Mutual Fund Performance at Long Horizons*

Hendrik Bessembinder
W. P. Carey School of Business,
Arizona State University

Michael J. Cooper
David Eccles School of Business,
University of Utah

Feng Zhang
Cox School of Business,
Southern Methodist University

This Draft: April 2022

* The authors thank Stephen Brown, Mikhail Chernov, Magnus Dahlquist, Wayne Ferson, Chris Hrdlicka, Paul Irvine, Haim Levy, Lars Lochstoer, Zhongjin (Gene) Lu, Seth Pruitt, Clemens Sialm, Harald Uhlig, Sunil Wahal, Morad Zekhnini, and seminar participants at the Virtual Asset Management Seminar, Arizona State University, Fordham University, University of Nebraska, Georgetown University, Rutgers University, Southern Methodist University, Stockholm School of Economics, Michigan State University, the University of Wyoming, the University of Utah and the University of Georgia for valuable comments and discussions on earlier versions of this paper. We also thank Professor Shuaiyu Chen for his assistance in matching CRSP mutual fund returns to those contained in the Morningstar database. Some results in this paper were contained in an earlier working paper titled "Return Horizon and Mutual Fund Performance."

Mutual Fund Performance at Long Horizons*

Abstract

The percentage of U.S. equity mutual funds that outperform the SPY ETF decreases substantially as the horizon over which returns are measured is increased. Further, some funds with positive monthly alpha estimates have negative compound long-horizon abnormal returns. These results reflect positive skewness in compound fund returns and highlight the limitations of conditional arithmetic means of short-horizon returns (including alpha). We tabulate pre-fee investor wealth enhancement of $1.23 trillion and fees of $2.26 trillion, or a wealth loss of $1.02 trillion to mutual fund investors in aggregate over our 30-year sample, when opportunity costs are based on SPY returns.

(*JEL* G10, G23)

Mutual Fund Performance at Long Horizons

1. Introduction

The majority of the research that considers investor outcomes reports on unconditional or conditional (as in alpha estimates) arithmetic means of returns that are measured over relatively short horizons, most often monthly. In contrast, investment and decision horizons can stretch to decades, and differ across investors.[1] We posit that many investors are concerned with the compound returns that accrue over longer horizons, propose that empirical measures of investment performance should therefore consider a variety of return measurement horizons, and report on compound returns to U.S. equity mutual funds from 1991 to 2020 at the monthly, annual, decade, and full sample horizons.[2] The results verify that compound long-horizon returns often contain important information that is not readily apparent in the distribution of short-horizon returns.

The literature that studies mutual fund return performance is vast, and numerous important empirical regularities have been documented (see Cremers, Fulkerson, and Riley, 2019, for a recent survey). However, these studies, like the broader literature, have mainly focused on return data measured over short (usually monthly) horizons.[3] Here, we study the frequency with which individual funds outperform benchmarks in terms of compound returns over various horizons, clarifying the respective

---

[1] For example, Ameriks and Zeldes (2004) report that nearly half of participants in a sample of defined contribution retirement plans made no changes to their allocations over a ten-year period.

[2] Our focus on return horizon is not unprecedented, but the prior literature has focused mainly on relatively short horizons ranging from daily to annual, where informational and trading frictions are most relevant, and has not considered mutual funds. Gilbert, Hrdlicka, Kalodimos, and Siegel (2014) estimate alphas and betas for equity portfolios over horizons ranging from daily to quarterly, and argue that differences across horizon are explained by differences in firms' opacity, i.e., in investors difficulty in assessing the value implications of events. Boguth, Carlson, Fisher and Simutin (2016) focus on slow information diffusion as an explanation for differing mean equity portfolio returns for horizons ranging from daily to annual. Kamara, Korajczyk, Lou, and Sadka (2016) also focus on heterogeneous stock price reactions and assess the extent to which systematic factors earn risk premia at some horizons from monthly to biannual, but not others. Kothari, Shanken, and Sloan (1995) estimate a positive return premium associated with CAPM betas when returns are measured at the annual horizon, but not at the monthly horizon. Our study differs from these not only because we study mutual fund returns while the prior studies consider equities, but because we focus on returns measured over longer horizons where these frictions are less important, to highlight the effect of horizon *per se*.

[3] An exception is the contemporaneous analysis of Bessembinder, Cooper, and Zhang (2022), who study estimated mutual fund alphas for returns that are measured at horizons ranging from monthly to decadal.

roles of return skewness and mutual fund expenses, and we tally the full-sample dollar gain or loss to mutual fund investors on a fund-by-fund basis and in aggregate.

We study a broad sample of nearly 8,000 U.S. equity mutual funds during the 1991 to 2020 period. While specific point estimates would differ if we considered alternative samples, e.g. non-domestic, balanced, or levered funds, the central point we emphasize, that compound long-horizon returns contain different information as compared to means of short horizons returns, is unlikely to be sample-specific. We show that the percentage of funds that outperform market benchmarks decreases with horizon over which returns are measured. In the monthly data, equity mutual fund returns exceed the matched-month return to the SPY ETF (taken as a proxy for the overall market that investors could readily have captured) for 47.2% of observations. The percentage of sample funds that generate buy-and-hold returns that exceed buy-and-hold returns to the SPY decreases to 41.1% at the annual horizon, 38.3% at the decade horizon, and 30.3% at the full-sample horizon. While the sample contains many small funds, these do not drive the results. When we focus on a subsample of the largest funds we find that only 29.6% outperform the SPY in terms of compound full-sample returns.

Rather, these results reflect a prominent dimension by which long-horizon returns contain different information than short-horizon returns: the cross-sectional distribution of long-horizon fund buy-and-hold returns is strongly positively skewed, while such skewness is not observable in monthly returns. This positive skewness in compound long-horizon returns is of substantial practical importance. Financial planning (e.g. at pension funds) is often based on assumptions regarding mean returns. Aside from the active debate as to whether the assumed mean returns are appropriate, in a positively skewed distribution a potentially large majority of possible future realizations are less than the mean outcome.

Of course, while positive skewness implies that many funds underperform, some funds perform very well. Out of 7,883 sample funds, 442 delivered a positive full-sample compound return more than twice as large as the compound return to the SPY over the matched months, and 160 delivered compound returns three times as large as the SPY during the matched months of the full sample.

While, as noted, the academic literature mainly studies the distribution of fund returns measured at the monthly horizon, the SEC requires mutual funds to report compound returns for the most recent 1, 5, and 10-year periods, on both a pre- and post-tax basis.[4] Further, S&P Dow Jones Indices produces well-publicized annual "scorecards" that compare compound fund returns to compound S&P 1500 index outcomes for horizons of up to twenty years. However, as we document, these comparisons are biased by the fact that funds that outperform the index but nevertheless exit the sample are counted as underperforming rather than outperforming.

We assess the role of fund fees by considering long-horizon outcomes in pre-fee returns. Consistent with the short-return-horizon evidence reported by Berk and van Binsbergen (2015) and Fama and French (2010), we find that *mean* pre-fee mutual fund returns exceed returns to market benchmarks at long horizons. The cross-fund mean full sample pre-fee buy-and-hold return to sample funds is 394%, while the mean SPY buy-and-hold return over matched periods is 298%. Nevertheless, only a minority (45.2%) of individual funds outperform the SPY in the long run even in a comparison of pre-fee fund returns to SPY returns that are net of fees.

The positive skewness in lifetime fund returns is attributable in part to differences in fund lives. Though our sample period spans thirty years, on average funds are contained in the database for just eleven years. Further, funds with worse performance tend to have shorter lives, implying that cross-sectional averages of fund-specific performance measures are affected by what Linnainmaa (2013) refers to as "reverse survivorship bias." To address this concern, we study outcomes to bootstrapped portfolios of mutual fund returns. We find that the percentage of bootstrapped mutual fund portfolios that outperform the SPY during the full thirty-year sample decreases from 47.5% at the monthly horizon to 5.5% at the 30-year horizon. These results indicate that our conclusions regarding the effect of return measurement horizon are robust to allowances for endogenous variation in fund lives.

[4] See https://www.sec.gov/rules/final/33-7941.htm.

We also assess the potential effects of funds' systematic risk exposures in explaining long-horizon outcomes. We rely on a simple single-factor market model because our main focus is on the effects of the compounding of random returns over long horizons, not on the widely-studied question of which benchmarks or factor models are most appropriate. In particular, we estimate "non-market" returns for each fund and month as the sum of the fund's estimated monthly alpha and the market-model residual for the month. We show that the compound non-market return is negative for the majority of funds. We also show the compound non-market return is negative for 12% of funds with positive monthly alpha estimates. Similarly, we show that about one of every six funds in our sample that has a positive monthly arithmetic mean market-adjusted return also has negative lifetime market-adjusted buy-and-hold returns. These outcomes reflect the well-known fact that the arithmetic mean exceeds the geometric mean in any return sample with positive volatility. Yet, the literature emphasizes conditional or unconditional arithmetic means of short-horizon returns, not only in the form of alpha estimates, but also when focusing on Sharpe ratios, fitted values from Fama-MacBeth or factor model regressions, and comparisons of average returns across portfolios. We document the extent to which this emphasis can be misleading for long-term investors.

We find that over two thirds of U.S. equity mutual funds underperform the post-fee SPY in terms of compound returns during the 1991-2020 period, and that over 20% of funds fail to even outperform one-month U.S. Treasury Bills during the sample period. While some funds perform very well, the net economic impact of this long-term underperformance is large. Using the SPY return as the opportunity cost and allowing for variation in beta estimates, we document an aggregate wealth loss to mutual fund investors of $1.02 trillion, as fund manager skill of $1.23 trillion is not sufficient to overcome fund expenses totaling $2.26 trillion, each measured as of the end of our 1991-2020 sample period.

2. Sample Overview and Predictions

2.1 The Monthly Mutual Fund Sample

We obtain data for the 1991 to 2020 period from the CRSP survivorship bias free Mutual Fund Database. We begin at 1991, as data regarding fund total net assets (TNA), which we use to aggregate fund returns across share classes, is not consistently available for earlier periods. We study domestic equity funds while excluding ETFs, target date funds, hedged funds, and leveraged funds. Specific fund filters are described in Appendix A. Prior studies (e.g., Elton, Gruber, and Blake, 2001) have documented the presence of errors in the CRSP mutual fund data. We identify and correct specific potentially influential errors, and omit a small number of funds with apparent data errors that we are not able to correct or verify from alternative sources, as described in Appendix A. We also exclude funds that have fewer than twelve months of non-missing return data.

Value-weighted market returns comprise a natural and widely-used benchmark. However, as noted by Pastor and Stambaugh (2012) and Berk and van Binsbergen (2015), investors cannot directly capture the value-weighted market return or returns to equity indices, since trades are required to initially enter and to exit positions, and also at the times of primary transactions, including dividends, stock repurchases, or new equity issues. While we report some outcomes relative to the value-weighted market, we focus on the SPY ETF as our primary market benchmark.[5] Since SPY returns are net of any fees, trading costs, or other expenses, investors could in principle have captured compound SPY returns with a simple buy-and-hold strategy with dividend reinvestment.

Table 1 presents summary statistics regarding the sample, which contains 7,883 domestic equity mutual funds. Of these, 525 are index funds. The sample includes 1,048,111 fund/months. The pooled (across funds and years) mean monthly fund return (net of fees) is 0.776%, while the mean monthly fee is

[5] The SPY ETF started trading in January of 1993. For 1991 and 1992, we use the return on the Vanguard S&P500 index fund (ticker symbol VFINX) instead.

0.095%.[6] The pooled mean of matched SPY returns is 0.835%, while the pooled mean of the matched value-weighed market returns is 0.882%. Mean TNA is $1.177 billion. However, the TNA distribution is strongly positively skewed, reflecting the presence of very large funds, and the sample median TNA is $149 million. In contrast, the pooled distribution of monthly fund returns is not strongly skewed; the sample skewness coefficient is -0.425, and the median return of 1.16% is greater than the mean return of 0.78%. Panel A of Figure 3 displays the frequency distribution of monthly returns, and gives visual indication of slight negative skewness, with marginally more observations in the vicinity of negative ten to twenty percent as compared to positive ten to twenty percent.

Figure 1 displays the number of funds contained in the sample and total TNA for sample funds on an annual basis. The number of domestic equity mutual funds increased rapidly from about 1,000 in 1991 to over 3,400 in 2002, remaining relatively constant until 2007, before expanding to approximately 4,300 in 2008. Sample funds' aggregate TNA not only rose rapidly in the early years of the sample period, from about $300 billion in 1991 to approximately $2.8 trillion in 2000, but continued to increase thereafter, to approximately $9.5 trillion in 2020.

To assess the performance of mutual funds at various horizons we compute the buy-and-hold return to the fund, obtained by compounding monthly fund returns. Since the return data includes any dividends or other cash distributions, the buy-and-hold return implicitly assumes that dividends and distributions are reinvested in fund shares at a price equal to the month-end net asset value.[7] In section 5 we report results that do not rely on the reinvested dividend assumption, and that also allow for investor flows. For comparison, we compute buy-and-hold returns to one-month US Treasury Bills, to the value-

---

[6] The associated CRSP variable name is *exp_ratio*, which is reported on an annual basis. The data as reported contain some errors (the maximum reported fee is 146%), and a few (about 0.2%) observations are missing. We replace missing observations and those that exceed 10% per year with the median expense ratio for sample funds during the year. We divide by 12 to convert the annual expense ratio to a monthly figure, and refer to the variable as the fund fee.

[7] Fama (1972) notes that the assumption that dividends and other distributions are reinvested is desirable when measuring performance over intervals that are longer than the elapsed time between such distributions, because of the implicit assumption that funds invested at the beginning of the sample remain invested throughout. However, he also notes that the approach is "less pure" than some alternatives, because it assumes a reinvestment policy "not followed in the (mutual fund) portfolio."

weighted market (obtained from Professor Kenneth French's website), and to the SPY ETF. The periods over which benchmark returns are computed are always matched to fund returns. If, for example, a given fund has return data for 105 months during a given decade then benchmark returns for that fund and decade are computed based on the same 105 months. We also follow Loughran and Ritter (1995) in computing for horizons longer than one month a "wealth ratio" for each fund, as one plus the fund buy-and-hold return divided by one plus the benchmark buy-and-hold return. The wealth ratio is the accumulated value of a given initial investment in the fund relative to the accumulated value of the same initial investment had it earned benchmark returns instead.

We construct indicator variables that equal one when a given fund's buy-and-hold return outperforms a benchmark's buy-and-hold return over a specified time period, and zero otherwise. Table 1 reports the means of these indicator variables in the pooled monthly data. Only a slight majority (60.2%) of the fund/month return observations exceed the one-month Treasury bill return in the same month, reflecting the high volatility of equity returns. A slight minority (46.3%) of fund-month returns exceed the value-weighed market return during the same month, while 47.2% of fund-month returns exceed the SPY ETF return during the same month.

2.2 Predictions Regarding Compound Fund Returns as Compared to Benchmarks

Our goal is to investigate relations between measures of mutual fund performance and the time horizon over which returns are measured. The majority of the existing literature focuses on monthly returns and constructs performance measures, e.g. Sharpe ratios, alphas, etc. that rely on the conditional or unconditional arithmetic mean of those monthly returns. Unfortunately, theory provides relatively little in the way of predictions as to how performance measures constructed from compound returns will differ from those based on short-horizon returns. One exception is Farago and Hjalmarsson (2021), who, assuming iid short-horizon returns, develop a closed form expression for the skewness of compound returns that implies greater skewness for longer return measurement horizons. However, their findings do not cleanly map into predictions of magnitudes for our sample, as they consider returns for individual assets compounded over an interval of T periods, while we study a pooled sample where the number of periods over which returns are compounded varies from twelve (the minimum for sample inclusion) to three hundred and sixty.

We therefore conduct a set of simulations to develop predictions as to the effects of return measurement horizon on fund performance outcomes. Our simulations are conducted under a set of simplifying assumptions, including (i) the number of months for which return data is available for a given fund is random, and independent of either fund performance or economic conditions, a single-factor market model applies, returns are normally distributed each period and independent over time, and that funds' true alphas, betas, and residual return volatility vary cross-sectionally, but are independent of each other and are time-invariant. Of course, these simplifying assumptions are inaccurate to varying degrees in the actual data. As a consequence, the simulations provide predictions as to the direct effects of moving from short horizons to compound long-horizon returns in a sanitized setting. Comparing performance measures obtained from the actual data to those obtained from the simulations therefore provides indication of the importance of various real-world complexities, including the endogeneity of fund launches and closures, non-normality and serial dependence in returns, potential correlations between alphas, betas, and volatility, etc., for long horizon performance measurement.

To facilitate these comparisons, we conduct the simulations using parameters that are calibrated to match the distribution of the monthly sample. We set the mean monthly fund beta to 1.024, the mean monthly (post-fee) alpha to -0.131%, the mean monthly management fee to 0.095%, and mean fund residual return volatility to 2.4%.[8] Although our empirical sample includes 360 months spanning January 1991 to December 2020, on average a fund appears in the data for just 132 months, with a standard deviation of 97 months. The effects of compounding depend in part on the number of months over which returns are compounded. To calibrate the simulations in this dimension we define a "failure" function whereby a fund fails in month *t*, and all subsequent returns for the fund are excluded from the simulation,

[8] Betas, alphas, and monthly fees are assigned to individual funds based on the specified means, and random draws from normal distributions with standard deviation of 0.2, 0.15%, and 0.045%, respectively. The standard deviation of residual returns is assigned to specific funds based on random draws from a uniform (to ensure only positive outcomes) distribution, such that outcomes range from .014 to .034.

if the month t outcome on a random variable distributed uniformly over the interval 0 to 1 exceeds $.035/\ln(1+t)$. This simple function delivers a distribution of fund lives that matches the actual data quite closely.

The distribution of fund returns depends in part on the distribution of market returns, through the effect of beta. Our goal is to make predictions regarding the properties of individual fund performance measures that can be assessed with measures constructed from sample data. The sample contains only a single times series history of market returns. Indeed, compound market outcomes that are matched to fund returns vary across funds only because the set of months for which return data is available varies across funds. We therefore implement the simulations using the actual history of the SPY ETF (our market proxy) returns over the 360 sample months spanning January 1991 to December 2020. The failure function described in the prior paragraph focuses on the number of months, t, since the beginning of the simulation, but is intended to capture in reduced form differential fund lives that result from the introduction and closure of funds at various points during the 360-month sample. As such, we do not wish to link the first month that a fund exists to the SPY return in the first sample month or any other fixed calendar month. To accomplish this, we randomize the order of the actual SPY returns in each round of the simulation. While the compound return to the SPY over the 360 months is not altered by such randomization, there is no remaining intertemporal linkage between calendar months for the SPY returns and t, the number of months since the fund first appears in the simulations.

While our empirical analysis necessarily focuses on the actual historical sample and the simulations described here are designed to make predictions regarding sample outcomes, it should be recognized that ex ante distributions of long-horizon fund returns are likely to display more variability and skewness than ex post distributions. While the sample data contains a single compound 30-year outcome for market proxy returns, other outcomes are possible ex ante. Fama and French (2018) rely on bootstrap simulations conducted with replacement and estimate the skewness of the value-weighted market return at the 30-year horizon to be 6.11. We implement bootstrap simulations using actual SPY returns from 1991 to 2020, also with replacement, and estimate that the skewness coefficient for 30-year

SPY returns is 3.30. The ex-ante skewness of market proxy returns implies that the ex-ante skewness of fund returns is considerably larger than the ex-post skewness of full-sample fund returns observed in the actual data. We implement simulations that parallel those described below, but that rely on a different bootstrapped sample of SPY returns on each iteration, and indeed observe much larger skewness coefficients for the distribution of fund returns. However, the percentage of fund returns that exceed the SPY over the 30-year horizon, which is our main focus, are virtually identical to those described below.

Within each round of the simulation we generate returns to 500 funds for t = 1 to 360 months. The simulation is repeated 10,000 times, resulting in a pooled distribution of 1.8 billion monthly fund returns and matched SPY returns that can be compounded across months to obtain longer horizon outcomes. However, after implementing the failure function described above and also (consistent with our sample methods) omitting funds that do not have at least 12 monthly observations, the number of simulated monthly returns included in our calculations is reduced to 478 million. We study compound returns to each fund, both as reported and after fees are added to obtain pre-fee returns, and compound returns on the market proxy over the matched months. Panels B to E of Table 2 report simulated means, medians, standard deviations, and skewness statistics for these variables at the indicated horizons.

The notable results that can be observed in Table 2, and that comprise predictions as to the effects of compounding across multiple periods to be observed in the actual sample, include:

1. The skewness of fund returns increases with the return measurement horizon. Specific return skewness parameters from the simulations are -0.28 at the monthly horizon, 0.56 at the annual horizon, 3.04 at the decade horizon, and 8.93 at the lifetime horizon.[9]
2. Reflecting this skewness, the median compound fund return is less than the mean compound fund returns at all horizons except monthly.
3. The percentage of funds that outperform the SPY is always less than 0.5, and decreases as the return measurement horizon is lengthened. Specific simulation outcomes are that 48.5% of funds

[9] The negative skewness in monthly fund returns is attributable to the well-documented negative skewness in short-horizon monthly market proxy returns.

outperform that market at the monthly horizon, 43.3% at the annual horizon, 36.2% at the decade horizon, and 35.3% at the lifetime horizon.

4. The percentage of funds that outperform the SPY when returns are measured on a pre-fee basis is also less than 0.5 at all horizons, and also decreases with return measurement horizon. Specific simulation outcomes are that 49.7% of fund pre-fee returns exceed the market at the monthly horizon, 48.3% at the annual horizon, 46.3% at the decade horizon, and 46.1% at the lifetime horizon. This predicted outcome, if verified in the sample data, implies that declining outperformance rates as the return measurement horizon is increased are not simply attributable to the accumulated effects of fund fees.
5. Wealth ratios are positively skewed at horizons longer than one month, and on average are less than one and decrease with return measurement horizon.

Implications (1) to (5) all follow from the fact that the compounding of random short-horizon returns induces positive skewness in long-horizon returns (even if short-horizon returns are symmetric), a result first formally demonstrated by Arditti and Levy (1975). Intuitively, this positive skewness arises because (i) reversals of similar percentage magnitudes lead to compound losses (e.g. returns of 5% and -5% in either order compound to -0.25%), while continuations of similar percentage magnitudes lead to larger gains than losses (e.g. continuations of 5% lead to accumulated returns of 10.25%, while continuations of -5% lead to accumulated returns of -9.75%). The most closely related prior research, including Bessembinder (2018) and Farago and Hjalmarsson (2021), has focused on individual stocks. However, unless the simplifying assumptions employed here are in some manner offset by fund manager actions or other complexities (e.g. serial dependence in returns) the same implications apply to compound mutual fund returns.

While comparisons of fund returns to market proxy returns are of inherent interest, outcomes of this comparison for any particular fund depend in part on specific market outcomes and also on the fund's sensitivity, or beta, with respect to the market. We therefore also conduct tests focused on funds' "non-market" returns, designed to be uncorrelated with either betas or market outcomes. To do so, we estimate monthly market model regressions for each fund within each simulation. Non-market returns are defined for each fund and month as the simulated fund return less the product of the estimated fund beta and the matched SPY return. (Equivalently, the non-market return is the estimated fund alpha plus the monthly residual from the market model regression). These non-market returns are then compounded across months. Outcomes reported on Table 2 support the following prediction:

6. Non-market returns are negative on average, are positive for less than half of funds, and are positively skewed for horizons longer than one month.

Finally, we report in panels F and G of Table 2 simulation outcomes at the "lifetime" horizon for subsamples that are delineated based on whether the estimate of alpha obtained from the monthly-horizon market model regression is positive or negative. While some outcomes of this comparison, e.g. that mean compound fund returns and wealth ratios are higher for funds with positive rather than negative monthly alpha estates, are rather self-evident, additional and more subtle implications emerge.

7. The distribution of compound non-market returns for funds with positive alpha estimates is strongly positively skewed, while that for funds with negative monthly alpha estimates is modestly negatively skewed.
8. Some funds (9.6% in the simulations) with positive alpha estimates obtained from monthly returns nevertheless have negative lifetime non-market returns.

Implication (7) reflects a degree of truncation implicit in dividing the sample based on estimated monthly alphas. Funds with negative monthly alpha estimates cannot have extreme right tail compound return outcomes, and funds with positive monthly alphas rarely have extreme left tail compound return outcomes.

Implication (8) highlights a fundamental shortcoming of alpha estimated from short-horizon returns as a performance measure for a long-horizon investor. Alpha is the arithmetic mean of non-market returns. Compound non-market outcomes are determined directly by the geometric mean of the short-horizon non-market returns, and it is well known that the arithmetic mean exceeds the geometric mean in every sample with positive volatility (and more so if returns are more volatile). Compound outcomes can therefore be negative even while arithmetic means (alphas) of short-horizon returns are positive. While the fact that arithmetic means can mislead regarding compound return outcomes is well known, the resulting shortcomings of alpha as a performance measurement for long term investors have not been emphasized in the literature.

3. Mutual Fund Performance at the Annual, Decade, and Full-Sample Horizons

We next report on equity mutual fund performance at three horizons: annual, decade, and "lifetime." The last designation refers to all months that the fund is contained in the sample, and does not literally equate to the lifetime of the fund in those cases where the fund was present prior to the 1991 sample start date or for those funds that continue after the 2020 sample end date. In those cases where a given mutual fund has data for only a portion of a given period the computation pertains only to the months with data, as the alternative of computing returns only for funds with data for the entire sample period would introduce survivorship biases. Tables 3, 4, and 5 present results at the annual, decade, and lifetime horizons, respectively.

3.1 Mutual Fund Performance at the Annual Horizon

Panel A of Table 3 shows that the mutual funds have return data for an average of 11.3 months in a given year, and that the pooled (across funds and years) mean annual return is 9.47%. By comparison, the mean (also pooled across funds and years) SPY return over the same months is 10.12% and the mean value-weighted market return is 10.69%. The mean annual fund wealth ratio is 0.990 when benchmarked against the value-weighted market and 0.994 when computed relative to the SPY. Each average wealth ratio differs significantly from a benchmark of 1.0, with a p-value less than 1%.[10] Annual fund returns are

[10] We compute p-values for wealth ratios using the bootstrapped skewness-adjusted t-statistics proposed (equation 5) by Lyon, Barber, and Tsai (1999), based on 1,000 bootstrap iterations.

moderately positively skewed; the estimated skewness coefficient is 0.566. The median fund return is 10.29%, as compared to a mean return of 9.47%.[11] Panel B of Figure 3 displays the frequency distribution of annual mutual fund returns. While the figure displays some irregularities, skewness is not strongly evident.

Only a minority of funds outperform market benchmarks at the annual horizon. In particular, 39.3% outperform the value-weighted market portfolio and 41.1% outperform the SPY ETF. By comparison, 68.8% of funds outperform one-month US Treasury Bills in a given year. Each rate of outperformance differs significantly from a benchmark of 50%, with p-value less than 1%.

Figure 2 displays the percentage of funds that outperform the market benchmarks and Treasury-bill benchmarks by calendar year. The majority of funds underperform the market benchmarks in most, but not all, years. In particular, more than half of funds outperformed market benchmarks in 2001 and 2009. The percentage of funds that outperform Treasury bills vary dramatically across years, from 2.5% in 2002 to over 90% in 1991, 1995-1997, 2003-2004, and eight of the eleven years from 2009 to 2020, depending on broad stock market performance during the year.

3.2 Mutual Fund Performance at the Decade Horizon

Table 4 reports results based on decade buy-and-hold returns. On average, return data is available for 71 months per decade, and the mean compound return pooled across funds and decades is 86.90%. By comparison, the mean SPY return over matched months of the same decade is 100.14% and the mean value-weighted market return over the same months of the same decade is 104.57%. Decade returns to mutual funds are more highly skewed than annual returns; the estimated skewness coefficient for decade returns is 2.64 (compared to 0.57 for annual returns), and the median decade-horizon fund

[11] Intuition may suggest that a positive skewness coefficient necessarily implies that the mean outcome exceeds the median. However, exceptions can occur, as here. See, http://ww2.amstat.org/publications/jse/v13n2/vonhippel.html. In this case the exception is attributable in part to unusually large kurtosis of 16.80.

return is just 39.38%, well below the mean of 86.90%. Positive skewness is readily apparent in the frequency distribution of decade-horizon mutual fund returns, displayed on Panel C of Figure 3.

Equity mutual funds outperform market benchmarks less often at the decade horizon as compared to the annual horizon. In particular, 34.1% of funds outperform the value-weighted market and 38.3% outperform the SPY ETF at the decade horizon, as compared to 39.3% and 41.1%, respectively, at the annual horizon. The mean wealth ratio for equity mutual funds at the decade horizon is 0.953 relative to the value-weighted market and 0.983 relative to the SPY. The fact that the mean wealth ratios for mutual funds are less than one can be attributed to the effect of fees and trading costs, while the fact that less than 50% of funds outperform market benchmarks reflects both the effects of accumulated fees and skewness in the return distribution. We present additional results to identify the distinct effects of skewness vs. accumulated fees in Section 4.

3.3 Mutual Fund Performance at Long Return Measurement Horizons

Panel A of Table 5 reports corresponding results for the lifetime (within the 1991 to 2020 database) horizon. Only 347 of the 7,883 sample funds have return data for all months from 1991 to 2020. On average across funds, return data is available for 133 months (median of 112 months). The mean lifetime buy-and-hold return for domestic equity funds is 294.35%. By comparison, the mean matched-period buy-and-hold return to the SPY is 297.69% and the mean matched-period return to the value-weighted stock market is 332.88%.[12] The mean wealth ratio for domestic equity funds over their lifetimes is 0.884 relative to the value-weighted stock market and 0.935 relative to the SPY ETF.

The distribution of lifetime equity mutual fund returns is strongly positively skewed. The estimated skewness coefficient is 6.40, and skewness is strongly observable in the frequency distribution displayed on Panel D of Figure 3. The median lifetime return to domestic equity mutual funds is 95.09%,

[12] We observe only one sample path for each SPY and VW market returns over the 360 sample months. The cumulative SPY return over the 360 months was 1935%, while the cumulative VW market return was 2245%. The means reported on Table 5 are lower because they reflect the average across funds of the market proxy returns during the months the fund was present in the data. All variation in matched SPY and VW market returns across funds is attributable to the fact that funds are present in the data for different subsets of the 360 months.

compared to a mean return of 294.35%. Only 24.1% of domestic equity funds outperform the value-weighted market return over the full sample, while 30.3% of outperform the SPY ETF, a benchmark that could have been captured by any investor who simply held the SPY and reinvested dividends. While the majority of domestic equity funds outperform T-bills over the full sample, 20.8% do not.

We also report on Panel A of Table 5 the 95th percentile outcome on each variable. While the cross-fund median buy-and-hold return is 95%, the 95th percentile of the distribution is 1,430%. The 95th percentile wealth ratio measured against the SPY is 1.598, which reflects a cumulative gross return fund return nearly 1.6 times as great as the cumulative gross return to the SPY over the fund's life. That is, while the presence of positive skewness in the distribution of compound fund returns implies that many funds underperform, it also implies that some funds exhibit very strong performance.

In Panel B of Table 5 we report lifetime results for four groups of funds delineated based on the length of time that the database contains return data for the fund. The four groups are funds with lives of 1 to 5 years (2,336 funds), 5 to 10 years (1,814 funds), 10 to 15 years (1,675 funds), and 15 to 30 years (2,058 funds). Not surprisingly, since poorly performing funds are likely to be shut down sooner, funds with shorter lives have worse average performance. For example, only 24.6% of funds with lives less than five years have lifetime buy-and-hold returns that outperform SPY, and only a slight majority (58.4%) of these funds outperform one-month Treasury bills. In contrast, almost all (97.7%) of funds with lives that exceed fifteen years outperform Treasury bills. The most informative fact observable in Panel B of Table 5 is that, even in the longest-lived group (a characteristic than can only be observed ex post), less than half (43.3%) of funds outperform the SPY, and less than a third (32.6%) outperform the value-weighted market.

Fama and French (2010) document that equity mutual funds as a class underperform market benchmarks when returns are measured at the monthly horizon. The results here show that the rate of underperformance increases at longer return horizons, a result that can be attributed at least in part to the fact that the compounding of returns over time leads to positive skewness in the distribution of long-horizon returns. However, the rate of underperformance is lower for funds with longer lives, which can be attributed to the fact that the poorest performing funds tend to be shut down before their lives are long.

As noted, the majority of the literature assessing fund performance studies returns measured at the monthly horizon, and summarizes performance across multiple months based on arithmetic means. Arithmetic means of monthly returns are then frequently compared across subsamples. Sharpe ratios rely on arithmetic mean returns in the numerator. Alphas are conditional (on zero factor outcomes) arithmetic mean returns. More broadly, fitted values from OLS regressions implemented in monthly returns, including Fama-MacBeth regressions and factor model regressions, deliver estimates of conditional (on explanatory variable outcomes) arithmetic mean returns.

We highlight the extent to which arithmetic means of monthly returns can be misleading for long term investors with results reported in Panels C of Table 5. For each fund, we compute both the arithmetic and geometric mean monthly return over its full life, as well as the arithmetic and geometric mean monthly returns to the SPY over the matched set of months. The first row of Panel C reports the cross-fund mean of the difference between the fund and the SPY arithmetic mean returns, and the percent of these observations that are positive, while the second row of each Panel reports the cross-fund mean of the difference between fund and benchmark geometric mean returns, and the percent of these observations that are positive.

Focusing on the first row of Panel C, the mean cross-fund difference between arithmetic mean fund returns and arithmetic mean SPY returns over matched months, is -0.17%. The arithmetic mean monthly fund return exceeds that of the SPY over matched months for 35.1% of funds. By comparison, the geometric mean fund return exceeds the geometric mean SPY return over matched months for 30.3% of sample funds (second row of Table 5, Panel C, and equivalent by construction to the proportion with greater buy-and-hold returns, as reported on Panel A). The next two rows of Table 5, Panel C report corresponding outcomes for the subsample of 2,769 funds with positive SPY-adjusted arithmetic mean returns. Among these, 15.2% underperform (84.8% outperform) the SPY in terms of compound (or geometric mean) monthly returns, despite the higher market-adjusted arithmetic average return. The corresponding figure when the value-weighted market is used as the benchmark instead (not tabulated) is that 17.8% of funds underperform the market in terms of compound returns even though the arithmetic mean of market-adjusted returns for the fund is positive.

Panel D of Table 5 reports outcomes that correspond to Panel C, except that the benchmark is the one-month Treasury bill rate rather than the SPY. The most striking result is that, among the subsample of 6,639 funds (84.2% of total) that have a positive sample risk premium, measured conventionally by an arithmetic mean monthly return that exceeds the mean monthly Treasury bill return, 6.0% actually underperform Treasury bills in terms of compound returns over their lives.

The term "volatility drag" has been coined by practitioners to describe the fact that compound returns are lower than might be anticipated based on the arithmetic mean of short-horizons returns. While it is broadly known that the geometric mean return in any sample with positive volatility is less than the arithmetic mean, to our knowledge the degree to which monthly mean returns are potentially misleading regarding long-horizon investor outcomes has not been widely discussed in the broad academic literature that reports arithmetic means of short-horizon returns.

3.4 Comparisons of Sample Outcomes to Simulation Outcomes and Predictions

The empirical results for mutual fund returns at the annual, decade, and "lifetime" horizons strongly support the implications obtained based on the simulations. Compound fund returns are positively skewed at all horizons other than monthly, and skewness increases with return measurement horizon. Median fund returns are less than mean fund returns at all horizons longer than monthly, and more notably so at longer horizons. The percentage of funds that deliver compound returns that exceed compound SPY returns is always less than half, and decreases with return measurement horizon.

Arditti and Levy (1975), Bessembinder (2018), and Farago and Hjalmarsson (2019) all observe that the compounding of random short-horizon returns introduces substantial positive skewness into long-horizon returns (even if short-horizon returns are symmetric). However, their results focus on the compounding of returns to individual securities. The results here verify that compounding generates substantial skewness in longer horizon returns for the periodically rebalanced (based on managers' disparate objective functions) portfolios that underlie mutual fund returns as well.

The specific empirical outcomes for compound fund returns as reported on Tables 3 to 5 are not only directionally consistent with the simulation-based implications, but sample parameter estimates are broadly similar to simulation parameters. For example, the skewness coefficients for fund returns in the actual data at the monthly, annual, decade, and lifetime horizons are -0.42, 0.57, 2.64, and 6.40, compared to -0.28, 0.56, 3.04, and 8.93 in the simulations. The percentages of fund returns that exceed matched-month SPY returns are 47.2%, 41.1%, 38.3%, and 30.3% at the monthly, annual, decade, and lifetime horizons in the sample, compared to 48.5%, 43.3%, 36.2%, and 35.3% in the simulations.

The simulation outcomes are based on numerous oversimplifications, including that fund life is independent of market conditions or past performance, that single period returns are distributed normally and are independent over time, that firm parameters such as betas and alphas are distributed randomly, etc. As such, the simulation outcomes illustrate the pure effects of time horizon and the compounding of random returns. The broad similarities of the sample-based outcomes to the simulation-based outcomes suggest that endogenous decisions to open and close funds, non-normality of monthly returns, and various forms of serial dependence are not the primary reasons that compound fund returns contain different information regarding fund performance as compared to measures constructed from short-horizon returns.

Still, it will be of interest for future research to identify the specific reasons that the empirical outcomes do not directly match the simulation outcomes that illustrate the pure effects of the compounding of random returns while ignoring several real-world complexities. The fact that the empirical full sample outcomes indicate a smaller skewness coefficient and a higher rate of underperformance relative to the SPY as compared to simulation outcomes is particularly intriguing.

3.5 The Role of Fund Size

The results presented in the preceding sections pertain to the full sample of 7,883 funds. However, as the data on Table 1 show, the distribution of fund sizes is highly skewed, with a mean fund TNA of $1.18 billion, compared to a median of $149 million. One possibility is that the high rates of underperformance documented here are primarily attributable to smaller funds. We next assess relations between fund size and long-run investment performance. However, numerous studies (see, for example, Cremers, Fulkerson, and Riley, 2019) have documented that investment flows respond to prior fund performance, implying that returns may cause fund size changes; i.e. fund size is endogenous. Further, beginning-of-sample fund size is not very informative about size over the course of a potentially long fund life.

To assess the role of fund size while avoiding biases attributable to endogenous fund flows, we adopt the following procedure. We first compile the cross-sectional distribution of fund sizes for each sample month, and record the percentile position of each fund within the size distribution for each month. We then determine if the size of a given fund exceeds the 25th, 50th, or 75th percentile of the size distribution in any month of the sample. Having done so, we construct three subsamples that exclude smaller funds. The first excludes any fund whose size never reaches the 25th percentile, the second excludes any fund whose size never reaches the 50th percentile, and the third excludes any fund whose size never reaches the 75th percentile. Within these subsamples, we only study returns after the first month that a fund is larger than the threshold size. The exclusion of returns prior to this date avoids lookahead bias stemming from endogenous fund flows.

Table 6 reports on monthly, annual, decadal, and full sample returns for funds in each of these size groups. For convenience Table 6 also reproduces corresponding outcomes for the full sample. This data shows that the distribution of compound returns is not as highly skewed for subsamples that contain only larger funds. In particular, the skewness of "lifetime" returns is 4.70 for the largest fund subsample, compared to 6.40 for the full sample. However, the finding that most funds underperform the SPY is quite uninform across size-based samples. At the annual horizon, for example, 41.1% of the full sample of funds outperform the SPY ETF, compared to 41.8% of funds in the largest subsample. At the "lifetime" horizon, 30.3% of funds in the full sample outperform the SPY, while slightly fewer, 29.6%, of funds in the subsample of largest funds outperform. On balance, these results demonstrate that the key conclusions supported by the results reported on Tables 3 to 5 are not attributable to the presence of small funds in the sample, but hold even in the subsample of the largest funds.

3.6 A Comparison to SPIVA

As noted, academic research focused on mutual fund performance has mainly reported conditional (e.g. alphas) and unconditional means of returns at short (most often monthly) horizons. In contrast, S&P Dow Jones Indices releases annual "Standard & Poor's Indices Versus Active Funds (SPIVA)" scorecards that compare compound fund returns to compound S&P 1500 index outcomes for horizons of up to twenty years. The 2020 SPIVA report states that only 14% of domestic equity funds outperformed the S&P Index over the 20-year period 2001 to 2020.[13] By comparison, we report (Table 5) that a proportion more than twice as large, 30.3%, of domestic equity funds outperform the SPY ETF over the 30-year period ending at the same date.

We investigate the extent to which this notable difference in outperformance rates is attributable to differing procedures for including funds in the sample, benchmark selection, sample period, or to other methodological choices. The Glossary of the 2020 SPIVA report indicates that "we use the opportunity set available at the beginning of the period as the denominator" to assess the percentage of funds underperforming the index. We identify 2,849 non-index funds that were present in our sample as of the end of 2000. We report in Table 7 the cross-fund mean compound return for these funds over the 2001-2020 period, as well as matched-month compound returns to the SPY ETF, the S&P 500 index, and the S&P 1500 index (each index return is inclusive of reinvested dividends). The results indicate that the S&P 500 index comprises a slightly higher hurdle (mean compound return over matching months equal to 127.64%) as compared to the post-fee SPY ETF (mean compound return over matching months equal to 124.66%), and that the S&P 1500 comprised a moderately higher hurdle yet (mean compound return over matching months equal to 140.33%).[14] However, the results reported on Table 7 show that 33.5% of

[13] The report is available at https://www.spglobal.com/spdji/en/documents/spiva/spiva-us-year-end-2020.pdf, and the comparison to the S&P 1500 Index is contained in the first row of "Report 1" on page 9.
[14] As noted, though, index returns cannot be directly captured, as trading would be required.

funds in this sample outperformed the S&P 1500 index in terms of compound returns over matched months during the years 2001-2020. That is, within the same twenty years, and focusing on the same benchmark and a comparable sample, we document a markedly higher fund outperformance rate as compared to that reported by SPIVA.

Delving further, the Glossary of the SPIVA report, under the heading "Percentage of Funds Underperforming the Index" states "We determine the count of funds that have survived and beat the index." As reported in Table 7, we find that 15.7% of the 2,849 funds present in our sample at the end of 2000 *both* outperformed the S&P 1500 index and survived to the end of 2020. This figure is comparable to the 14.0% reported by SPIVA, with remaining differences likely attributable to slightly divergent sample selection methods. However, a larger proportion, 17.8%, of funds outperformed the index during the months the funds were present, but exited the sample prior to 2020. The departure of over 500 outperforming funds from the sample may be related to the fact that overall stock market performance was weak from 2001 until early 2009; in particular the compound SPY return from December 2000 to February 2009 was -35.3%. In any case, these results indicate that rates of outperformance vs. the index as are affected quite notably by a methodological decision that counts outperforming but non-surviving funds as underperforming.

3.7 Potential Reverse Survivorship Bias

While the comparisons of sample outcomes to bootstrap-simulation-based outcomes are suggestive that endogenous fund lives are not the primary reason for different fund performance outcomes at longer versus shorter horizons, specific point estimates will be affected by endogenous lives. The positive skewness observed in the pooled distributions of compound fund returns reflects not only the effects of compounding, but also cross-sectional variation in the number of months that funds are present in the sample. Further, as Linnainmaa (2013) emphasizes and our results reported in Panel B of Table 5 verify for the present sample, funds with poor performance on average have shorter lives and, as a result, cross-sectional averages of fund-specific outcomes are potentially misleading regarding average mutual fund performance.

To provide evidence regarding the effect of return measurement horizon that is unaffected by endogenous fund lives, we measure returns at the monthly, annual, decade, and full sample horizon for portfolios of mutual funds. As Linnainmaa (2013) notes, returns to portfolios of mutual funds that are computed for the full sample period are not afflicted by the reverse survivorship issues he highlights. Specifically, we form a portfolio of ten randomly selected funds during each of the 360 months from January 1991 to December 2020, and compute buy-and-hold portfolio returns for each of the 30 years, each of the three decades, and for the full sample period. We repeat this process 10,000 times, to obtain bootstrap distributions of 3.6 million monthly portfolio returns, 300,000 annual portfolio returns, 30,000 decadal portfolio returns, and 10,000 "lifetime" portfolio returns.

Results are reported on Table 8. The skewness of equal-weighted bootstrap portfolio returns increases from -0.68 at the monthly horizon to 0.58 at the 30-year (full-sample) horizon, while the skewness of value-weighted bootstrap portfolio returns increases from -0.62 at the monthly horizon to 0.90 at the full sample horizon.[15] The percentage of equal-weighted mutual fund portfolios that outperform the SPY during the matched months decreases from 47.5% at the monthly horizon to 5.5% at the 30-year (full-sample) horizon, while the percentage of value-weighted mutual fund portfolios that outperform the SPY decreases from 48.1% at the monthly horizon to 16.6% at the full-sample horizon. The fact that the percentage of bootstrapped fund returns that exceed the SPY over their full lifetimes is substantially lower than corresponding percentage for the actual sample reflects in part that the average sample fund life is just eleven years, while the bootstrap simulation portfolios always involve a full thirty years of compounding. These results indicate that our conclusion that the choice of return measurement horizon significantly affects inference as to whether mutual funds outperform market benchmarks is robust to complications arising from the fact that the life of individual funds is endogenously related to fund performance.

---

[15] The fact that the skewness of portfolio returns is less than that of the portfolio's component returns has been previously noted, e.g., by Albuquerque (2012). The negative skewness of shorter-horizon mutual fund portfolio returns in our sample mirrors negative skewness in the distribution of shorter-horizon market returns over the 1991-2020 sample period.

3.8 Fund Characteristics and Long Horizon Mutual Fund Performance

In Figure 4 we display the percentage of sample funds that outperform benchmarks over their lifetimes, when funds are sorted into deciles based on a variety of fund characteristics. Panels A, B, and C report results when the benchmark is the value-weighted market, the SPY ETF, and one-month Treasury bill, respectively. Decile portfolios are defined based on (i) the R-squared statistic in a regression of monthly fund returns on monthly SPY returns (which measures the extent to which the fund simply tracks the overall market, (ii) the slope coefficient (beta estimate) from the same regression, (iii) lifetime average fund fees and expenses (based on the CRSP "expense ratio" variable) as a percentage of TNA, (iv) the volatility (in particular, the standard deviation) of monthly fund returns, (v) the skewness of monthly fund returns, and (vi) the lifetime average fund size, based on TNA.[16]

The most notable conclusion that can be drawn based on the results displayed in Figure 4 is that long term fund performance generally does bear any simple or linear relation to the characteristics. While funds in the first decile by these measures often underperform, underperformance rates are generally not monotone across deciles. Focusing on estimated fund beta, for example, the rate of outperformance against SPY is just 15.7% for the lowest beta decile (mean beta = 0.523), compared to 40.6 for decile eight (mean beta = 1.134), but only 22.1% for decile ten (mean beta = 1.576). The relation between fees and rates of outperformance is also not monotone. While the decile of funds with the highest fees and expenses (average expense ratio of 0.187% per month) displays a low rate of outperformance (20.5%) relative to SPY, the decile of funds with the lowest fees and expenses (average expense ratio of 0.018% per month) also has a relatively low (33.6%) rate of outperformance relative to SPY, as this category is dominated by index funds. The highest rate of outperformance relative to SPY (33.3%) is for funds in the sixth decile of fees and expenses (average 0.100% per month). The non-linear relations documented in Figure 4 caution against the presumption that long run fund performance can be well explained by simple univariate relations between fund characteristics and performance.

[16] We measure average fund size based on percentile ranks. For each month, we assign a percentile rank to all funds with return data. We then compute the time series average of percentile ranks for each fund.

4. Why do so many Equity Mutual Funds Underperform in the Long Run?

The literature that studies mutual fund returns has documented that mutual fund returns on average lag behind market benchmarks. This empirical regularity is often attributed to the fact that funds charge management fees, and, in addition, incur trading costs. Alternatively, behavioral biases could exacerbate the effect of fees. Further, some fund managers may be skilled, but not sufficiently so to fully overcome fees and other costs. In addition, fund's systematic risk exposures are relevant. If a given fund has a market beta that is less than one then it might be expected that the fund would be more likely underperform even in the absence of fees or expenses, for any period with a positive ex post market excess return. In any case, we document that a strikingly high percentage of US equity mutual funds underperform market benchmarks in the long run. We next assess the extent to which these high rates of underperformance can indeed be attributed to fees, managerial skill, or factor exposures.

4.1 Fees and Managerial Skill

We first assess whether the high rates of underperformance can be attributed to management fees, which average 0.095% per month, as documented on Panel A of Table 9. To do so, for each fund and month we add the fund's monthly percentage fee to the reported return to obtain the equivalent pre-fee return. The left panel of Table 9 Panel B reports results obtained when we compound pre-fee returns over the full sample.

Notably, the mean compound pre-fee buy-and-hold mutual fund return is 393.6%, which exceeds the mean compound return on both the SPY and the value-weighted market over the matched time periods, which are 297.7% and 332.9%, respectively. This long-return-horizon result complements the short-return-horizon results reported by Berk and van Binsbergen (2015) and Fama and French (2010), who report positive alpha estimates based on pre-fee monthly returns. The mean wealth ratio relative to the value-weighted market when pre-fee returns are compounded is 1.011. That is, domestic equity mutual funds on average outperform the value-weighted market in the long run, when pre-fee returns are considered. Outperformance in pre-fee mutual fund returns relative to the SPY is larger: the mean wealth ratio is 1.074.

However, median pre-fee fund returns are much lower than means, and only a minority of funds outperform the market benchmarks, even on a pre-fee basis. Specifically, pre-fee buy-and-hold returns exceed the buy-and-hold return to the value-weighted market for 37.6% of funds and exceed the (post-fee) SPY buy-and-hold return for 45.2% of funds. By comparison, the simulations (Table 2) predicted that 46.1% of funds would outperform the SPY on a pre-fee basis. On balance, these results imply that, while the accumulated effect of fees contributes, skewness in the distribution of long-horizon compound returns is a key reason that so many mutual funds underperform benchmarks at long horizons.

In addition to fees, mutual fund returns are affected by trading costs and managerial skill. We next focus on estimated fund alphas, reasoning that alphas reflect the combined effect of fees, trading costs, and management skill. In particular, we estimate alpha for each fund by a regression of excess fund returns on excess SPY returns, using the full time series of available post-fee monthly returns. The mean monthly alpha estimate in our 30-year sample of domestic equity funds (Panel A of Table 9) is -0.131% per month.[17] We then subtract the (negative on average) fund alpha estimate from each post-fee monthly return for the fund. The resulting return series for every fund are, by construction, characterized by zero monthly alphas within the sample. We then compound these “zero-alpha” monthly returns and compare outcomes to market benchmarks.

The central finding of this exercise is that the majority of funds underperform the value-weighted market, even when zero-alpha returns are compounded. In particular, only 36.2% of funds outperform the SPY when alpha estimates are deducted before returns are compounded. Management fees, other expenses, and trading costs no doubt contribute to the empirical fact that less than half of mutual funds outperform market benchmarks in the long run. This analysis shows that these explanations are not

[17] Berk and van Binsbergen (2015) argue that foreign funds, which are not included in our sample, perform better than domestic funds, and report alphas for their broader sample that do not differ significantly from zero when using returns to an array of Vanguard index funds as benchmarks.

complete, however. Rather, the positive skewness in the distribution of long horizon mutual fund returns is also important.

4.2 Allowing for Factor Exposures

We document in that the majority of U.S. equity mutual funds underperform both the value-weighted stock market and the SPY ETF in terms of compound returns over the full time period that the funds are included in our sample drawn from the CRSP mutual fund database. However, underperformance relative to these benchmarks could be attributable in part to factor exposures. Cash balances in particular could contribute: a hypothetical fund that earned returns identical to the overall market on the portion of its portfolio invested in risky assets but that also kept some funds in cash would have a market beta less than one and would also underperform the market during any time period where the market return exceeded the cash return.

To assess this possibility, we focus on fund returns after removing the effect of market exposures and market outcomes. We focus on a simple single-factor model, as our main emphasis is on the effects of the compounding of random returns, not on the effects of alternative benchmarks, which have been widely studied. In particular, we identify for each period t and for each fund the sum of the fund's alpha, estimated by a time series regression of excess fund returns on excess SPY returns, and the time t residual from the same regression. This measure parallels the risk-adjusted return estimated by Brennan, Chordia, and Subrahmanyam (1998) for individual stocks. In light of the fact that the measure adjusts for both estimated market exposures and ex-post market outcomes we adopt the label "non-market return." The measure can be interpreted as the excess fund return relative to the return that would have been predicted based on the fund's estimated market exposure and the observed market outcome for the period.

We then compound these non-market returns across all months that each fund is contained in the database. Panel C of Table 9 reports on the results. Consistent with the simulation outcomes reported in Section 2, the mean compound non-market return is modestly negative, equal to -5.6%. However, the distribution of compound non-market returns is also strongly positively skewed, with skewness coefficient equal to 7.5. Less than one-quarter (23.6%) of funds have positive compound non-market returns, a finding also consistent with the implications of the simulation. We conclude that market exposures have relatively little explanatory power for the observation that the majority of funds underperform market benchmarks over their full lives.

We also report on Table 9 Panel C outcomes for subsamples of funds based on the sign of the estimated monthly alpha. These results also confirm the implications of the simulation conducted in Section 2. The skewness of fund returns is strongly positive for funds with positive monthly alpha estimates, while skewness is negative for funds with negative alpha estimates. While compound non-market returns for those funds with negative monthly alpha estimates are negative by construction, the sign of compound non-market returns for funds with positive monthly alpha estimates need not be positive. Specifically, 12.2% of the funds with positive monthly alpha estimates nevertheless have negative compound non-market returns. This result reinforces the point that short return-horizon alpha, as a conditional arithmetic mean, can be misleading regarding compound longer-horizon performance.[18]

5. Mutual Fund Investment and Investor Wealth

We document that compound mutual fund returns most often underperform market benchmarks, and that compound non-market returns to mutual funds are negative on average. However, we also document that some funds deliver very large returns. Both fund sizes and fund lives differ in the cross section, and are related to fund performance. In addition, our outcomes to this point, as well as the broader mutual fund literature, focus on fund returns inclusive of dividends. As such, the results rely on an implicit assumption that dividends are reinvested in the fund. In practice investors may reinvest dividends fully in some funds or during some periods, but not others. Further, investors in aggregate may invest or withdraw additional funds (fund "flows"), the effects of which are not captured in either the

[18] To estimate long-return-horizon alphas directly requires accommodation of the fact (see for example, Levhari and Levy, 1977 and Longstaff, 1989) that betas also depend on return measurement horizon. Bessembinder, Cooper, and Zhang (2022) discuss some of the challenges that arise in estimating long-horizon alphas.

compound buy-and-hold returns studied herein to this point or in the arithmetic mean returns studied by the broader literature.

5.1 A Framework for Measuring Aggregate Mutual Fund Outcomes

We rely on the following framework to quantify mutual fund investors' aggregate experience, while considering fund size, fund life, periodic investor fund flows, as well as opportunity costs. Let W_0 denote investor's beginning-of-sample wealth, and assume that returns are studied over T periods. Investors each period allocate wealth between a mutual fund with return $R_t = R_{ct} + R_{dt}$, where R_{ct} is the capital gain component of the period t return, and R_{dt} is the dividend component, and an alternative investment that pays a return denoted R_{at}. The return on the alternative investment comprises the opportunity cost for funds invested in the mutual fund. Investors potentially makes a time t investment (or flow) into the fund in the amount F_t (with a withdrawal of funds or return of capital denoted by $F_t <$ 0). Let W_t, A_t, and M_t, denote time t outcomes for investors' wealth, the value of positions in the alternative asset, and the value of positions in the mutual fund, respectively, with $W_t = A_t + M_t$.

The value of the alternative asset position evolves according to $A_t = A_{t-1}(1+R_{at}) + M_{t-1}*R_{dt} - F_t$, as investors earn a return on the alternative asset, collect any mutual fund dividend, and potentially increase or decrease the mutual fund investment. The value of the investors' position in the mutual fund evolves according to $M_t = M_{t-1}*(1+R_{ct}) + F_t$, based on the capital gains return and any net new investment flow to the fund. Investors' wealth at time t can be expressed as $W_t = A_{t-1}(1+ R_{at}) + M_{t-1}*(1+R_t)$, and the increase in wealth since the prior period due to investing some funds in the mutual fund instead of the alternative asset is:

$$W_t - W_{t-1}*(1+ R_{at}) = M_{t-1}*(R_t - R_{at}). \quad (1)$$

Let $FV_{t,T} = (1+ R_{at+1})*(1+ R_{at+2})* (1+ R_{at+3})*\ldots *(1+ R_{aT})$ denote a future value factor obtained by compounding realized returns on the alternative asset from time t to time T. Applying (1) iteratively leads to the following expression:

$$W_T - W_0*FV_{0,T} =$$

$$M_0*(R_1–R_{a1})\, FV_{1,T} + M_1*(R_2–R_{a2})\, FV_{2,T} + \ldots + M_{T-2}*(R_{T-1}–R_{at-1})*FV_{T-1,T} + M_{T-1}*(R_T–R_{aT}). \quad (2)$$

The first line of expression (2) is investors' actual end-of-sample wealth, in excess of the wealth they would have attained if all capital had remained invested in the alternative asset. The second line of expression (2) shows that this dollar amount can be computed as the sum of the future values (using the alternative asset returns to compound forward) of the period-by-period wealth differentials specified by the right side of expression (1).[19]

An alternative method of quantifying investors' aggregate experience that also allows for periodic investor flows is the "dollar-weighted return," which is obtained as the internal rate of return of the time series of cash flows to investors in aggregate. Dollar-weighted returns for mutual funds are studied by Friesen and Sapp (2007), while Dichev and Yu (2011) consider dollar-weighted returns for hedge funds and Dichev (2007) computes dollar-weighted returns for several international equity indices. We show in Appendix B that the enhancement in end-of-period investor wealth due to investing in the mutual fund instead of the alternative asset is equivalently the end-of-sample value of the same series of cash flows that defines investors' dollar-weighted return.

Bessembinder (2018) implements expression (2) for individual common stocks, using the one-month U.S. Treasury Bill as the alternative asset. Here, we implement (2) using three alternative assets in turn: the one-month U.S. Treasury Bill, the SPY ETF, and the value-weighted U.S. stock market. We measure investors' aggregate mutual fund investment at each time t as the fund's Total Net Assets (TNA), as reported by CRSP.

We also implement a version of expression (2) where the simple differential $R_t - R_{at}$ is replaced for each period t by the "non-market" return for the month, obtained as the sum of the alpha estimated from a market model regression of excess fund returns on excess alternative asset returns and the time t residual from that regression, denoted $\alpha + \varepsilon_t$. In this implementation, the product $M_{t-1}(\alpha + \varepsilon_t)$ can be interpreted as the dollar excess return to the fund investor, after allowing for the exposure (beta) of the

[19] This measure is similar to the "Public Market Equivalent" introduced by Kaplan and Schoar (2005), in that each measure considers outcomes relative to an opportunity cost defined by compound returns to an alternative investment.

fund to alternative asset returns.[20] When implemented based on pre-fee fund returns, this time t quantity is conceptually similar to the "Realized Value Added (RVA)" defined by Berk and van Binsbergen (2015) in their expression (5). However, Berk and van Binsbergen estimate their RVA measure for each month t as $M_{t-1}^{*}\alpha$, where M_{t-1}^{*} is inflation-adjusted M_{t-1}, and study the mean across months of RVA for each fund. They do not, therefore, capture the effects of the compounding of random returns over multiple periods that are our focus.

5.2 Aggregate Outcomes to Mutual Fund Investors

Table 10 reports outcomes obtained when expression (2) is implemented for each of the 7,883 mutual funds in the sample, based both on actual net-of-fee returns and also on returns prior to fees. All outcomes are measured as of end-of-sample, December, 2020. The outcome for each fund pertains to aggregate investors in that fund, and the column of Table 10 labeled sum represents the aggregate outcome across all funds in the sample.

If the opportunity cost is defined by the one-month Treasury bill return, mutual fund investment increased investors' end-of-period wealth by \$8.66 trillion on a post-fee basis and by \$9.50 trillion on a pre-fee basis. These outcomes can be viewed as measures of the ex-post dollar risk premium earned by equity mutual fund investors for taking on the risk of fund investment and verify that equity mutual fund investment has allowed investors to participate substantially in the aggregate increase in wealth attributable to the public stock markets.

On the other hand, equity mutual fund investment reduced shareholder wealth, even on a pre-fee basis, if the opportunity cost is defined by the return to the value-weighted stock market. Specifically,

[20] For consistency, we also restate in this specification the compounding factors denoted $FV_{t,T}$ using alternative asset returns that are adjusted for the estimated beta risk of the fund. Specifically, $FV_{t,T} = [1 + R_{ft+1} + \beta^{*} (R_{at+1} - R_{ft+1})]^{*} [1 + R_{ft+2} + \beta^{*} (R_{at+2} - R_{ft+2})]^{*} [1 + R_{ft+3} + \beta^{*} (R_{at+3} - R_{ft+3})]^{*} \ldots {}^{*}[1 + R_{fT} + \beta^{*} (R_{aT} - R_{fT})]$, where β is the fund's beta estimated from the time-series regression. If the estimated beta is one the final outcome is identical to that given by expression (2).

actual post-fee returns led to a $2.89 trillion reduction in mutual fund investors' aggregate wealth, while hypothetical pre-fee returns were associated with a $0.47 trillion wealth reduction.

However, as noted, it is not possible to directly capture the return to the overall market, in part because tracking the market as a whole requires periodic primary market transactions. Investors could in principle have captured SPY returns with a simple buy-and-hold and reinvest dividends strategy.[21] When the SPY return is used in the role of the alternative asset, our computations reveal that mutual fund investors' aggregate wealth was decreased by $1.31 trillion during our thirty-year sample period. If we focus instead on hypothetical pre-fee fund returns, the outcome is that mutual fund investors' aggregate wealth was increased by $0.94 trillion during the sample period.

The difference in pre-fee vs. post-fee investor wealth enhancement (relative to the SPY return benchmark) of $2.24 trillion is the accumulated impact of mutual fund fees, inclusive of opportunity costs on the fees paid. This total is equivalent to 23.6% of $9.51 trillion end-of-sample market value of the mutual funds in our sample. The nominal total of fees paid during our thirty-year sample, computed based on the CRSP expense ratio and the TNA reported by CRSP for each fund and month, is $0.73 trillion. The remaining $1.51 trillion reflects the SPY-based opportunity cost, as funds used to pay fees did not earn SPY returns thereafter. By comparison, French (2008) estimates that all forms of active management, including expenses incurred by mutual fund, ETF, hedge fund, and closed-end fund investors, institutional manager costs, in combination with aggregate trading costs, led to total active management costs amounting to 0.67% of stock market capitalization per year.

We also report on Table 10 the percentage of funds with positive wealth enhancement outcomes, when returns are measured on both a pre- and post-fee basis, for each of the three alternative investment benchmarks. These percentages differ from those focusing on buy-and-hold returns reported on Table 5, because it is possible for a given fund's buy-and-hold return to exceed the buy-and-hold return to the

[21] It is, however, worth noting that the actual investors in the SPY ETF did not earn the buy-and-hold return to the SPY, as in aggregate they did not follow a buy-and-hold strategy. The annualized geometric mean return earned by SPY investors from 1993 to 2020 was 10.03%, while the annualized dollar-weighted return was 8.49%.

benchmark even while wealth is reduced rather than enhanced relative to benchmark outcomes. In particular, this can occur if larger fund returns tend to occur during periods when fund TNA is smaller, as might be anticipated if abnormally large returns that attract investor flows tend to be followed by normal or below-normal returns.

The results indicate that this scenario occurs more often than the reverse. In particular, while 24.1% of funds have buy-and-hold returns that exceed compound returns to the value-weighted market over matched intervals (Table 5), 20.0% of funds increased the wealth of their investors in aggregate as compared to the value-weighted market alternative. Similarly, while 30.3% of funds' buy-and-hold returns exceed the SPY compound return over matched intervals, 25.4% of funds increased the wealth of their aggregate investors when the SPY is the alternative asset. Finally, while 79.2% of funds outperform 1-month T-bills in terms of buy-and-hold returns, 74.4% of funds enhanced aggregate investor wealth relative to the one-month T-bill return benchmark. Stated alternatively, the final result implies that, despite stellar overall equity market returns during the 1991-2020 period, over one quarter of U.S. equity mutual funds decreased their aggregate investors' wealth as compared to the wealth they would have attained if they had instead earned one-month Treasury-bill returns. While management fees and trading costs contribute to this striking outcome, the timing of investor flows and the skewness of compound fund returns are also important.

The aggregate outcomes reported to this point are based on expression (2), which relies in each period on the simple difference between the fund return and the alternative asset return. However, this differential can include a component attributable to differences in systematic risk across the fund and the alternative asset. To allow for such, we also implement expression (2) when each simple differential $R_t - R_{at}$ is replaced by the "non-market" return for the month, obtained (with the implicit assumption that the alternative asset comprises an adequate proxy for the market), as the sum of the alpha estimated from a market model regression of fund returns on alternative asset returns and the time t residual from that regression, denoted $\alpha + \varepsilon_t$.

The last two rows of Table 10 report the resulting wealth enhancement outcomes when the alternative asset is the SPY ETF, the return on which could have been captured by investors. This specification indicates improved fund outcomes. In particular, end-of-sample wealth enhancement is improved by \$0.28 trillion (from -\$1.31 trillion to -\$1.02 trillion) when considering the effect of estimated beta as compared to expression (2) outcomes that implicitly rely on a beta equal to one.

The end-of-sample wealth improvement from mutual fund investment as compared to SPY returns on a pre-fee basis is \$1.23 trillion in this specification. In terms of the reasoning advanced by Berk and van Binsbergen (2015), this figure can be interpreted as a measure of the value of aggregate mutual fund manager skill during the 1991 to 2020 period, measured as of the end of the sample. As such, our long-horizon outcomes support the emerging consensus (surveyed by Cremers, Fulkerson, and Riley, 2019) that mutual fund managers possess skill. However, this figure, while notable, is less than the \$2.26 trillion in fees (also measured as of the end of sample) charged, and imply that investors in domestic equity funds during the 1991-2020 period suffer an aggregate wealth reduction of \$1.02 trillion, relative to the SPY benchmark, after allowing for beta exposures.

6. Conclusions

The literature that studies funds' return performance (including mutual funds, hedge funds, pension funds, etc.) is vast, but most of the evidence is based on conditional or unconditional arithmetic means of returns measured over short (e.g. monthly) horizons. However, we know of no compelling reason to believe that parameters estimated from short-horizon returns are the most relevant to diverse investors who may hold positions for long horizons. These investors will be concerned with the compound returns that accrue over horizons that are relevant to them. In this paper, we focus attention on the fact that simple measures of investment performance differ depending on the horizon over which returns are measured.

We study U.S. equity mutual funds for the 1991 to 2020 period, and show that the percentage of funds that outperform market benchmarks decreases with return horizon. In the monthly data, equity mutual fund returns exceed the matched-month return to the SPY ETF (taken as a proxy for the overall market that investors could readily have captured) for 47.2% of observations. The percentage of sample funds that generate buy-and-hold returns that exceed buy-and-hold returns to the SPY decreases to 41.1% at the annual horizon and 30.3% at the (fund-specific) full-sample horizon. Outperformance rates are slightly worse for the largest fund subsample, and the percentage of funds that improve shareholder wealth as compared to a buy-and-hold investment in the SPY is lower yet. Further, even some of those funds with positive alphas estimated from monthly returns deliver negative compound non-market returns at long horizons.

We also show that these horizon effects cannot simply be attributed to the cumulative effects of fees and other expenses, nor can they be attributed to market exposures. Rather, these outcomes reflect two important facts. First, the cross-sectional distribution of long-horizon fund buy-and-hold returns is strongly positively skewed, while such skewness is not observable in the pooled distribution of short-horizon (monthly) returns. Financial planning (e.g. at pension funds) is often based on assumed mean outcomes. Aside from the active debate as to whether assumed means are realistic, in a positively skewed distribution a potentially large majority of the possible future realizations are less than the mean outcome. Second, alpha is a (conditional) arithmetic mean, and arithmetic means are well-known to exceed the geometric means that determine compound returns.

Of course, while strong positive skewness implies that many funds underperform, some funds perform very well. Out of 7,883 sample funds, 442 delivered a positive full-sample compound return more than twice as large as the compound return to the SPY over the matched months, and 160 delivered compound returns three times as large as the SPY during the matched months of the full sample.

We also assess the aggregate effect of mutual fund investing on shareholder wealth in aggregate. When focusing on SPY ETF returns (which could have been captured by investors) to define the opportunity cost, and allowing for variation in estimated betas, we compute that pre-fee mutual fund returns were associated with $1.23 trillion in enhanced investor wealth, measured as of the end of our thirty-year sample. This figure can, in the spirit of Berk and van Binsbergen (2015) be viewed as a measure of the value of aggregate fund manager skill. However, fees reduced end-of-sample wealth by a larger amount, $2.26 trillion, so that the net wealth of mutual fund investors was reduced by $1.02 trillion in aggregate during the 30-year period we study.

A central conclusion supported by our analysis is that inference regarding fund performance differs when focusing on compound long-horizon returns rather than conditional or unconditional means of short horizon returns. While this conclusion is likely to be robust, point estimates and dollar amounts will be altered if the sample is broadened, e.g. to include international, target date, and levered funds, or is narrowed to focus on specific fund categories. Outcomes will also vary if multi-factor or style-specific benchmarks are employed, providing ample opportunity for additional research.

The results reported here imply that the evaluation of fund performance is intrinsically linked to return horizons: a given fund's performance relative to benchmarks can be positive over short horizons and negative over long horizons, even when results are measured from a single dataset. Our results imply that the assessment of fund performance cannot be evaluated independent of information regarding the return horizon that is most relevant to investors, and we provide estimates of U.S. mutual fund performance over various horizons, including monthly, annual, decadal, and full sample.

References

Albuquerque, Rui, 2012, Skewness in stock returns: reconciling the evidence on firm versus aggregate returns, *Review of Financial Studies*, 25, 1631–1673.

Ameriks, John, and Stephen Zeldes, 2004, "How Do Household Portfolio Shares Vary with Age?" Working Paper, Columbia University.

Arditti, Fred and Haim Levy, 1975, Portfolio efficiency in three moments: The Multiperiod case", *Journal of Finance*, 30, 797-809.

Bessembinder, Hendrik, 2018, Do stocks outperform treasury bills? *Journal of Financial Economics*, 129, 440-457.

Bessembinder, Hendrik, Michael Cooper, and Feng Zhang, 2022, What you see may not be what you get: Return Horizon and Investment Alpha, working paper. Downloadable at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3550284.

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

Berk, Jonathan, and Jules H. van Binsbergen, 2015, Measuring skill in the mutual fund industry, *Journal of Financial Economics*, 118, 1–20.

Boguth, Oliver, Murray Carlson, Adlai Fisher, and Mikhail Simutin, 2016, Horizon effects in average returns: The role of slow information diffusion, *Review of Financial Studies*, 29, 2241-81.

Brennan, Michael, Tarun Chordia, and Avanidhar Subrahmanyam, 1998, Alternative factor specifications, security characteristics, and the cross-section of expected stock returns, *Journal of Financial Economics*, 49, 345-373.

K.J. Martijn Cremers, Jon A. Fulkerson & Timothy B. Riley, 2019, Challenging the Conventional Wisdom on Active Management: A Review of the Past 20 Years of Academic Literature on Actively Managed Mutual Funds, Financial Analysts Journal, 75, 8-35.

Dichev, Ilya, 2007, What are stock investors' actual historical returns? Evidence from dollar-weighted returns, *American Economic Review* 97, 386–401.

Dichev, Ilya, and Gwen Yu, 2011, Higher risk, lower returns: What hedge fund investors really earn, *Journal of Financial Economics* 100, 248-263.

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

Fama, Eugene, 1972, Components of Investment Performance, *Journal of Finance*, 27, 551-567.

Fama, Eugene, and Kenneth French, 2018, Long-horizon returns, *Review of Asset Pricing Studies*, 8, 232-252.

Fama, Eugene, and Kenneth French. 2010. Luck versus skill in the cross-section of mutual fund performance. *Journal of Finance,* 65, 1915-1947.

French, Kenneth, 2008, Presidential address: The cost of active investing, Journal of Finance, 63, 1537-1573.

Farago, A., Hjalmarsson, E., 2019, Compound returns, Working paper, Department of Economics, University of Gothenburg. Download at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3398501.

Friesen, Geoffrey C., and Travis R.A. Sapp, 2007, Mutual fund flows and investor returns: An empirical examination of fund investor timing ability, *Journal of Banking & Finance*, 31, 2796-2816.

Gilbert, Thomas, Christopher Hrdlicka, Jonathan Kalodimos, and Stephan Siegel, 2014, Daily data is bad for beta: Opacity and Frequency-Dependent Betas, *Review of Asset Pricing Studies*, 4, 78-117.

Kamara, Avraham, Robert Korajczyk, Xiaoxia Lou, and Ronnie Sadka, 2016, Horizon pricing, *Journal of Financial and Quantitative Analysis,* 51, 1769-1793.

Kaplan, Steven, and Antoinette Schoar, 2005, Private Equity Performance: Returns, Persistence, and Capital Flows, Journal of Finance, 60, 1791-1823.

Kothari, S.P., Jay Shanken, and Richard Sloan, 1995, Another Look at the Cross-section of Expected Stock Returns, *Journal of Finance*, 50, 185-224.

Levhari, David, and Haim Levy, 1977, The capital asset pricing model and the investment horizon, *Review of Economics and Statistics*, 54, 92–104.

Linnainmaa, Juhani, 2013, Reverse Survivorship Bias, *Journal of Finance*, 68, 789-813.

Longstaff, Frances, 1989, Temporal aggregation and the continuous-time Capital Asset Pricing Model, *Journal of Finance*, 44, 871-887.

Loughran, Tim and Jay Ritter, 1995, The new issues puzzle, *Journal of Finance*, 50, 23–51.

Lyon, John D., Brad M. Barber, and Chih-Ling Tsai, 1999, Improved methods for tests of long-run abnormal stock returns, *Journal of Finance*, 54, 165–201.

Pástor, Lubos and Robert Stambaugh, 2012, On the Size of the Active Management Industry, *Journal of Political Economy*, 120, 740–781.

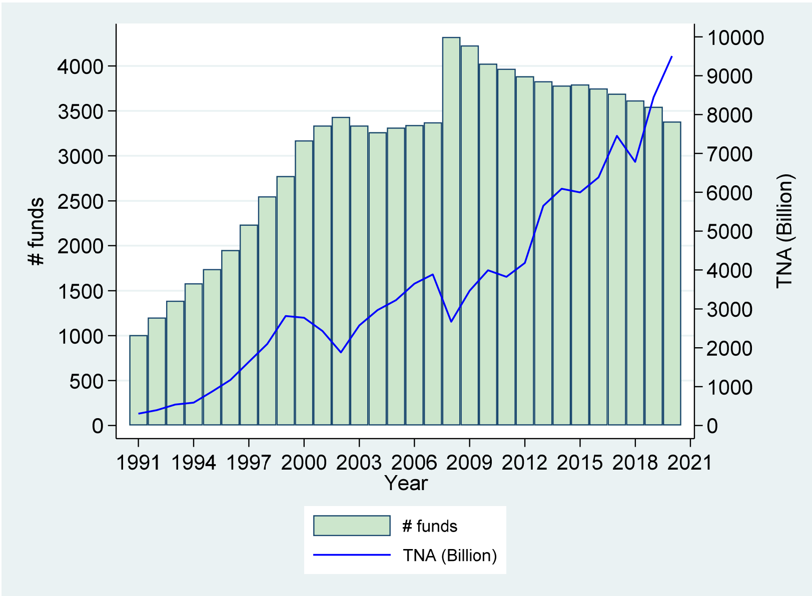

Figure 1
Number of funds and aggregate TNA, by year
This figure plots the annual number of active equity funds (left axis) and the aggregate TNA in $Billion (right axis) in each year.

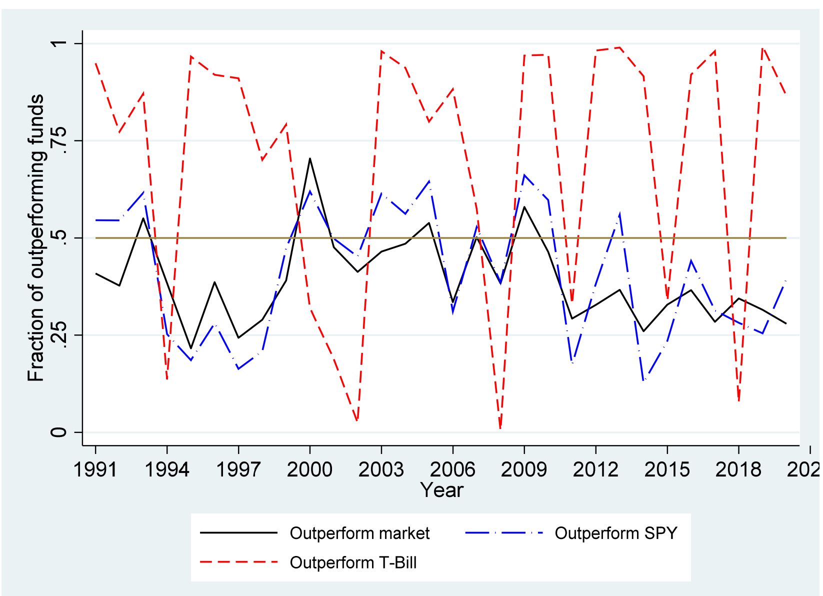

Figure 2
Fraction of funds that outperform three benchmarks, by year
This figure plots the fraction of funds that outperform each of three benchmarks in each year. The benchmarks are the CRSP value-weighted market return, the SPDR S&P 500 ETF return (SPY), and the one-month T-Bill rate.

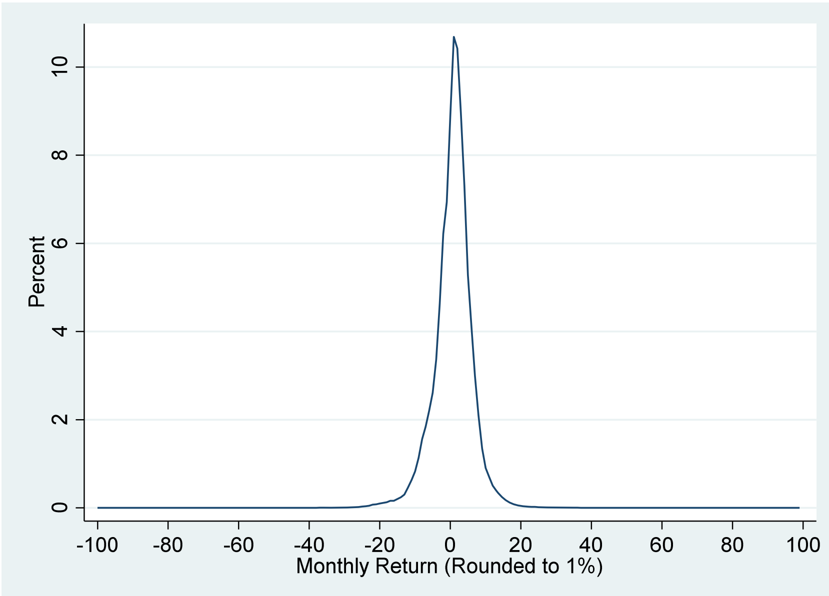

Figure 3, Panel A
Frequency distribution of monthly mutual fund returns
This figure plots the frequency distribution of monthly mutual fund returns from 1991 to 2020, rounded to the nearest 1%. Monthly returns above 100% are winsorized to 100%. See Table 1 for summary statistics of the monthly fund returns.

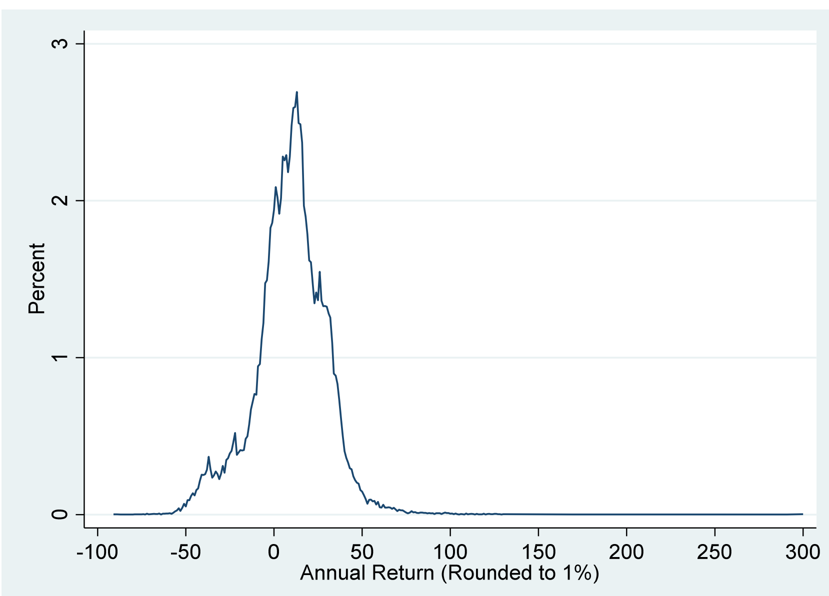

Figure 3, Panel B
Frequency distribution of annual mutual fund returns
This figure plots the frequency distribution of annual mutual fund returns in each year from 1991 to 2020, rounded to the nearest 1%. Annual returns above 300% are winsorized to 300%. See Table 3 for summary statistics of the annual fund returns.

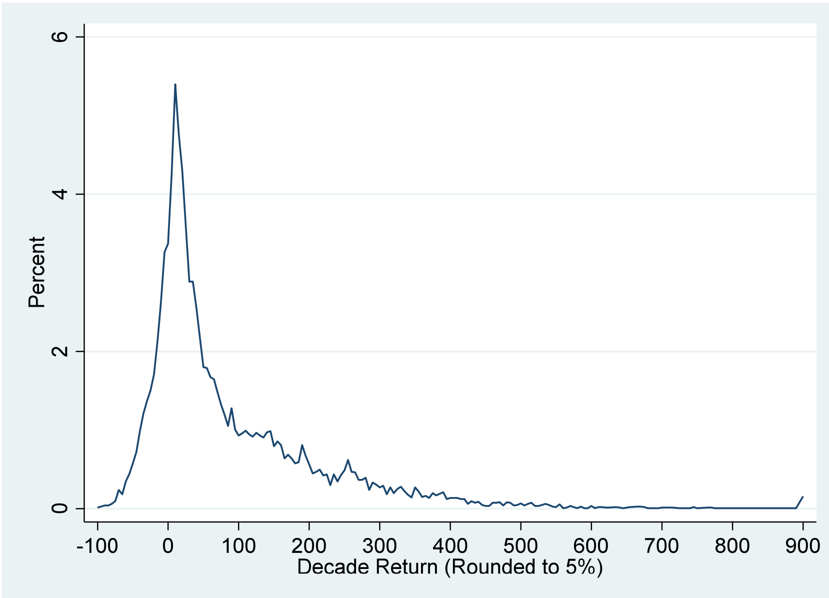

Figure 3, Panel C
Frequency distribution of decade mutual fund returns
This figure plots the frequency distribution of decade mutual fund returns, rounded to the nearest 5%, over each of the three ten-year periods: 1991-2000, 2001-2010, and 2011-2020. Returns above 900% are winsorized to 900%. See Table 4 for summary statistics of the decade fund returns.

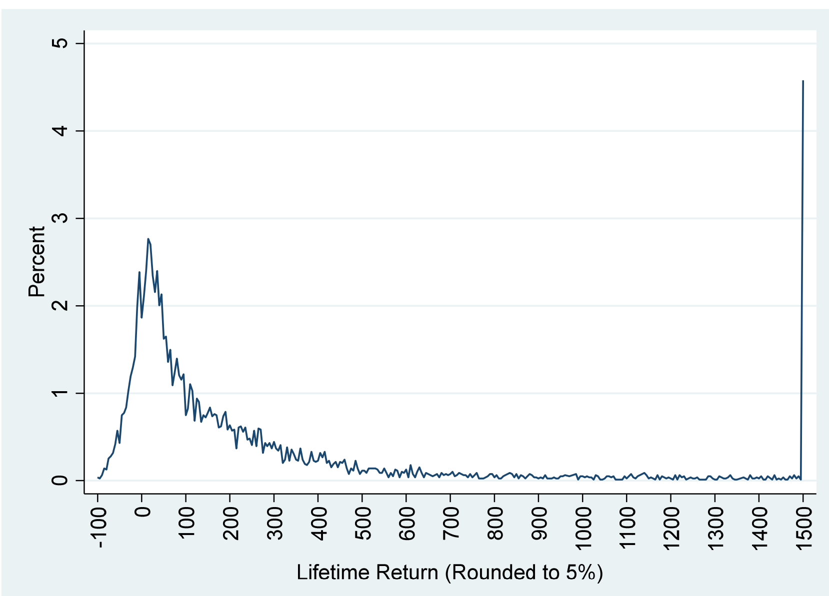

Figure 3, Panel D
Frequency distribution of lifetime mutual fund returns
This figure plots the frequency distribution of lifetime returns, rounded to the nearest 5%, to sample mutual funds from 1991 to 2020. Returns above 1500% are winsorized to 1500%. See Table 5 for summary statistics of the lifetime fund returns.

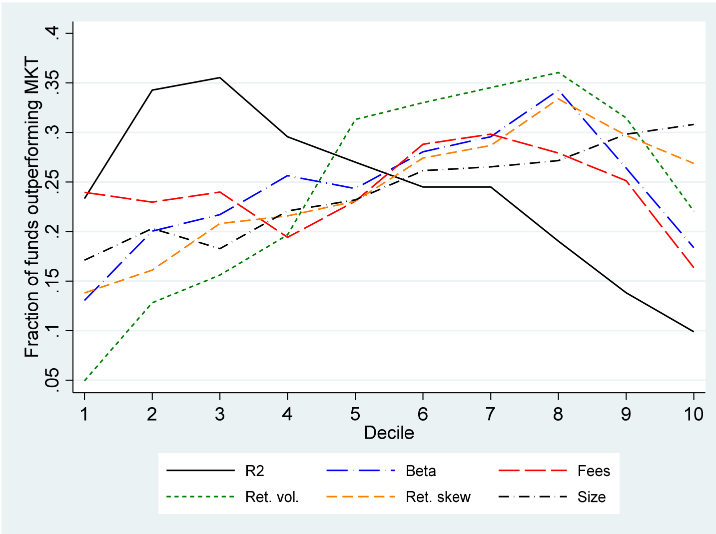

Figure 4, Panel A

Fraction of funds that outperform the market return over their lives

This figure plots the fraction of funds that outperform the CRSP value-weighted market return over their lives, for each decile of funds sorted on each of six fund characteristics. Fund R-squared is the R-squared of the regression of monthly excess fund return on monthly excess SPY return; Fund beta is the coefficient on excess SPY return. Fees refer to the average monthly fund expense ratio. Fund return volatility and skewness are the standard deviation and the skewness of monthly fund returns. Fund size percentile rank is the fund's average monthly percentile rank by fund TNA among all funds alive in the month.

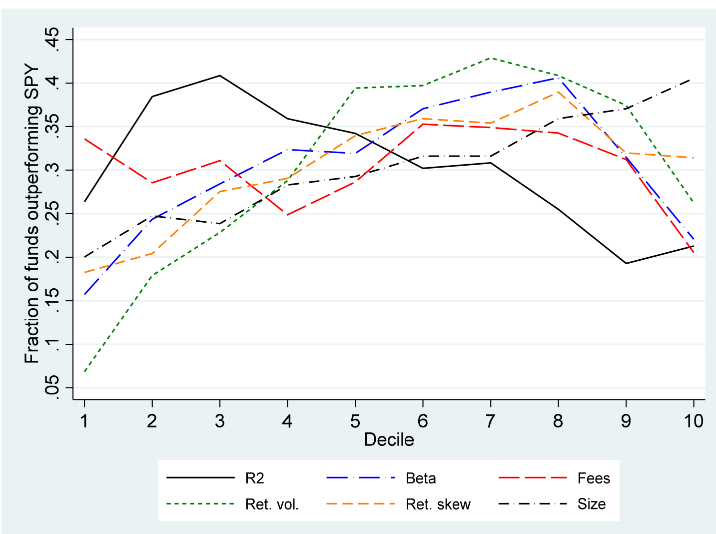

Figure 4, Panel B

Fraction of funds that outperform the SPY ETF over their lives

This figure plots the fraction of funds that outperform the SPDR S&P 500 ETF return (SPY) over their lives, for each decile of funds sorted on each of six fund characteristics. Fund R-squared is the R-squared of the regression of monthly excess fund return on monthly excess SPY return; Fund beta is the coefficient on excess SPY return. Fees refer to the average monthly fund expense ratio. Fund return volatility and skewness are the standard deviation and the skewness of monthly fund returns. Fund size percentile rank is the fund's average monthly percentile rank by fund TNA among all funds alive in the month.

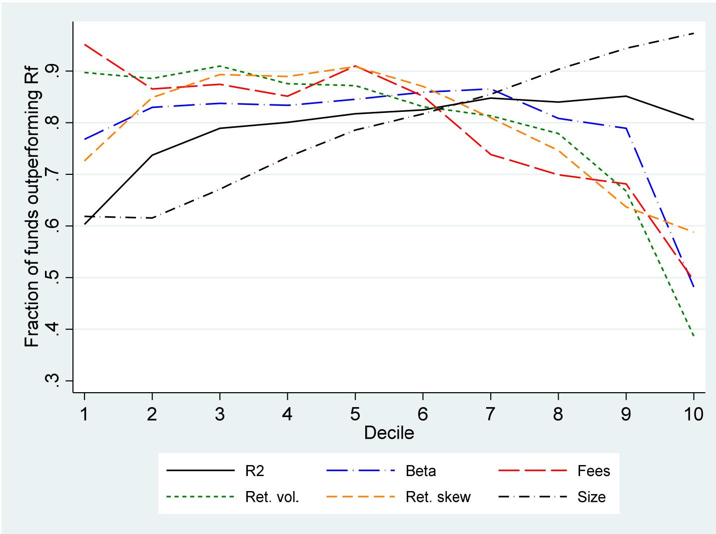

Figure 4, Panel C
Fraction of funds that outperform one-month T-Bills over their lives
This figure plots the fraction of funds that outperform the one-month T-Bill over their lives, for each decile of funds sorted on each of six fund characteristics. Fund R-squared is the R-squared of the regression of monthly excess fund return on monthly excess SPY return; Fund beta is the coefficient on excess SPY return. Fees refer to the average monthly fund expense ratio. Fund return volatility and skewness are the standard deviation and the skewness of monthly fund returns. Fund size percentile rank is the fund's average monthly percentile rank by fund TNA among all funds alive in the month.

Table 1: Pooled Sample monthly returns, expense ratios and TNA

This table reports summary statistics of fund expense ratios and Total Net Assets (TNA) at the fund-month level, as well as monthly fund returns and matching monthly returns to three benchmarks: the CRSP value-weighted market return, the SPDR S&P 500 ETF return (SPY), and the one-month T-Bill rate. A fund outperforms a benchmark in a month if its monthly return is greater than that of the benchmark. Our sample includes U.S. equity mutual funds from 1991 to 2020.

Variable	# unique funds	# fund-months	Mean	Median	Std. dev.	Skewness
Fund return (%), monthly	7,833	1,048,111	0.776	1.158	5.419	-0.425
Market return (%), monthly	7,833	1,048,111	0.882	1.380	4.496	-0.626
SPY return (%), monthly	7,833	1,048,111	0.835	1.328	4.332	-0.616
T-bill return (%), monthly	7,833	1,048,111	0.166	0.120	0.168	0.626
Outperform market	7,833	1,048,111	0.463	0.000	0.499	0.150
Outperform SPY	7,833	1,048,111	0.472	0.000	0.499	0.113
Outperform T-Bill	7,833	1,048,111	0.602	1.000	0.490	-0.415
Fees (%), monthly	7,833	1,048,111	0.095	0.094	0.049	1.583
TNA ($B), monthly	7,833	1,048,111	1.177	0.149	7.703	42.553

Table 2: Simulation Parameters and Outcomes

Panel A presents the parameters used in our simulations to demonstrate the effects of compounding on fund performance over various investment horizons. Within each round of simulation, we generate returns to 500 funds for t = 1 to 360 months. We define a "failure" function whereby a fund is limited to t 360 observations, so that the average fund life in the simulations matches the average fund life of 71.1 months in the sample. The simulation is repeated 10,000 times. Panels B-G present summary statistics of fund performance across the 10,000 simulations at the monthly, annual, decade, and lifetime horizons. We rely on randomly ordered SPY (the SPDR S&P 500 ETF) returns to proxy for market returns. Wealth ratio is the ratio of one plus fund buy-and-hold return to one plus SPY buy-and-hold return over different investment horizons: month, year, decade, or fund life. A fund outperforms a benchmark over an investment horizon if its buy-and-hold return is greater than that of the benchmark over matching months of the investment horizon. Non-market return in a month equals fund return in the month minus fund beta times SPY return in the month. Panels F and G report simulation outcomes at the "lifetime" horizon for subsamples that are delineated based on whether the estimate of alpha obtained from the monthly-horizon market model regression is positive or negative.

A. Parameters used in our simulations

	Mean	Std. Dev.
Monthly SPY return (%)	0.929	4.193
Monthly fund beta	1.024	0.200
Monthly fund alpha (%)	-0.131	0.150
Monthly residual fund return (%)	0.000	2.400
Monthly fund fees (%)	0.095	0.045

B. Simulated Monthly Returns

-	Mean	Median	Std. dev.	Skewness
SPY return (%)	0.929	1.330	4.188	-0.589
Fund return (%)	0.804	1.010	5.577	-0.277
Pre-fee fund return (%)	0.899	1.100	5.577	-0.277
Non-market fund return (%)	-0.149	-0.150	3.437	0.000
Fund outperforms SPY indicator	0.485	0.000	0.500	0.059
Pre-fee fund outperforms SPY indicator	0.497	0.000	0.500	0.014
Non-market return > 0 indicator	0.482	0.000	0.500	0.072
Wealth ratio	0.999	0.999	0.035	-0.012

C. Simulated Annual Returns

-	Mean	Median	Std. dev.	Skewness
SPY return (%)	11.228	10.460	15.648	0.263
Fund return (%)	10.140	8.460	19.286	0.557
Pre-fee fund return (%)	11.347	9.630	19.509	0.559
Non-market fund return (%)	-1.483	-1.720	8.463	0.276
Fund outperforms SPY indicator	0.433	0.000	0.495	0.272
Pre-fee fund outperforms SPY indicator	0.483	0.000	0.500	0.068
Non-market return > 0 indicator	0.411	0.000	0.492	0.364
Wealth ratio	0.989	0.986	0.091	0.310
Months with data per year	11.523	12.000	1.865	-4.175

D. Simulated Decade Returns

_	Mean	Median	Std. dev.	Skewness
SPY return (%)	119.722	91.430	110.958	1.363
Fund return (%)	107.005	63.780	140.074	3.041
Pre-fee fund return (%)	126.361	76.830	159.126	3.010
Non-market fund return (%)	-9.008	-9.460	24.307	0.878
Fund outperforms SPY indicator	0.362	0.000	0.481	0.575
Pre-fee fund outperforms SPY indicator	0.463	0.000	0.499	0.150
Non-market return > 0 indicator	0.313	0.000	0.464	0.807
Wealth ratio	0.942	0.926	0.288	1.318
Months with data per decade	80.530	0.930	41.007	-0.411

E. Simulated "Lifetime" Returns

_	Mean	Median	Std. dev.	Skewness
SPY return (%)	385.485	110.268	591.638	1.859
Fund return (%)	349.936	82.944	882.178	8.935
Pre-fee fund return (%)	473.974	98.588	1258.825	9.433
Non-market fund return (%)	-11.927	-11.363	33.679	2.417
Fund outperforms SPY indicator	0.353	0.000	0.478	0.614
Pre-fee fund outperforms SPY indicator	0.461	0.000	0.498	0.158
Non-market return > 0 indicator	0.300	0.000	0.458	0.873
Wealth ratio	0.938	0.912	0.466	5.831
Months with data per "lifetime"	125.643	87.000	107.542	1.001

F. Simulated "Lifetime" Returns, subsample with estimated monthly alpha > 0

_	Mean	Median	Std. dev.	Skewness
SPY return (%)	318.025	85.491	538.001	2.233
Fund return (%)	546.266	111.998	1333.009	6.943
Pre-fee fund return (%)	730.435	126.630	1904.673	7.354
Non-market fund return (%)	19.925	10.524	32.466	6.094
Fund outperforms SPY indicator	0.809	1.000	0.393	-1.574
Pre-fee fund outperforms SPY indicator	0.900	1.000	0.300	-2.669
Non-market return > 0 indicator	0.904	1.000	0.294	-2.747
Wealth ratio	1.267	1.124	0.584	7.323
Months with data per "lifetime"	110.244	71.000	101.174	1.254

G. Simulated "Lifetime" Returns, subsample with estimated monthly alpha < 0

_	Mean	Median	Std. dev.	Skewness
SPY return (%)	418.966	126.070	613.781	1.708
Fund return (%)	252.497	69.295	503.953	4.444
Pre-fee fund return (%)	346.692	85.168	722.344	4.626
Non-market fund return (%)	-27.736	-22.531	20.509	-0.869
Fund outperforms SPY indicator	0.127	0.000	0.333	2.240
Pre-fee fund outperforms SPY indicator	0.242	0.000	0.428	1.203
Non-market return > 0 indicator	0.000	0.000	0.000	
Wealth ratio	0.775	0.810	0.274	0.718
Months with data per "lifetime"	133.285	96.000	109.767	0.891

Table 3: Annual fund returns

In each year from 1991-2020, we compute buy-and-hold returns to the fund, as well as to CRSP value-weighted market portfolio, the SPDR S&P 500 ETF (SPY), and the one-month T-Bill during the matched months. The sample includes 92,844 fund/years. The Wealth ratio is the ratio of one plus fund buy-and-hold return to one plus market/SPY/T-Bill buy-and-hold return; it measures the year end wealth of an investor who invested $1 in the fund at the beginning of the year relative to the year-end wealth of another investor who invested $1 in the market/SPY/T-Bill at the beginning of the year. A fund outperforms a benchmark in a year if its buy-and-hold return is greater than that of the benchmark over the matched months. This table presents summary statistics for the pooled sample of fund-years. We test whether the wealth ratio equals one using the bootstrapped skewness-adjusted t-test proposed by Lyon, Barber, and Tsai (1999); we carry out t-test of whether the likelihood of a fund outperforming the market/SPY/T-Bill equals a half. ***, **, and * correspond to statistical significance at the 1%, 5% and 10% levels, respectively.

Variable	Mean	Median	Std. dev.	Skewness
Fund life (months)	11.3	12.0	2.2	-3.2
Wealth ratio w.r.t. market	0.990***	0.985	0.117	2.648
Wealth ratio w.r.t. SPY	0.994***	0.989	0.120	2.775
Outperform market	0.393***	0.000	0.488	0.438
Outperform SPY	0.411***	0.000	0.492	0.360
Outperform T-Bill	0.688***	1.000	0.463	-0.813
Fund buy-and-hold return (%)	9.465	10.287	21.114	0.566
Market buy-and-hold return (%)	10.685	12.362	17.332	-0.734
SPY buy-and-hold return (%)	10.121	12.680	16.763	-0.796

Table 4: Decade fund returns

In each of three decades (1991-2000, 2001-2010, 2011-2020) we compute buy-and-hold returns to the fund, as well as to CRSP value-weighted market portfolio, the SPDR S&P 500 ETF (SPY), and the one-month T-Bill during the matched months. The sample includes 14,710 fund/decades. Wealth ratio is the ratio of one plus fund buy-and-hold return to one plus market/SPY/T-Bill buy-and-hold return; it measures the year end wealth of an investor who invested $1 in the fund at the beginning of the year relative to the year-end wealth of another investor who invested $1 in the market/SPY/T-Bill at the beginning of the year. A fund outperforms a benchmark in a year if its buy-and-hold return is greater than that of the benchmark over the matched months. This table presents summary statistics for the pooled sample of fund-decades. We test whether the wealth ratio equals one using the bootstrapped skewness-adjusted t-test proposed by Lyon, Barber, and Tsai (1999); we carry out t-test of whether the likelihood of a fund outperforming the market/SPY/T-Bill equals a half. ***, **, and * correspond to statistical significance at the 1%, 5% and 10% levels, respectively.

Variable	Mean	Median	Std. dev.	Skewness
Fund life (months)	71.3	70.0	42.5	-0.1
Wealth ratio w.r.t. market	0.953***	0.928	0.401	8.476
Wealth ratio w.r.t. SPY	0.983*	0.951	0.436	9.123
Outperform market	0.341***	0.000	0.474	0.673
Outperform SPY	0.383***	0.000	0.486	0.481
Outperform T-Bill	0.740***	1.000	0.439	-1.093
Fund buy-and-hold return (%)	86.897	39.382	128.857	2.640
Market buy-and-hold return (%)	104.568	36.456	122.455	0.926
SPY buy-and-hold return (%)	100.140	31.199	121.835	0.926

Table 5: Lifetime fund returns

Panel A presents summary statistics of lifetime fund returns for all 7,883 sample funds while Panel B presents lifetime outcomes for subsamples defined by fund life. We compute lifetime buy-and-hold returns to the fund, and to the CRSP market portfolio, the SPDR S&P 500 ETF (SPY), and the one-month T-Bill over each fund's life. Wealth ratio is the ratio of one plus lifetime fund return to one plus the benchmark return. A fund outperforms a benchmark if its lifetime return is greater than benchmark returns over the fund's life. We test whether the wealth ratio equals one using the bootstrapped skewness-adjusted t-test proposed by Lyon, Barber, and Tsai (1999); we carry out t-test of whether the likelihood of a fund outperforming the market/SPY/T-Bill equals a half. ***, **, and * correspond to statistical significance at the 1%, 5% and 10% levels, respectively. Panels C and D compare monthly arithmetic and geometric mean return of mutual funds to the SPDR S&P 500 ETF (SPY), and the one-month T-Bill, over the fund's life, respectively.

Panel A: Lifetime fund returns

Variable	Mean	Median	Std. dev.	Skewness	5th pctl	95th pctl
Fund life (months)	133.0	112.0	97.5	0.8	19.0	347.0
Wealth ratio w.r.t. market	0.884	0.848	0.407	9.085	0.447	1.464
Wealth ratio w.r.t. SPY	0.935*	0.886	0.445	7.762	0.467	1.598
Outperform market	0.241***	0.000	0.428	1.209	0.000	1.000
Outperform SPY	0.303***	0.000	0.460	0.858	0.000	1.000
Outperform T-Bill	0.792***	1.000	0.406	-1.438	0.000	1.000
Fund buy-and-hold return (%)	294.354	95.093	636.757	6.398	-32.705	1420.665
Market buy-and-hold return (%)	332.887	152.636	514.585	2.644	-14.606	1642.569
SPY buy-and-hold return (%)	297.693	146.603	448.472	2.555	-16.850	1474.489

Panel B: Lifetime fund returns by fund life grouping

Fund life	N	Fund life (months)	Wealth ratio w.r.t. market	Wealth ratio w.r.t. SPY	Outperform market	Outperform SPY	Outperform T-Bill
[1y, 5y]	2336	34.5	0.880	0.896	0.208	0.246	0.584
(5y, 10y]	1814	88.1	0.837	0.865	0.195	0.244	0.754
(10y, 15y]	1675	147.7	0.853	0.907	0.235	0.287	0.896
(15y, 30y]	2058	272.3	0.955	1.064	0.326	0.433	0.977

Panel C: Differences in monthly arithmetic/geometric mean return between mutual funds and SPY

Variable	N	Mean	Fraction positive
All funds			
Fund arithmetic mean return minus SPY arithmetic mean (%)	7883	-0.168	0.351
Fund geometric mean return minus SPY geometric mean (%)	7883	-0.229	0.303
Funds whose arithmetic mean > SPY arithmetic mean			
Fund arithmetic mean return minus SPY arithmetic mean (%)	2769	0.255	1.000
Fund geometric mean return minus SPY geometric mean (%)	2769	0.182	0.848

Panel D: Differences in monthly arithmetic/geometric mean return between mutual funds and T-bills

Variable	N	Mean	Fraction positive
All funds			
Fund arithmetic mean return minus T-bill arithmetic mean (%)	7883	0.456	0.842
Fund geometric mean return minus T-bill geometric mean (%)	7883	0.295	0.792
Funds whose arithmetic mean > T-bill arithmetic mean			
Fund arithmetic mean return minus T-bill arithmetic mean (%)	6639	0.721	1.000
Fund geometric mean return minus T-bill geometric mean (%)	6639	0.580	0.940

Table 6: Fund size and fund returns at monthly, annual, decade, and lifetime horizons

We compile the cross-sectional distribution of fund sizes for each sample month from 1991 to 2020, and record the percentile position of each fund within the size distribution for each month. We then determine if the size of a given fund exceeds the 25th, 50th, or 75th percentile of the size distribution in any month of the sample. We then construct three subsamples that exclude smaller funds. The first ("all but small" funds) sample excludes any fund whose size never reaches the 25th percentile, the second ("medium and larger" funds) sample excludes any fund whose size never reaches the 50th percentile, and the third ("large" funds) excludes any fund whose size never reaches the 75th percentile. Within these subsamples, we study returns beginning with the first month that a fund is larger than the threshold size.

Variable	# Fund-periods	Mean	Median	Skewness	% > SPY
	Monthly fund return (%)				
All funds	1,048,111	0.776	1.158	-0.425	0.472
All but small funds	876,872	0.800	1.192	-0.420	0.474
Medium and larger funds	661,122	0.810	1.211	-0.421	0.474
Large funds	387,924	0.821	1.223	-0.462	0.474
	Annual fund return (%)				
All funds	92,844	9.465	10.287	0.566	0.411
All but small funds	77,158	9.780	10.703	0.358	0.417
Medium and larger funds	57,780	9.942	11.012	0.109	0.418
Large funds	33,666	10.161	11.281	0.051	0.418
	Decade fund return (%)				
All funds	14,710	86.897	39.382	2.640	0.383
All but small funds	11,669	94.483	46.022	2.601	0.396
Medium and larger funds	8,422	99.780	50.996	2.219	0.399
Large funds	4,717	107.250	60.754	2.136	0.397
	Lifetime fund return (%)				
All funds	7,883	294.354	95.093	6.398	0.303
All but small funds	6,021	330.736	134.793	6.336	0.311
Medium and larger funds	4,206	356.314	170.903	4.782	0.303
Large funds	2,307	390.510	196.326	4.704	0.296

Table 7: Fund performance over the 2001-2020 period

This table reports summary statistics (mean and median across funds) for fund returns between 2001 and 2020, for the 2,849 domestic non-index equity mutual funds that existed at the beginning of January 2001, as well as returns to four benchmarks over matching months. The four benchmarks are: the SPY ETF, the S&P 500 index return (including dividends), the S&P 1500 index return (including dividends), and the one-month T-Bill.

Variable	Fund	SPY	SP 500	SP 1500	T-bill
Mean Return (%)	137.01	124.66	127.64	140.33	20.74
Median Return (%)	31.46	27.02	28.06	35.36	24.07
% of Funds that Outperformed the Benchmark		40.75	39.66	33.45	54.20
% of Funds that Outperformed, survived to 2020		18.74	17.97	15.69	34.10
% of Funds that Outperformed, exited early		22.01	21.69	17.76	20.00

Table 8: Returns to portfolios of mutual funds at monthly, annual, decade, and lifetime horizons

We randomly draw a portfolio of 10 mutual funds for each month from 1991 to 2020, and compute the equal-weighted (EW) and value-weighted (VW) monthly portfolio returns. We then calculate the portfolio's buy-and-hold returns in each year from 1991 to 2020, over each of three decades (1991-2000, 2001-2010, 2011-2020), and over the full sample lifetime (30 years from 1991-2020). The process is repeated 10,000 times. The left part of Panel A presents summary statistics for EW portfolio returns at the monthly, annual, decade, and lifetime horizons. The right part of Panel A presents the mean of the indicator for whether the fund portfolio outperforms the CRSP market portfolio, the SPDR S&P 500 ETF (SPY), and the one-month T-Bill at each of the four horizons. A fund portfolio outperforms a benchmark if its buy-and-hold return is greater than that of the benchmark over the same horizon. Panel B presents the same summary statistics for the VW portfolio returns. We carry out t-test of whether the likelihood of a fund outperforming the market/SPY/T-Bill equals a half. ***, **, and * correspond to statistical significance at the 1%, 5% and 10% levels, respectively.

A. Summary statistics of EW fund portfolio returns

	# portfolio-	EW Fund Portfolio Return				Outperform		
Variable	periods	Mean	Median	Std. dev.	Skew.	Market	SPY	T-bill
EW fund monthly return (%)	3,600,000	0.866	1.312	4.531	-0.682	0.444***	0.475***	0.620***
EW fund annual return (%)	300,000	11.045	12.941	17.121	-0.857	0.328***	0.416***	0.737***
EW fund decade return (%)	30,000	182.242	183.790	131.368	0.159	0.192***	0.322***	0.843***
EW fund lifetime return (%)	10,000	1460.078	1439.192	274.399	0.583	0.008***	0.055***	1.000***

B. Summary statistics of VW fund portfolio returns

	# portfolio-	VW Fund Portfolio Return				Outperform		
Variable	periods	Mean	Median	Std. dev.	Skew.	Market	SPY	T-bill
VW fund monthly return (%)	3,600,000	0.874	1.308	4.651	-0.617	0.454***	0.481***	0.620***
VW fund annual return (%)	300,000	11.213	13.043	17.923	-0.791	0.372***	0.430***	0.742***
VW fund decade return (%)	30,000	186.902	198.982	140.029	0.305	0.209***	0.308***	0.808***
VW fund lifetime return (%)	10,000	1510.810	1451.783	463.178	0.902	0.070***	0.166***	1.000***

Table 9: Lifetime fund returns after adding back fund expense or subtracting fund alpha

We estimate each fund's alpha by regressing its monthly excess (over the T-bill rate) return on the excess SPDR S&P 500 ETF (SPY) return. Panel A presents summary statistics for estimated monthly alphas. Panel B presents summary statistics for lifetime fund returns, when either fund expenses are added to or estimated fund alpha is subtracted from each monthly fund returns. We compute lifetime buy-and-hold returns to the fund, and to the CRSP market portfolio, the SPDR S&P 500 ETF (SPY), and the one-month T-Bill over the fund's life. Wealth ratio is the ratio of one plus lifetime fund return to one plus market/SPY/T-Bill buy-and-hold return over the fund's life. A fund outperforms a benchmark if its lifetime return is greater than benchmark returns over the fund's life. We test whether the wealth ratio equals one using the bootstrapped skewness-adjusted t-test proposed by Lyon, Barber, and Tsai (1999); we carry out t-test of whether the likelihood of a fund outperforming the market/SPY/T-Bill equals a half. ***, **, and * correspond to statistical significance at the 1%, 5% and 10% levels, respectively.

We compute each fund's compound non-market return (alpha plus the residual in the preceding regression) over its life. Panel C presents summary statistics of the compound non-market lifetime return, for the full sample as well as for subsamples where the estimated alphas are positive vs. negative. We carry out t-tests of whether the average compound non-market lifetime return equals zero and whether the fraction of funds with positive compound non-market lifetime returns equals a half.

A. Summary statistics of fund expense and alpha's

Variable	N	Mean	Median	Std. dev.	Skewness
Fees (%), monthly	1,048,111	0.095	0.094	0.049	1.583
Fund alpha against SPY (%)	7,883	-0.131	-0.081	0.476	-1.164

B. Lifetime fund returns after adding back fund expense or subtracting fund alpha

Variable	Mean	Median	Std. dev.	Skew	Mean	Median	Std. dev.	Skew
	Add fees				Subtract fund alpha against SPY			
Wealth ratio w.r.t. market	1.011	0.932	0.524	7.846	0.902***	0.929	0.169	0.125
Wealth ratio w.r.t. SPY	1.074***	0.971	0.585	6.841	0.949***	0.979	0.175	0.328
Outperform market	0.376***	0.000	0.484	0.513	0.207***	0.000	0.405	1.450
Outperform SPY	0.452***	0.000	0.498	0.191	0.362***	0.000	0.481	0.574
Outperform T-Bill	0.827***	1.000	0.378	-1.730	0.825***	1.000	0.380	-1.707
Fund buy-and-hold return (%)	393.636	115.572	891.499	6.634	269.794	126.116	416.902	2.803
Market buy-and-hold return (%)	332.887	152.636	514.585	2.644	332.887	152.636	514.585	2.644
SPY buy-and-hold return (%)	297.693	146.603	448.472	2.555	297.693	146.603	448.472	2.555

Panel C: Compound non-market return over fund life.

	N	Mean	Median	Std. dev.	Skewness	Fraction positive
All Funds	7883	-0.056***	-0.094	0.413	7.496	0.236***
Funds with Monthly Alpha > 0	2849	0.265***	0.125	0.506	9.542	0.880***
Funds with Monthly Alpha < 0	5034	-0.238***	-0.199	0.177	-1.144	0.000***

Table 10: Aggregate wealth enhancement due to equity mutual fund investment, measured at end-of-sample.

The first six rows of this table presents summary statistics for the increase in investor wealth, measured in $ billions, resulting from mutual fund investment, as defined by text expression (2) and measured as of December 2020, when the alternative asset that defines the opportunity cost of invested funds is based on three alternative benchmarks: the CRSP market portfolio return, the SPDR S&P 500 ETF return, and the one-month Treasure-bill return. Reported are the sum, mean, median, standard deviation, skewness of the fund-by-fund wealth increase outcomes, and the fraction of the 7,883 funds with positive aggregate wealth creation. The last two rows present summary statistics for the increase in investor wealth after adjusting for the fund's risk exposures. In particular, $(R_t – R_{at})$ in text expression (2) is replaced by $(\alpha + \varepsilon_t)$, where α is the fund's alpha estimated from time-series regression of the fund's monthly excess return on the SPY excess return and ε_t is the regression residual in month t, and the future value function is replaced by $FV_{t,T} = [1+ R_{ft+1} + \beta^* (R_{at+1} - R_{ft+1})]^* [1+ R_{ft+2} + \beta^* (R_{at+2} - R_{ft+2})]^* [1+ R_{ft+3} + \beta^* (R_{at+3} - R_{ft+3})]^* \dots ^*[1+ R_{fT} + \beta^* (R_{atT} - R_{fT})]$, where β is the fund's beta estimated from the time-series regression.

Variable	N	Sum	Mean	Median	SD	Skewness	% Pos.
Simple Difference vs. Benchmark							
Post-fee Fund Returns, MKT benchmark	7883	-2887.5	-0.366	-0.013	3.139	-1.901	0.200
Pre-fee Fund Returns, MKT benchmark	7883	-470.4	-0.060	-0.004	3.299	18.433	0.306
Post-fee Fund Returns, SPY benchmark	7883	-1308.4	-0.166	-0.007	2.869	13.040	0.254
Pre-fee Fund Returns, SPY benchmark	7883	936.0	0.119	-0.001	3.573	25.759	0.390
Post-fee Fund Returns, T-Bill benchmark	7883	8664.9	1.099	0.023	10.314	34.263	0.744
Pre-fee Fund Returns, T-Bill benchmark	7883	9500.9	1.205	0.033	10.651	32.849	0.780
Beta-Adjusted Difference vs. Benchmark							
Post-fee Fund Returns, SPY benchmark	7883	-1024.2	-0.130	-0.006	3.099	15.041	0.263
Pre-fee Fund Returns, SPY benchmark	7883	1234.6	0.157	-0.001	3.833	27.384	0.409

Appendix A: Sample Construction and Data Filters

We obtain data for the 1991 to 2020 period from the CRSP survivorship bias free Mutual Fund Database. We begin at 1991, as data regarding fund total net assets (TNA), which we use to aggregate fund returns across share classes, is not consistently available for earlier periods. We rely on the CRSP share class group number (crsp_cl_grp) in the fund names file. For funds without a CRSP share class group number, we identify share classes of the same fund based on fund names. When funds have multiple share classes CRSP fund names contain "/" or ";". The part of the fund name after the last "/" or ";" refers to the sub share class, while the prior part refers to the main fund name. For example, the fund named "MainStay Funds: MainStay Small Cap Growth Fund; Class A Shares" is Class A of the MainStay Small Cap Growth Fund; the fund named "Alliance Strategic Balanced Fund/A" is Class A of the Alliance Strategic Balanced Fund.

We study domestic equity funds (CRSP fund style code starting with "ED"), while excluding exchange traded funds, exchange traded notes (those with CRSP *et_flag* equal to "F" or "N"), funds that take short positions (CRSP fund style "EDYS"), commodity funds (CRSP fund style "EDSC") and real estate funds (CRSP fund style "EDSR"). We exclude target date funds, since these hold substantial non-equity positions. To exclude target date funds and college savings funds we remove all funds with names that contain a four-digit number between 1990 and 2050 and the word "target", except that we do not exclude funds with "Russell 2000" or "Russell2000" in their names.

We further exclude hedged funds (CRSP fund style of "EDYH" and Lipper fund style code of "LSE"), market neutral funds (CRSP fund style of "EDYH" and Lipper fund style code of "EMN") and absolute return funds (CRSP fund style of "EDYH" and Lipper fund style code of "ABR"). We also screen some funds CRSP style code starting with ED, but with names that are inconsistent with this categorization. Specifically, we exclude a fund with "VIX" in its name, funds with "Long/Short", "Long-Short", and "OTC/Short" in their names, funds whose name includes "ETF" or "ETN", leveraged funds with “1.25x”, “1.5x”, “2x”, “2.5x”, “3x”, and “4x” in their names, and one fixed income fund with “Government Portfolio” in its name.

Prior studies (e.g., Elton, Gruber, and Blake, 2001) have documented the presence of errors in the CRSP mutual fund data. We mitigate the effect of potentially influential errors by comparing large reported returns to those contained in the Morningstar Mutual Fund database, or if Morningstar data is not available to the returns implied by percentage changes in CRSP-reported NAV or TNA. Specifically, we identify 836 extreme fund returns based on a deviation from the same-month CRSP value- weighted market return of 30% or more. With the help of Professor Shuaiyu Chen, we are able to match 633 of these to monthly return data in the Morningstar mutual fund database. For 524 of these cases, the CRSP and Morningstar returns differ by less than 1% and we retain the CRSP return. For the remaining 109 cases we retain the Morningstar return, which in every instance is less extreme that the CRSP return. For the 203 instances that cannot be matched to Morningstar, we focus on the percentage change in the CRSP reported net-asset-value (NAV) as well as the TNA. We retain observations where the reported return deviates from both the NAV and TNA-implied returns by less than 30%. For 75 observations from 53 funds the deviation exceeds 30%, and we delete the associated funds from the sample. Finally, we exclude funds that have fewer than twelve months of non-missing return data. The sample employed here is also used by Bessembinder, Cooper, and Zhang (2022).

Appendix B The Relation Between Fund Wealth Enhancement and the “Dollar-Weighted” Fund Return

We rely here on the notation defined in section 5 of the text. The time series of cash flows experienced by investors in aggregate from time 0 to time T can be expressed as follows. The initial acquisition of fund shares results in $CF_0 = -M_0$. Interim cash flows, from $t = 1$ to $T - 1$, comprised of dividends and other net fund distributions, are $CF_t = D_t + F_t = M_{t-1}(1 + R_t) - M_t$. The final period ($T$) cash flow is the sum of the final dividend and distribution, as well as the liquidation value of the final position, $CF_T = D_T + F_T + M_T = M_{T-1}(1 + R_T)$. The dollar-weighted return is the internal rate of return of this series of cash flows.

Alternatively, one can focus on the compound "future value" of the cash flows to shareholders in aggregate. As in Section 5 of the text, let $FV_{t,T}$ denote a future value factor obtained by compounding returns on the alternative asset from time t to time T. The end-of-sample (time T) value of each cash flow is $CF_t \times FV_{t,T}$. Summing across t, noting that for that for periods $t = 1$ to $t = T\text{-}1$ we can express $CF_t \times FV_{t,T} = M_{t-1}(1 + R_t)FV_{t,T} - M_t(1 + R_{at+1})FV_{t+1,T}$, we obtain:

$$\sum_{t=0}^{T} CF_t\, FV_{t,T} =$$

$$M_0*(R_1 – R_{a1})\, FV_{1,T} + M_1*(R_2 – R_{a2})\, FV_{2,T} + \ldots + M_{T-2}*(R_{T-1} – R_{at-1})*FV_{T-1,T} + M_{T-1}*(R_T – R_{aT}). \qquad (3)$$

The right side of (3) is identical to the right side of text expression (2). That is, the enhancement in end-of-period investor wealth due to investing in the mutual fund instead of the alternative asset is also the compound future value of the series of cash flows that defines investors' dollar-weighted return, when the return on the alternative asset is used to compound forward.