[−16.18]	[−7.73]	[8.25]	[19.37]	[27.73]	[11.94]
Panel B. Gross t-Statistic of Return in Excess of Stated S&P or Russell Benchmark							
INDEX	0.69	0.28	0.51**	0.73**	0.88**	1.97	4.28*
	[1.73]	[1.15]	[3.01]	[4.96]	[6.51]	[1.30]	[2.42]
CONSTANT	−2.75**	−1.68**	−0.43**	0.31**	1.00**	2.03**	2.75**
	[−15.03]	[−26.45]	[−10.24]	[10.08]	[23.16]	[39.53]	[25.76]
No. of obs.	1,479	1,479	1,479	1,479	1,479	1,479	1,479

For performance measured in excess of the stated benchmark, an appropriate comparison requires scaling the excess return by the residual risk. The bottom panel of Table 10 shows that the t-statistics for the index fund distribution are at least as large as those of the active fund distribution at each quantile tested. More importantly, the most extreme positive t-statistics are significantly greater for index funds. On a per-unit-of-risk basis, even under a simple benchmark adjustment, active funds underperform index funds.

C.   Stochastic Dominance Tests

Our quantile regression results are useful in assessing how percentiles of the index and active fund distributions compare, but assessing the aggregate amount of incremental skill in the active space is challenging due to multiple comparison issues. To overcome these issues, we ask whether there exists enough incremental performance in the active fund space to warrant a risk-averse investor choosing an active fund over an index fund. We answer this question by testing the null that active funds stochastically dominate index funds and vice versa. We can easily reject the nulls that either distribution dominates the other in a first-order sense (i.e., the distributions cross), so we focus our discussion on second-order stochastic dominance.[25]

It is not obvious that gross index fund performance should second-order stochastically dominate that of active funds. Consider the alpha distribution. The index fund distribution has less variance than the active distribution while exhibiting a higher average alpha. However, neither distribution is symmetric. Although mean-variance investors would obviously prefer the index fund distribution with

---

[25]Recall that distribution G stochastically dominates distribution F in a second-order sense if the area under the CDF G is less than the area under the CDF F for all values of the outcome space.

its higher mean and lower variance, some investors may prefer possible skewness in the active fund distribution. Indeed, documented preferences for skewness for individual investors could make active funds attractive if they exhibit sufficient upside potential. Thus, understanding whether the upside of the active funds is economically sufficient to outweigh their underperformance on the downside is important.

Our tests of stochastic dominance follow the bootstrap-based test of Barrett and Donald (2003), which is based on Kolmogorov–Smirnov tests comparing the distributions at all points.[26] The intuition of the test of the null that distribution G second-order stochastically dominates distribution F is that a number of bootstrapped draws from F will generate a distribution of CDFs that can be compared with the empirical distribution G. The deviations between F and G (across all points) can be compared with those found between F and the bootstrapped CDFs to determine the likelihood of observing the deviations between F and G by chance.

Table 11 displays p-values associated with tests of stochastic dominance for the distributions of alpha and $t(\alpha)$. For alphas, we cannot reject the null that index funds dominate active funds, but we can strongly reject the null that active funds dominate index funds for all benchmark models. When using $t(\alpha)$, we again cannot reject the null that index funds dominate active funds for any benchmark model. For 4 of the 6 benchmarks, we can again reject the null that active funds dominate index funds.

Economically, the results indicate that the magnitude of any active fund outperformance relative to index funds is insufficient to outweigh the active funds' underperformance relative to the worst index funds. A risk-averse investor

---

TABLE 11
Stochastic Dominance Tests

Table 11 presents tests of second-order stochastic dominance (SD2) of index funds versus active funds. p-values are tabulated from bootstrap tests of the null hypothesis that the index fund gross performance distribution dominates the active fund gross performance distribution and vice versa. The bootstrap tests follow Barrett and Donald (2003), and each contains 1,000 draws.

Model	Null Tested	
	Index SD2 Active	Active SD2 Index
Panel A. Alphas		
Excess of S&P 500	0.868	0.000
CAPM	0.894	0.000
Fama–French–Carhart	0.819	0.000
Vanguard	0.807	0.000
CPZ 7-Factor	0.832	0.000
Ferson–Schadt	0.831	0.000
Panel B. $t(\alpha)$		
Excess of S&P 500	0.904	0.000
CAPM	0.659	0.262
Fama–French–Carhart	0.890	0.003
Vanguard	0.228	0.750
CPZ 7-Factor	0.896	0.000
Ferson–Schadt	0.880	0.001

---

[26]Specifically, we report the KSB1 p-values.

facing a random draw from the distribution of active funds versus the distribution of index funds should prefer the index fund lottery based on benchmark-adjusted performance, even before fees. In the Internet Appendix, we report additional stochastic dominance tests comparing index funds to more active funds based on active share (Cremers and Petajisto (2009)) or return gap (Kacperczyk et al. (2008)). Consistent with prior work, these active funds perform better, but the stochastic dominance tests lead to similar conclusions.

V.   Conclusion

The number of index funds and underlying indices has grown tremendously over the last 20 years. We exploit this growth to extend the intuitive comparison of passive and active fund performance to distributional tests. These tests imply that index fund skill exists, is persistent, and is found in proportions similar to those found in active funds. Additionally, investors seem to recognize differences in performance; index funds with larger past gross returns receive greater capital flows. We also document the underlying sources of dispersion in index fund performance. Heterogeneity in underlying benchmarks drives dispersion in alphas, whereas heterogeneity in tracking error contributes to dispersion in $t(\alpha)$.

Dispersed performance of passive funds has important implications for the evaluation of active managers. We use the index fund performance distribution as a set of counterfactual funds to assess incremental performance in the active fund distribution. When performance is measured using before-fee model alphas and compared across the cross-sectional distribution, any active fund performance advantage is substantially less than one would conclude from benchmarking to average index fund performance. Moreover, any advantage of the top active managers over the top index funds is much less than the advantage of the worst index funds over the worst active funds. When performance accounts for residual risk, active funds no longer outperform index funds. Stochastic dominance tests show that index funds dominate active funds in a second-order sense for either alphas or $t(\alpha)$, indicating that no risk-averse investor should prefer a random draw from the active fund distribution to one from the index fund distribution.

Our results highlight the importance and challenges of choosing an appropriate benchmark for managed funds. Should performance evaluation give asset managers credit for selecting a benchmark index that outperforms the market? If one uses alphas from benchmark models, then the implicit answer is yes, and passive funds appear attractive relative to active funds. If one answers no, then any outperformance by index funds relative to their underlying indices is economically small, but investors and researchers are left with the challenge of taking a stand on the unobservable true benchmark return for an active fund. Although more work is required to fully address this challenge, future work on benchmarking performance of managers should consider the implications of dispersion in passive fund performance.

References

Angrist, J. D., and J.-S. Pischke. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press (2008).

Barras, L.; O. Scaillet; and R. Wermers. "False Discoveries in Mutual Fund Performance: Measuring Luck in Estimated Alphas." *Journal of Finance*, 65 (2010), 179–216.

Barrett, G. F., and S. G. Donald. "Consistent Tests for Stochastic Dominance." *Econometrica*, 71 (2003), 71–104.

Berk, J., and J. H. van Binsbergen. "Measuring Skill in the Mutual Fund Industry." *Journal of Financial Economics*, 118 (2015), 1–20.

Berk, J., and J. H. van Binsbergen. "Assessing Asset Pricing Models Using Revealed Preference." *Journal of Financial Economics*, 119 (2016), 1–23.

Berk, J., and R. Green. "Mutual Fund Flows and Performance in Rational Markets." *Journal of Political Economy*, 112 (2004), 1269–1295.

Blocher, J., and R. E. Whaley. "Two-Sided Markets in Asset Management: Exchange-Traded Funds and Securities Lending." Working Paper, Vanderbilt University (2016).

Bollen, N. B. P., and J. A. Busse. "On the Timing Ability of Mutual Fund Managers." *Journal of Finance*, 56 (2001), 1075–1094.

Carhart, M. M. "On Persistence in Mutual Fund Performance." *Journal of Finance*, 52 (1997), 57–82.

Chen, H.-L.; N. Jegadeesh; and R. Wermers. "The Value of Active Mutual Fund Management: An Examination of the Stockholdings and Trades of Fund Managers." *Journal of Financial and Quantitative Analysis*, 35 (2000), 343–368.

Chevalier, J., and G. D. Ellison. "Risk Taking by Mutual Funds as a Response to Incentives." *Journal of Political Economy*, 105 (1997), 1167–1200.

Choi, J. J.; D. Laibson; and B. C. Madrian. "Why Does the Law of One Price Fail? An Experiment on Index Mutual Funds." *Review of Financial Studies*, 23 (2010), 1405–1432.

Cremers, K. J. M., and A. Petajisto. "How Active Is Your Fund Manager? A New Measure That Predicts Performance." *Review of Financial Studies*, 22 (2009), 3329–3365.

Cremers, M.; A. Petajisto; and E. Zitzewitz. "Should Benchmark Indices Have Alpha? Revisiting Performance Evaluation." *Critical Finance Review*, 2 (2013), 1–48.

Daniel, K.; M. Grinblatt; S. Titman; and R. Wermers. "Measuring Mutual Fund Performance with Characteristic-Based Benchmarks." *Journal of Finance*, 52 (1997), 1035–1058.

Del Guercio, D., and J. Reuter. "Mutual Fund Performance and the Incentive to Generate Alpha." *Journal of Finance*, 69 (2014), 1673–1704.

Elton, E. J.; M. J. Gruber; and C. R. Blake. "The Persistence of Risk-Adjusted Mutual Fund Performance." *Journal of Business*, 69 (1996), 133–157.

Elton, E. J.; M. J. Gruber; and J. A. Busse. "Are Investors Rational? Choices among Index Funds." *Journal of Finance*, 59 (2004), 261–288.

Elton, E. J.; M. J. Gruber; S. Das; and M. Hlavka. "Efficiency with Costly Information: A Reinterpretation of Evidence from Managed Portfolios." *Review of Financial Studies*, 6 (1993), 1–22.

Evans, R. B. "Mutual Fund Incubation." *Journal of Finance*, 65 (2010), 1581–1611.

Fama, E. F., and K. R. French. "Common Risk Factors in the Returns on Stocks and Bonds." *Journal of Financial Economics*, 33 (1993), 3–56.

Fama, E. F., and K. R. French. "Luck versus Skill in the Cross-Section of Mutual Fund Returns." *Journal of Finance*, 65 (2010), 1915–1947.

Ferson, W. E., and Y. Chen. "How Many Good and Bad Fund Managers Are There, Really?" Working Paper, University of Southern California (2015).

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

Glode, V. "Why Mutual Funds Underperform." *Journal of Financial Economics*, 99 (2011), 546–559.

Grinblatt, M., and S. Titman. "Mutual Fund Performance: An Analysis of Quarterly Portfolio Holdings." *Journal of Business*, 62 (1989), 393–416.

Grinblatt, M., and S. Titman. "The Persistence of Mutual Fund Performance." *Journal of Finance*, 47 (1992), 1977–1984.

Grinblatt, M., and S. Titman. "Performance Measurement without Benchmarks: An Examination of Mutual Fund Returns." *Journal of Business*, 66 (1993), 47–68.

Gruber, M. J. "Another Puzzle: The Growth in Actively Managed Mutual Funds." *Journal of Finance*, 51 (1996), 783–810.

Hausman, J.; Y. Luo; and C. Palmer. "Errors in the Dependent Variable of Quantile Regression Models." Working Paper, Massachusetts Institute of Technology (2014).

Hortacsu, A., and C. Syverson. "Product Differentiation, Search Costs, and Competition in the Mutual Fund Industry: A Case Study of S&P 500 Index Funds." *Quarterly Journal of Economics*, 119 (2004), 403–456.

Hunter, D.; E. Kandel; S. Kandel; and R. Wermers. "Mutual Fund Performance Evaluation with Active Peer Benchmarks." *Journal of Financial Economics*, 112 (2014), 1–29.

Investment Company Institute. "2015 Investment Company Fact Book." White Paper (2015).

Jensen, M. C. "The Performance of Mutual Funds in the Period 1945–1964." *Journal of Finance*, 23 (1968), 389–416.

Jiang, G. J.; T. Yao; and T. Yu. "Do Mutual Funds Time the Market? Evidence from Portfolio Holdings." *Journal of Financial Economics*, 86 (2007), 724–758.

Jiang, H.; M. Verbeek; and Y. Wang. "Information Content When Mutual Funds Deviate from Benchmarks." *Management Science*, 60 (2014), 2038–2053.

Kacperczyk, M.; C. Sialm; and L. Zheng. "Unobserved Actions of Mutual Funds." *Review of Financial Studies*, 21 (2008), 2379–2416.

Kacperczyk, M.; S. van Nieuwerburgh; and L. Veldkamp. "Time-Varying Fund Manager Skill." *Journal of Finance*, 69 (2014), 1455–1484.

Keim, D. B. "An Analysis of Mutual Fund Design: The Case of Investing in Small-Cap Stocks." *Journal of Financial Economics*, 51 (1999), 173–194.

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

Malkiel, B. G. "Returns from Investing in Equity Mutual Funds 1971 to 1991." *Journal of Finance*, 50 (1995), 549–572.

Pastor, L.; R. Stambaugh; and L. Taylor. "Scale and Skill in Active Management." *Journal of Financial Economics*, 116 (2015), 23–45.

Petajisto, A. "Active Share and Mutual Fund Performance." *Financial Analysts Journal*, 69 (2013), 73–93.

Sensoy, B. A. "Performance Evaluation and Self-Designated Benchmark Indexes in the Mutual Fund Industry." *Journal of Financial Economics*, 92 (2009), 25–39.

Sirri, E., and P. Tufano. "Costly Search and Mutual Fund Flows." *Journal of Finance*, 53 (1998), 1589–1622.

Storey, J. D. "A Direct Approach to False Discovery Rates." *Journal of the Royal Statistical Society*, 64 (2002), 479–498.

Wermers, R. "Mutual Fund Performance: An Empirical Decomposition into Stock-Picking Talent, Style, Transactions Costs, and Expenses." *Journal of Finance*, 55 (2000), 1655–1703.