Scale Constrains on Value Creation in the Mutual fund industry

Georgios Spais and Nikolaos Kourogenis*

Department of Banking and Financial Management

University of Piraeus

August 2025

Abstract

This paper investigates the role of diseconomies of scale in constraining value creation within the U.S. mutual fund industry. While traditional performance measures such as net alpha typically suggest little or no evidence of managerial skill, the concept of added value—the product of gross alpha and fund size—offers a more robust measure of a manager's ability to extract value from capital markets. Building on the theoretical framework of Berk and Green (2004), we employ a novel bootstrapping methodology inspired by Fama and French (2010) to compare the empirical distribution of added value with a simulated distribution constructed under the null of no scale effects. Using a comprehensive dataset of 2,331 actively managed domestic (US) equity mutual funds from 1993 to 2022, we find that fund size significantly constrains the capacity of skilled managers to generate added value. The effect is particularly pronounced in the upper tail of the distribution, where managers with strong skills are disproportionately affected, whereas unskilled managers exhibit little sensitivity to scale. These results provide strong empirical support for the presence of diseconomies of scale in active management and contribute to the literature by emphasizing value creation over return-based metrics as the appropriate measure of managerial skill.

1 Introduction

A significant strand of the financial literature focuses on the investigation of whether skill exists in the active mutual fund industry. Skill typically refers to a manager's ability to select securities and time the market effectively. Two primary measures are commonly used to assess managerial skill: net alpha (or simply alpha), which captures the risk-adjusted return delivered to investors in excess of a passive benchmark, and added

*Corresponding author: Nikolaos Kourogenis, Department of Banking and Financial Management, University of Piraeus. nkourog@unipi.gr +306974632550

value, defined as the average amount of money a manager extracts from the capital markets, measured as the average product of gross alpha and the fund's lagged size.

These two measures often lead to different conclusions regarding the existence of skill in the mutual fund industry. The average net alpha is typically zero or, in some cases, negative (e.g., Jensen 1968; Fama and French 2010; Barras, Scaillet, and Wermers 2010; Agyei-Ampomah et al. 2015; Leippold and Rueegg 2020), with only a small fraction of mutual funds generating significantly positive alphas (e.g., Kosowski et al. 2006; Fama and French 2010; Barras, Scaillet, and Wermers 2010; Huang et al. 2023). Moreover, when positive alphas do occur, the outperformance tends to persist only over short horizons — typically less than one year (e.g., Jensen 1968; Hendricks, Patel, and Zeckhauser 1993; Malkiel 1995; Carhart 1997; Cuthbertson, Nitzsche, and O'Sullivan 2022) — suggesting weak or negligible evidence of skill.

In contrast to this evidence, the average added value is generally positive and often persists over much longer horizons, extending up to ten years (e.g., Berk and Van Binsbergen 2015; Barras, Scaillet, and Wermers 2010; Cuthbertson, Nitzsche, and O'Sullivan 2022), providing stronger evidence of managerial skill.

These empirical findings give rise to three theoretical challenges. First — and most obviously — why do these two measures of skill yield conflicting conclusions? Second, an important principle in economics, which suggests that economic agents earn economic rent if and only if they have a competitive advantage, seems to be violated, given the net alpha-related evidence. In this case, mutual fund managers extract high fees from investors despite the absence of a significant competitive advantage (since they do not deliver superior risk-adjusted returns relative to passive benchmarks). Third, mutual fund investors tend to chase past performance (as documented in, e.g., Chevalier and Ellison 1997; Sirri and Tufano 1998; Del Guercio and Tkac 2002; Berk and Van Binsbergen 2016; Ben-David et al. 2022), even though outperformance rarely persists. This behavior possibly suggests a degree of irrationality among mutual fund investors.

A possible explanation of this counter-intuitive evidence may be sought in the theoretical model of Berk and Green 2004 and its assumption of 'diseconomies of scale'. The term diseconomies of scale implies that as a fund grows, the manager's ability to generate superior returns diminishes. In this framework, investors infer managerial skill from past performance and allocate capital accordingly. As capital flows into skilled managers' funds, the resulting increase in size erodes their ability to outperform, driving alphas to zero in equilibrium. Consequently, alpha becomes a misleading indicator of skill, and the added value emerges as the more appropriate measure of managerial ability (Berk and Van Binsbergen 2015).

In this paper, we provide empirical evidence that supports Berk's assumption, namely, the existence of diseconomies of scale, in the mutual fund industry. Previous studies examining scale effects typically focus on the relationship between fund size and performance, most often measured by alpha. However, if expected alpha is driven to zero in equilibrium, as theoretical models suggest, the empirical identification of a relationship between fund size and performance (as measured by alpha) becomes fundamentally challenging.

For example, J. Chen et al. 2004 provides evidence of a negative relationship between fund size and performance, attributing this pattern to liquidity constraints. In contrast, Elton, Gruber, and Blake 2012 report a positive but statistically insignificant association, arguing that larger funds may benefit from advantages such as access to skilled traders or increased attention from high-quality analysts. Furthermore, several recent studies that attempt to address the potential endogeneity between size and performance also fail to reach a consensus, suggesting that empirical results are highly sensitive to the methodology employed.[1]

Our approach deviates from prior research by focusing on the effect of size on added value rather than solely on alpha. If diseconomies of scale are present in the mutual fund industry, then size should impose constraints on the value a manager can extract from the market. Added value is the average product of the gross alpha with lag size ($V = E(a_t^g q_{t-1})$ and can be written as the product of average gross alpha with average lag size plus the covariance of gross alpha and lag size ($V = E(a_t^g)E(q_{t-1}) + Cov(a_t^g, q_{t-1})$). Under the assumption of diseconomies of scale, the first term reflects value created through skill, while the second term (covariance) is expected to be negative, indicating value destruction due to constrains imposed by size.

We further deviate from the existing literature by adopting a bootstrapping methodology motivated by the luck vs skill approach of Fama and French 2010, where we compare the actual estimated distribution of mutual fund added value, with a simulated distribution of added value constructed under the null of no diseconomies of scale. The distributional comparison allows us to account for a potential heterogeneity in the size effect, assess differences not just on average, but across various parts of the distribution and for deviations from normality in the distribution of added value.[2]

Our analysis reveals that, for most funds, the value a manager can add is constrained by size — consistent with diminishing performance as fund size increases. This effect is particularly strong among skilled managers who generate positive added value, while it largely disappears for unskilled managers who destroy value. These findings suggest that managerial skill and the strength of the scale effect are positively correlated.

We contribute to the literature in two key directions. First, we offer new insights into the scale effect puzzle by shifting the focus from return performance (alpha) to value creation, and examining how diseconomies of scale constrain managers' ability to generate added value. Second, we introduce a novel methodological approach for studying diseconomies of scale in the active management industry. Specifically, we extend the simulation-based framework of Fama and French 2010 (originally developed to distinguish skill from luck in alpha performance) to test for evidence of scale effects using the added value measure and addressing important limitations identified in recent critiques (e.g., Harvey and Liu 2022).

The remainder of the paper is structured as follows. Section 2 discusses the theoretical implications

[1] A more detailed review of the literature on diseconomies of scale is provided in the next section.

[2] This approach also allows us to account for a "lucky" relationship between size and added value. For instance, a fund that experiences one or two years of abnormally high (but luck-driven) performance will likely attract significant inflows, thereby increasing its size. Subsequently, if performance deteriorates and added value decreases, a spurious negative effect of size could be observed. A more detailed discussion of this issue is provided in the Methodology section.

of diseconomies of scale in the mutual fund industry and reviews the related empirical literature. Section 3 outlines the methodology employed. Section 4 describes the data. Section 5 presents the main results. Section 6 provides robustness tests, and Section 7 concludes the paper.

2 Theory and Related Literature

2.1 Theoretical Background

In this subsection, we briefly discuss the theory presented in Berk and Green 2004 and Berk and Van Binsbergen 2015 theory, to highlight the importance and theoretical implications of economies of scale in the mutual fund industry.

Suppose the market consists of two types of investment opportunities: actively managed mutual funds and passively managed mutual funds. Actively managed funds differ in their managers' ability to generate expected returns in excess of those provided by passively managed mutual funds with similar risk.

The return-generating process of the mutual fund i is given by:

$$a_t^n = \phi - c(q_{t-1}) - f + \epsilon_t \tag{1}$$

where:

- a_t^n (the net alpha) is the excess return earned by an investor of the mutual fund, relative to passive benchmark with similar risk.
- ϕ reflects the stock selection skill of the fund manager. It is unknown to both the mutual fund manager and investors.
- $c(q_{t-1}) = cq_{t-1}$ is a unit cost function of total net assets q_{t-1}, reflecting the costs associated with active management.
- f represents the fees charged to investors as a percentage of total net assets under management.
- ϵ_t is a fund-specific return term with zero mean.

We assume that $c > 0$, which reflects the notion that larger trades will be associated with larger price impact and in turn lower returns. It is further assumed that investors and managers are symmetrically informed and form rational expectations.

Investors compete with each other to find positive alpha funds ($a_t^n > 0$), withdraw money from funds with negative alphas and allocate capital to funds with positive alphas. As a result the cost of active investing

$c(q)$ (decreasing returns to scale) drives the average net alpha to zero:

$$E_{t-1}(a_t^n) = 0 \tag{2}$$

These simple assumptions lead to three very important implications.

1. From an investor's perspective, it is equivalent to invest in actively or passively managed mutual funds.

2. Although a mutual fund investor earns on average a zero net alpha, the average (gross) alpha of the fund is positive.
$$a_t^g = a_t^n + f \Rightarrow E_{t-1}(a_t^g) = f > 0$$

3. The net alpha does not measure managerial skill; rather, it only indicates whether a mutual fund is overfunded or underfunded.

Since net alpha cannot measure skill, how should we measure skill? An obvious alternative is the gross alpha, which measures the additional return earned by the fund relative to the benchmark. However, as we stated earlier, the average gross alphas are equal to the fees an investor is charged. Do managers choose the fees charged to investors in way that reflects their skill? As we formally show, gross alpha does not provide a skill measure.

Berk and Van Binsbergen 2015 propose a different measure: The value a fund manager extracts from the capital markets, which is the gross alpha of the fund multiplied by the total net assets of the fund:

$$V = E(a_t^g q_{t-1}) \tag{3}$$

This measure is referred to in the literature as added value, and measures the average value a fund manager extracts from the market with active investing.

We will formally show that added value can measure skill, whereas gross alpha cannot. Before we move on to the relative proposition, we need to define what we mean by a "measure of skill". We will use the definitions of Berk and Van Binsbergen 2015:

Definition 1. *A variable is said to measure skill if it quantitatively measures the amount of money a manager extracts from the market.*

The main problem with this definition is that skill, is measured in dollars and almost every time a return measure will not measure skill. So, to evaluate the usefulness of return measures, Berk and Van Binsbergen 2015 proposed an additional definition:

Definition 2. *A proxy of skill is any positive linear transformation of a skill measure.*

With Definition 2, a proxy for skill is any measure that is proportional to a skill measure. This allows return-based measures to be transformed into skill measures through an appropriate linear transformation.

Proposition 1. *Under the assumptions of rational investors, competitive financial markets, decreasing returns to scale ($a^g(q) = a_0 - cq$), managers that optimize value and are allowed to index part of the portfolio, the following hold:*

1. *The only condition under which the gross alpha measures skill, is if all managers manage \$1, and the only condition under which the gross alpha is a proxy of skill, is if the managers set their fees to ensure that all funds have the same total net assets.*

2. *Value added, the product of total net assets and gross alpha, always measures skill.*

Proof. A simple proof is provided in the appendix. □

This proposition highlights the importance of scale in measuring skill. They only case the gross alpha measures skill, is when all managers manage funds of the same size, making scale unimportant.

2.2 Previous Empirical Research

Empirical research on mutual fund performance provides results that are broadly consistent with the implications of economies of scale, as described in the model of Berk and Green 2004. Studies examining average net mutual fund performance often report zero or negative average alphas (e.g., Jensen 1968, Fama and French 2010, Barras, Scaillet, and Wermers 2010, Agyei-Ampomah et al. 2015, Leippold and Rueegg 2020). Negative average alphas are typically found in studies that use factor-based benchmarks — most notably the four-factor model of Carhart 1997. In contrast, studies employing market index-based benchmarks tend to report average alphas closer to zero.

However, the reliability of alpha estimates from factor-based models has come under scrutiny. Berk and Van Binsbergen 2015 and Agyei-Ampomah et al. 2015 argue that risk factor based benchmarks can underestimate mutual fund performance. The alpha estimated from a benchmark factor model is as the excess return that can not be attributed to the risk implied by the factor model. Moreover, the extent to which factor models price risk correctly is a subject of debate in the recent financial literature. Specifically, market factors used in the risk models are interpreted as alternative passive investment opportunities (or trading strategies) rather than risk factors. Yet, this alternative framing also presents problems: factor portfolios, though theoretically tradable, are not always accessible or even known to investors, especially considering that mutual fund datasets often begin well before these factor strategies were formally identified.

In addition to the findings that average mutual fund performance is generally not positive, several studies suggest that performance persistence, defined as the continuation of superior alpha over time, is short-lived.

Typically, any positive alpha persists for less than one year, indicating that past outperformance is not a reliable predictor of future success. Studies supporting this view include Jensen 1968, Hendricks, Patel, and Zeckhauser 1993, Malkiel 1995, Carhart 1997, and Cuthbertson, Nitzsche, and O'Sullivan 2022.

Overall, the empirical evidence points to weak or nonexistent skill among mutual fund managers, at least under the assumption that alpha is an appropriate measure of skill. At the same time, these findings are consistent with the existence of decreasing returns to scale in active management.

Berk and Van Binsbergen 2015 reexamined the evidence of skill in the mutual fund industry and introduced the added value as a new measure of managerial skill. They found that, the average fund generates positive added value, with significant cross-sectional differences in the persistence of value created by the fund managers. Notably, this persistence lasts much longer than return-based performance measures—extending up to 10 years.

Subsequent studies support these findings. Barras, Gagliardini, and Scaillet 2022 analyze the distribution of added value using a non-parametric kernel density approach, while Xu and T. Chen 2024 apply a methodology inspired by the residual income model from the accounting literature to assess value creation. Furthermore, Binsbergen et al. 2024 examine added value across different investment horizons, showing that low-turnover[3] funds generate value only over longer horizons, whereas high-turnover funds create substantial value within very short horizons (often within the first two weeks).

Overall, the empirical findings on managerial skill discussed so far align with the implications of decreasing returns to scale. However, they can also hold even in their absence. The lack of persistence in mutual fund returns may suggest a lack of skill, while the persistence of added value could stem from non-stationarity in total net assets. Additionally, positive added value may be partly driven by effective marketing. For instance, Roussanov, Ruan, and Wei 2021 argue that marketing is nearly as important as performance and fees in determining fund size. Given these considerations, a substantial body of research directly examines empirical evidence of decreasing returns to scale.

Early studies primarily use portfolio sorting and cross-sectional regression methodologies. J. Chen et al. 2004 find evidence of negative relations between size and performance, attributing this relation to liquidity constraints. This relationship is more pronounced in funds that invest in small-cap stocks (stocks of smaller companies), suggesting that larger funds face difficulties trading these stocks without significantly impacting prices. Furthermore, Yan 2008 analyze stock transactions data along with detailed stock-holdings of mutual funds Their findings support the results of J. Chen et al. 2004 and further highlight that the negative relation between size and performance is stronger among funds that hold less liquid portfolios.

In contrast, Adams, Hayunga, and Mansi 2018 provide evidence that the negative relation between size and perfomance dissappears in J. Chen et al. 2004 paper after accounting for multivariate outliers. Similarly,

[3]The turnover ratio measures a mutual fund's annual trading activity. Low-turnover funds trade infrequently within a year, whereas high-turnover funds trade more actively.

Elton, Gruber, and Blake 2012 document a positive but statistically insignificant relation between size and performance. They contend that the negative effect of scale may be offset by declining expense ratios as fund size increases. Moreover, they suggest that larger funds might benefit from better access to top traders or receive more attention from high-quality analysts, which could contribute to a positive relationship between size and performance. In the same spirit of a positive relation between size and performance, Ma, Tang, and Gomez 2019 find that managers of larger and more complex mutual funds are more often compensated with explicit performance-based incentives. They suggest that the improved performance of these funds may be a result of stronger incentives for managers to perform well. Bhojraj, JUN CHO, and Yehuda 2012 suggest that larger fund families performed better before 2000 due to access to private information from investment banks, giving them an edge over smaller funds. This advantage disappeared after the SEC implemented fair disclosure rules, which leveled the playing field.

Ferreira et al. 2013, in a cross country study, find that decreasing return scale holds only in the US mutual fund market, international mutual funds and mutual funds from other countries are not affected by their size. They argue that international funds are less affected by a lack of new investment opportunities as the fund grows, as they are not restricted to invest in their local market. The evidence of decreasing return to scale in the US mutual fund industry is related to liquidity constraints faced by funds that, by virtue of their style, have to invest in small and domestic stocks.

A concern with all these studies is that the fund size and performance relation may be endogenous. For example managerial skill might influence both fund size and fund performance. Thus, an estimation procedure that does not account for skill will suffer from omitted variable bias. Recognizing this, more recent studies employ methodologies specifically designed to correct for endogeneity.

To deal with endogeneity, Pástor, Stambaugh, and Taylor 2015 apply a fixed effect recursive demeaning methodology. They find an insignificant relation between size and performance (alpha) at the fund level, but report a significant negative relation at the industry level. Phillips, Pukthuanthong, and Rau 2018, using an instrumental variable approach motivated by investor responses to changes in mutual funds' holding-period reported returns, also find an insignificant relation between fund size and performance. Reuter and Zitzewitz 2021, use a regression discontinuity approach and find no statistically significant relation between size and performance, concluding that the lack of performance persistence in mutual fund returns implies either a lack of managerial skill, or that it may arise from competitive pressures among rival mutual funds. Additionally, Hoberg, Kumar, and Prabhala 2018 provide evidence that funds that outperform generate positive future alpha when they face less competition.

Other studies also account for endogeneity concerns and find evidence of decreasing return to scale. For instance, Zhu 2018 propose a correction of the recursive demeaning approach of the Pástor, Stambaugh, and Taylor 2015 and find a negative and statistical significant relation between size and performance. They also find much stronger evidence of decreasing returns to scale with a log-linear specification (t-statistic of -13.32) compared to a linear specification (t-statistic of -2.03). McLemore 2019, uses mutual fund mergers as an exogenous shock to fund size. He finds that acquiring funds' performance deteriorates following the size increase caused by mergers. He also shows that this decline is not driven by higher performance before the merger, nor by integration costs after the event, but rather by a decreasing returns to scale effect.

As the above discussion suggests, research on decreasing returns to scale does not provide uniformly conclusions. Previous studies mainly focus on the effect of size to alpha, but decreasing return to scale drive expected alphas to zero, making particularly difficult for a researcher to establish with consistency a negative relation between performance and size. Barras, Gagliardini, and Scaillet 2022 use a fund-level analysis of skill and scale. They find that most funds' performance is negatively affected by size. They attribute the difficulty of previous studies to establish a negative relation between size and performance to heterogeneity in the scale effect across funds. Pástor, Stambaugh, and Taylor 2020 and Busse et al. 2021, focus on the mechanism that may cause the decreasing return to scale and find evidence that larger funds are cheaper and trade less, better diversified funds hold less liquid stocks that are larger and trade more. They argue that these relationships between fund characteristics provide strong evidence of decreasing returns to scale.

This paper contributes to the existing literature by examining the effect of returns to scale on added value. While previous research has largely focused on the relationship between fund size and alpha, often yielding inconsistent conclusions due to alpha's tendency toward zero, added value exhibits substantial cross-sectional variation. This characteristic reduces the difficulty of empirically investigating a negative relationship between fund size and performance, providing the tools for a more concrete investigation of whether fund size impacts a manager's ability to create value.

3 Methodology

In this section we present the methodology we will follow to investigate whether dis-economies of scale are present in the mutual fund industry (a question for which prior research has yielded contradictory results). Contrary to previous research that focus on return performance, we focus on the effect of dis-economies of scale on added value.

Added value is a metric that captures the dollar value generated by a manager and varies significantly across funds. The effect of scale on added value stems directly from the definition of the metric:

$$V = E(a_t^g q_{t-1}) = E(a_t^g)E(q_{t-1}) + Cov(a_t^g, q_{t-1}) \tag{4}$$

This identity shows that the added value generated by a fund manager equals the product of the fund's average gross alpha and its average size, plus the covariance between gross alpha and fund size. In the presence of dis-economies of scale (negative relation between performance and scale), this covariance term should be negative, reflecting the value destroyed due to constrains imposed on the value a manager can extract from the capital markets, due to increasing scale diminishes returns. Conversely, if scale has no effect on returns, the covariance term should be zero and value creation is proportional to the size of the fund. These two implications form the core intuition behind my empirical strategy.

To investigate whether dis-economies of scale are present in the mutual fund industry, we adopted a bootstrapping methodology inspired by the luck vs skill approach of Fama and French 2010. My approach compares the empirical distribution of added value across the entire cross-section of mutual funds with the distribution from a bootstrapped universe in which added value is not constrained by size. This comparison allows me to infer about evidence of dis-economies of scale, account for potential heterogeneity of a size effect across funds, and more importantly to account for a regression-to-the-mean bias. For instance, one could alternatively estimate a panel fixed-effects model, regressing added value on fund size and its square. However, such an approach assumes a homogeneous scale effect and is prone to regression-to-the-mean bias. A fund that experiences one or two years of abnormally high (but luck-driven) performance will likely attract inflows, increasing its size. If added value then reverts to the mean, the model would falsely attribute this decline to the effect of scale, rather than to chance. My simulation-based approach account for such a "lucky" relation by comparing the actual distribution with a simulated distribution constructed under the null of no scale constrains on added value.

The remainder of this section is structured as follows. The first subsection describes the methodology employed to measure added value. The second subsection outlines the procedure for generating the bootstrapped cross-sectional distribution of added value under the null hypothesis of no scale constrains. The third subsection addresses undersampling issues that can arise in bootstrap simulations of the Fama and French 2010 procedure, as highlighted by Harvey and Liu 2022. Finally, The fourth subsection describes the statistical tests used to compare the empirical and simulated distributions, which provides the basis for inference about decreasing returns to scale.

3.1 Estimation of the Actual Added Value Distribution

Added value is the average product of the gross alpha of a fund with its lag size ($V = E(a_t^g q_{t-1})$), thus to empirically estimate average added value, first we must estimate the gross alpha of a fund.

To estimate the alpha for each fund, we follow the suggestions of Berk and Van Binsbergen 2015, Pástor, Stambaugh, and Taylor 2015 and Zhu 2018 and benchmark each fund with a tradable market index that corresponds to its fund style. An index reflects better an alternative passive investment to investors that is tradable and marketable in the full length of our sample, where factors from factor models (e.g the four factor model of Carhart 1997) reflect trading strategies that are not marketable in my full sample. For example most factors are discovered in the late 90s, and usually are available to investors in a ETF form much later.

The empirical estimation of average added value ($V = E(a_t^g q_{t-1})$) necessitates a preliminary estimation of each fund's gross alpha (a_t^g). Consistent with the suggestions of Berk and Van Binsbergen 2015, Pástor, Stambaugh, and Taylor 2015 and Zhu 2018, we benchmark each fund against a tradable market index commensurate with its investment style. This approach is preferred given that market indexes represent a fully tradable and marketable passive investment alternative throughout the entirety of my sample period. Conversely, factors derived from traditional factor models (e.g., Carhart's (1997) four-factor model) often reflect trading strategies that lacked marketability across my full sample, with widespread investor access via ETF structures typically occurring significantly later than their initial discovery in the late 1990s

To estimate the gross alpha of the fund we estimate with OLS the following regression:

$$r_t^g - r_t^f = a^g + \beta(r_t^B - r_t^f) + \epsilon_t \tag{5}$$

where r_t^g is the return of the fund before expenses and r_t^B the return of the benchmark index, consistent with the style of the fund.

The gross alpha of a fund at time t is given by the following equation:

$$a_t^g = r_t^g - \hat{\beta} r_t^B \tag{6}$$

where $\hat{\beta}$ is the estimation of the beta in the regression (5) and the added value of a fund at time t is given by:

$$V_t = a_t^g q_{t-1} \tag{7}$$

The average added value of fund is estimated by $\overline{V} = \sum_{t=1}^{T} \frac{V_t}{T}$ and the associated t-statistic by:

$$t(V) = \overline{V} / \hat{\sigma}(\overline{V}) \tag{8}$$

$$\hat{\sigma}(\overline{V}) = \sqrt{\sum_{t=1}^{T} \frac{(V_t - \overline{V})^2}{T(T-1)}} \tag{9}$$

where T is the sample life of the fund.[4]

we estimate the added value for each mutual fund in my sample to construct the actual distribution. We focus on the t-statistic of the average added value, rather than the mean itself. This choice allows me to account for differences in estimation precision across funds. Furthermore the distribution is presented

[4]This t-statistic does not adjust for potential autocorrelation in the time series of $(V_{it})_t$. In unreported analysis, we find that the time series of value added for the average fund exhibits negligible autocorrelation. Although, in the robustness section we re-estimate the t-statistics with robust heteroskedasticity and autocorrelation standard errors.

through percentiles. We construct percentiles in two ways. First, by pooling the entire cross-section of added value observations across all funds, we obtain what we refer to as the ex ante distribution—this represents the distribution from which added value is drawn. Second, we estimate percentiles by weighting each fund according to the number of months it appears in the sample, which gives rise to the ex post distribution—the distribution that reflects the value added of surviving funds.

3.2 Estimation of the Simulated Added Value Distribution

My bootstrapping methodology is inspired by Fama and French 2010, who attempt to disentangle skill from luck in mutual fund performance. They do so by comparing the actual empirical distribution of mutual fund alphas with a distribution resulting from bootstrapping simulations under the null hypothesis that all funds have zero alpha. A key innovation in their methodology is the resampling of time indexes across the panel, which preserves the cross-sectional correlation structure of mutual fund returns—unlike standard fund-by-fund resampling approaches (e.g., Kosowski et al. 2006).

Building on and extending Fama and French 2010's time-index resampling strategy, we adapt it to test for diseconomies of scale in the mutual fund industry. Specifically, we compare the empirical distribution of mutual fund added value with a distribution that is the result of bootstrapping simulations under the null that size does not impose constraints on the value a fund manager can extract from the capital markets.

The simulated distribution is the result of 1,000 bootstrapping simulations from the mutual fund data. Compared to the Fama and French 2010 setup, my extension faces two new key challenges. First, we must eliminate the potential correlation between lagged fund size and returns while preserving the effect of returns on fund size to remain consistent with the real-world dynamics of mutual fund growth. Second, we need to maintain the cross-sectional correlation among fund sizes without distorting the strong persistence observed in the time-series behavior of fund size.[5]

Resampling the same set of time indices for both returns and fund sizes preserves several important features: the cross-sectional correlation among fund sizes, the cross-sectional correlation among returns, and the contemporaneous correlation between returns and sizes. Additionally, this procedure sets the correlation between lagged fund sizes and returns to zero, which is essential to construct the added value distribution without scale constraints. However, it does not preserve the strong persistence typically observed in the time-series dynamics of fund sizes.

To maintain the desirable properties of resampling the same set of time indices for both returns and fund sizes, we replace fund sizes with fund growth rates in the resampling process. Fund growth rates does not posses strong persistent (autocorrelation) structure. Then reconstruct the size series recursively using the resampled growth rates and the initial size of each fund. This approach resamples with keeping the empirical

[5]Evidence on the persistent structure of fund size is provided in the appendix to this paper.

properties of the data intact and maintain the strong persistent of the size time series process.

In more details the fund growth rates are defined as:

$$fgrow_t = \frac{q_t}{q_{t-1}} - 1 \tag{10}$$

and given a simulated sequence of fund growth rates $fgrow_1^b, fgrow_2^b, \ldots$, the simulated fund size is generated recursively as:

$$q_0^b = q_0 \tag{11}$$

$$q_t^b = q_{t-1}^b(1 + fgrow_t^b), \quad \forall t \geq 1 \tag{12}$$

where q_0 is the starting fund size from the original sample.

Resampling returns and fund growth rates preserves the desirable sample properties discussed above, but it implicitly assumes that both series follow an independent and identically distributed (i.i.d.) time-series structure. In the appendix section of this paper, we examine potential departures from this assumption by analyzing autocorrelation in the time series of returns and fund growth rates. While the median fund does not exhibit significant autocorrelation in either returns or fund growth, a closer examination of the distribution of first-order autocorrelation coefficients reveals some weak deviations. In particular, a subset of funds displays weak positive autocorrelation in fund growth rates. For example, the 75th percentile of the distribution of the first-order autocorrelation coefficient for fund growth is 0.17, whereas return series generally show no such pattern.

Since the median fund does not exhibit evidence of autocorrelation in the time series of either returns or fund growth rates, in the baseline simulation procedure, we assume that returns and fund growth rates are independently and identically distributed over time. This assumption is relaxed in the robustness section, where we incorporate weak autocorrelation in growth rates.

To clarify the resampling procedure used to construct the simulated distribution of added value—under the assumption that scale does not impose constraints—I summarize the main steps below:

- For each fund $i \in \{1, 2..., n\}$ (where n is the number of funds in the sample), we estimate the growth rates ($(fgrow_{it})_{t\in\tau}$ where $\tau = \{1, 2, 3, ..., T\}$ is the set of the time indexes of the full sample) with equation (10) and keep the starting size (q_{i0}) of each fund in the sample.

- Estimate the excess returns for each fund i ($(r_{it})_{t\in\tau}$) and the excess benchmark returns ($(r_{B_i t})_{t\in\tau}$).

- For each bootstrap iteration $b \in \{1, 2, ..., N_B\}$, (where N_B the number of bootstrap iterations, set to 1,000) :

1. we re-sample with replacement from the time-index set τ and create the bootstrapped set of indexes τ^b, which leads to the bootstrapped samples of mutual fund excess returns $(r^b_{it})_{t\in\tau^b_i}$, fund growth rates $(fgrow^b_{it})_{t\in\tau^b_i}$ and benchmark index excess returns $(r^b_{B_it})_{t\in\tau^b_i}$.
2. Construct the re-sampling fund size process $(q^b_{it})_{t\in\tau^b_i}$ for each fund we recursively, using equation (12) with the re-sampled growth rate process $(fgrow^b_{it})_{t\in\tau^b_i}$.
3. For each fund i, we estimate the regression (5) with the simulated excess returns, to estimate the gross alpha of each fund $a^b_{it} = r^b_{it} - \hat{\beta}^b r^b_{B_it}$ for $t \in \tau^b_i$.
4. Estimate added value for each fund i, $V^b_t = a^b_{it}q^b_{it}$ and the relative t-statistic $t^b_i(V)$
5. we drop t-statistics from $(t^b_i)_{i\in\{1,2...,n\}}$ that are fall outside fund-specific bandwidths, as suggested by Harvey and Liu 2022. The details about the this step, are discussed in a later subsection.

3.3 Comparison Tests

I compare the distribution of actual added value with a simulated distribution constructed under the assumption that size does not impose constrains to the value a fund manager can extract from the capital markets. As shown in equation (4), if size impose constrains on added value, the percentiles of the simulated added value distribution should systematically exceed the corresponding percentiles of the actual distribution.

To statistically test for the presence of a size effect, we employ two distinct approaches: a percentile-based comparison and formal distributional tests. In the percentile comparison approach, we first directly compare the percentiles of the actual and simulated distributions and then estimate the likelihood that the simulated percentile exceeds the corresponding actual percentile. To further assess differences between the two distributions, we apply two complementary tests: the Anderson-Darling (AD) test for equality of distributions, and the first-order stochastic dominance (SD1) test proposed by Barrett and Donald 2003. In what follows, we provide more details on both approaches.

I start with the percentile comparison. Following the procedure of the previous subsections, we obtain $1,000$ (N_B) simulated cross-sections of added value t-statistics, all constructed under the null hypothesis that scale does not constrain added value. For each of those $1,000$ simulated cross-section, we estimate ex ante and ex post percentiles, those percentile are average away in the $1,000$ simulations to construct the ex ante and ex post distribution. We estimate the ex ante and ex post actual distribution and simple compare the simulated and actual percentiles.

For a given percentile p, we estimate the likelihood that the simulated percentile will exceed the actual percentile, by computing the proportion of simulations in which the simulated percentile exceeds the corresponding actual percentile of the added value t-statistics. The statistic is defined as:

$$S(p) = \frac{1}{N_B} \sum_{b=1}^{N_B} 1_{\{t_p^b > t_p^{Act}\}} \tag{13}$$

where t_p^b is the p-th percentile of the simulated distribution of the added value t-statistics in bootstrap iteration b, and t_p^{Act} is the corresponding percentile from the actual distribution. A high value of $S(p)$ (> 0.90), suggests that actual added value realizations are systematically lower than what we would expect in the absence of a size effect—providing evidence in favor of a lager size impose constrains on the value a fund manager can extract from the capital markets, due to a larger size diminishing returns. In contrast, values of $S(p)$ near 0.5 suggest no difference between the actual and simulated distributions, implying size does not constrain added value and size does no effect returns. Low values of $S(p)$ (below 0.10) suggest that an increase in size has a positive effect on add value and returns.

To further investigate potential size effects, we employ two robust distributional tests: the Anderson-Darling (AD) test for equality of distributions and the first-order stochastic dominance (BD SD1) test proposed by Barrett and Donald (2003). For both tests, it's crucial to prepare the data appropriately. We construct two independent samples, each containing 10,000 observations:

- Actual Distribution Sample: This is created by resampling with replacement directly from the original observed added value t-statistics.
- Simulated Distribution Sample: This sample is formed by first combining all 1,000 simulated cross-sections of added value t-statistics generated during the percentile comparison phase. From this combined dataset, we then resample with replacement to obtain the 10,000 observations.

This resampling approach ensures that the AD and BD SD1 tests can be validly applied to compare the actual and simulated distributions. It's important to note that we construct both ex ante and ex post distribution. The ex post distribution is constructed by resample with weights the relative sample life of each observation. These tests will provide a complementary perspective to the percentile comparisons, offering a more holistic view of whether scale indeed imposes constraints on a fund manager's ability to extract value from the market.

The Anderson-Darling test the null hypothesis of equality between the simulated and actual distribution ($F_{Sim} = F_{Act}$). This test compares the two ECDFs by evaluating a weighted sum of squared differences between them across the joint support. The weights are inversely related to the variance of the joint ECDF at each point, giving more emphasis to differences in the tails. The test statistic is formally defined as:

$$AD = \frac{N_{Act} N_{Sim}}{(N_{Act} + N_{Sim})^2} \sum_{x \in Z} \left(\frac{\mid \hat{F}_{Sim}(x) - \hat{F}_{Act}(x) \mid}{\sqrt{\hat{G}(x)(1 - \hat{G}(x))}} \right) \tag{14}$$

where N_{Sim} and N_{Act} denote the number of observations in the simulated and actual samples respectively, Z is the set of all unique values in the combined sample, and $\hat{G}$ is the ECDF of the pooled sample.

To compute the p-value of the Anderson-Darling test, we follow the standard re-sample approach: we randomly split the pooled sample into two groups of the same sizes as the original samples and compute the AD statistic across many such permutations. The empirical p-value is then given by the proportion of resampled statistics that exceed the original AD statistic.

While rejection of the null hypothesis in the Anderson-Darling test provides evidence of a statistically significant difference between the actual and simulated distributions, it does not reveal the direction of the difference. To address this, we adopt the first-order stochastic dominance test proposed by Barrett and Donald 2003. This test allows me to assess whether the simulated distribution F_{Sim} consistently lies under the actual distribution F_{Act}. Which in turn suggest. Formally, the null hypothesis of the test is that F_{Sim} first-order stochastically dominates F_{Act}, meaning that for all x, $F_{Sim}(x) \leq F_{Act}(x)$. In the context of this analysis, this corresponds to evidence that size impose constrains on value creation.

The Barrett and Donald 2003 is Kolmogorov-Smirnov based that is conducted with the following statistic:

$$\hat{S}_1 = (\frac{N_{Act}N_{Sim}}{N_{Act} + N_{Sim}})^{\frac{1}{2}} \sup_{\{x \in Z\}} (\hat{F}_{Sim}(x) - \hat{F}_{Act}(x)) \tag{15}$$

where N_{Sim} and N_{Act} denote the number of observations in the simulated and actual samples, respectively, and Z is the set of all unique values in the joint sample. The statistic captures the largest vertical distance between the two empirical cumulative distribution functions, in the direction consistent with first-order stochastic dominance.

To assess statistical significance, p-values are computed using a resampling procedure analogous to that used in the Anderson-Darling test—by drawing repeated samples from the combined distribution.

Together, these two approaches—the percentile-based analysis and the distributional tests—offer complementary insights. The percentile method highlights localized deviations at specific points in the distribution, while the distributional tests assess broader, global differences in shape. Their combination enables a robust and nuanced evaluation of whether added value is constrained by size.

3.4 Issues with the Bootstrapping Simulations

The time-index bootstrapping strategy we adopt preserves the cross-sectional correlation between fund returns and fund sizes. However, it introduces certain issues that must be addressed. In particular, many funds in the sample have missing observations, primarily because they are not active throughout the entire sample period. As a result, the number of observations in a fund's bootstrapped sample may differ from the number of observations in the actual data, which may lead to a difference in the distribution of t-statistics.

Although the oversampling of some funds should, in aggregate, roughly offset the undersampling of others, the impact on individual t-statistic distributions—and consequently on the cross-sectional distribution of t-statistics—can differ substantially between the two cases.[6] This issue is particularly important because, as the degrees of freedom increase, the Student's t-distribution converges to the standard normal distribution. Therefore, the distortion caused by differences in sample size is negligible for funds with a large number of observations but can be significant for those with relatively few. For example, for a fund with 24 observations, oversampling to 36 has a minimal effect on the shape of the t-distribution, whereas undersampling to 12 produces a distribution with fatter tails.

Harvey and Liu 2022 examine these implications in the original Fama and French 2010 framework and propose remedies. They find that time-index bootstrapping can introduce bias in the tails of the simulated distribution, potentially leading to a failure to reject the null hypothesis of zero alpha. In contrast, in my context, the problem may work in the opposite direction: undersampling could produce fatter tails in the simulated distribution, and since we am rejecting the null of no scale constrains, when the simulated t-statistics are larger than the actual, this potential may lead to over-rejecting the null hypothesis. To address this issue, we follow the corrective procedures suggested by Harvey and Liu 2022.

Firstly, we required a minimum of 12 return observations for a fund to be included in the analysis. More importantly, within each bootstrap iteration-where re-sampled return histories are generated for each fund (with the time index strategy), we retain only those funds that have at least 12 unique return observations in that iteration. [7]

Secondly, we extend and adapt the thresholding approach of Harvey and Liu 2022 to specifically address the undersampling distortions introduced by the time-index bootstrapping procedure. Before applying the time-index strategy, we perform a standard per-fund bootstrapping simulation to generate realistic bandwidths for the distribution of t-statistics.

For each fund i, we focus only on months where returns are observed and perform 1000 bootstrap resamples in line with the methodology outlined in subsection 1.3.2, computing the added-value t-statistic in each iteration. Let $\hat{q}_{25,i}$ and $\hat{q}_{75,i}$ denote the 25th and 75th percentiles of the resulting bootstrapped t-statistic distribution for fund i. We define the corresponding bandwidth for fund we as:

$$\hat{band}(i) = (\hat{q}_{25,i} - thres \times [\hat{q}_{75,i} - \hat{q}_{25,i}], \hat{q}_{75,i} + thres \times [\hat{q}_{75,i} - \hat{q}_{25,i}]) \tag{16}$$

where *thres* is a threshold parameter set to 2, following the simulation-based guidance in Harvey and Liu 2022.

Given these bandwidths across the cross-section of funds, we proceed with the time-index bootstrapping. In each bootstrapping iteration, we compare the simulated added-value t-statistic of each fund to its corre-

[6] Here, "oversampling" refers to bootstrap samples that contain more observations than the original data for a given fund, while "undersampling" refers to cases where the bootstrap sample includes fewer observations than in the original data.

[7] I also test how robust my results are to the minimum number of observations requirement.

sponding bandwidth. If a bootstrapped t-statistic falls outside this range, we discard it from the simulated cross section of added value t-statistics (step 5 in the resampling process discussed in section 1.3.2). This procedure ensures that extreme values—potentially driven by undersampling—do not distort the shape of the simulated distribution, particularly its tails.

It is worth noting that the ex post distribution can further mitigate issues related to undersampling, as it assigns smaller weights to funds with short return histories—which are the primary source of the problem.

4 Data

My mutual fund data is sourced from LSEG Data & Analytics (formerly Refinitiv). We restrict the sample period to begin in January 1993 and end in December 2022. This restriction is due to the fact that total net assets were only reported quarterly prior to 1993, whereas my analysis requires monthly observations to accurately align fund returns with size. Consistent monthly reporting of total net assets is therefore essential for the validity of the scale-related tests.

The dataset includes both live and dead mutual funds, ensuring that survivorship bias is minimized. We focus exclusively on actively managed domestic equity mutual funds in the U.S. market. To that end, we include only funds classified under the following Lipper categories: Large-Cap Core (LCC), Large-Cap Growth (LCG), Large-Cap Value (LCV), Mid-Cap Core (MCC), Mid-Cap Growth (MCG), Mid-Cap Value (MCV), Multi-Cap Core (MuCC), Multi-Cap Growth (MuCG), Multi-Cap Value (MuCV), Small-Cap Core (SCC), Small-Cap Growth (SCG), and Small-Cap Value (SCV)

Most mutual funds offer multiple share classes, and mutual fund databases such as LSEG report data at the share class level. While these share classes may differ in terms of fee structures and investor clientele, they represent claims on the same underlying portfolio. Following standard practice in the literature, we aggregate share classes at the fund level to avoid double-counting and to ensure that performance metrics reflect portfolio-level characteristics. This aggregation results in a sample of 2,331 distinct mutual funds.[8]

In the aggregation process, we compute three key fund-level variables: total net assets (fund size), net return, and total expense ratio (fees charged to investors). A fund's total net assets is calculated by summing the total net assets across all its share classes. The fund's net return is computed as a weighted average of the net returns of its constituent share classes, where the weights are given by each share class's beginning-of-period total net assets, scaled by the total across all share classes. The fund's total expense ratio is constructed analogously, as the weighted average of share-class-level expense ratios using the same weighting scheme.

To measure relative performance (alphas) for each fund, we follow the suggestions of Pástor, Stambaugh,

[8]The construction of the added value measure requires information on returns, total net assets, and total expense ratios. We exclude funds that lack complete time-series data for any of these variables.

and Taylor 2015 and Zhu 2018 and benchmark each fund with a tradable market index that corresponds to its Lipper category (fund style). Table 1 presents the fund styles included in the sample with the corresponding assigned benchmarks.

Table 1: Stock market indexes used as benchmarks, to adjust for fund style risk.

Fund Style	Assigned Benchmark
Large Cap Core	Russell 1000 TR USD
Large Cap Growth	Russell 1000 Growth TR
Large Cap Value	Russell 1000 Value TR
Mid Cap Core	S&P Mid Cap 400 TR
Mid Cap Growth	Russell MidCap Growth TR
Mid Cap Value	Russell MidCap Value TR
Multi Cap Core	Russell 3000 TR
Multi Cap Growth	Russell 3000 Growth TR
Multi Cap Value	Russell 3000 Value TR
Small Cap Core	Russell 2000 (EOD) TR
Small Cap Growth	Russell 2000 Growth TR
Small Cap Value	Russell 2000 Value TR

To ensure that fund size measurements are comparable over time, we adjust all total net asset values for inflation and express them in December 31, 2022 dollars. Inflation adjustments are based on data obtained from the Federal Reserve's website.

Having constructed the key variables, we now turn to a descriptive analysis of the sample. Table 2 presents summary statistics for the main fund-level variables. A fund is included in the calculations only if it has at least one year of gross return data.

The distributions of net and gross returns are left-skewed, with average monthly net returns of 0.77% and gross returns of 0.86%. The monthly expense ratio is calculated as one-twelfth of the annual gross expense ratio reported in the LSEG database, yielding an average of 10 basis points per month. Fund size data are available starting from December 31, 1992. The distribution of fund size is highly right-skewed, with a median value of $304.87 million and a 75th percentile of $1,332.11 million. Finally, we define fund sample age as the number of years since a fund's first gross return observation, measured at each month-end. On average, funds in the sample have 16 years of available data.

Table 2: Summary statistics.

FundVariable	N	Pct5	Pct25	Pct50	Mean	Pct75	Pct95	sd
Net r (%)	464488	-8.38	-1.92	1.19	0.77	3.79	8.36	6.12
Gross r (%)	448297	-8.29	-1.83	1.28	0.86	3.89	8.45	5.63
Expence ratio (%)	464132	0.04	0.07	0.09	0.1	0.11	0.17	0.04
Fund Size (mil $)	492206	4.27	64.06	304.87	2193.72	1332.11	8647.96	8772.58
Fund Sample Age	445125	2.42	8.25	16.25	16.03	23.33	30	8.87

5 Results

5.1 First Impressions

I estimate the ex ante and ex post empirical cumulative distributions of the t-statistic of added value ($t(V)$) for each fund in the sample, including only those funds with at least 12 return observations. In addition, we construct a separate set of 1,000 simulated cross-sections of $t(V)$ under the assumption of no size constrains on added value, as described in the methodology section and compute the average ex ante and ex post simulated empirical cumulative distributions of $t(V)$.

Figure 1 compares the ex ante distributions of the simulated and actual distributions of the t-statistic of added value ($t(V)$) and Figure 2 the ex post distribution. In both figures, from approximately the 15th percentile onward, the simulated distributions lie below the corresponding actual distributions. This pattern suggests that, for the majority of funds, the value a manager can add is constrained by size, as increasing size diminishes returns. To the left of the 15th percentile, the simulated distribution is almost identical to the actual distribution, suggesting that unskilled managers that destroy value are not constrain by size and destroy value simple because they do not have skill.

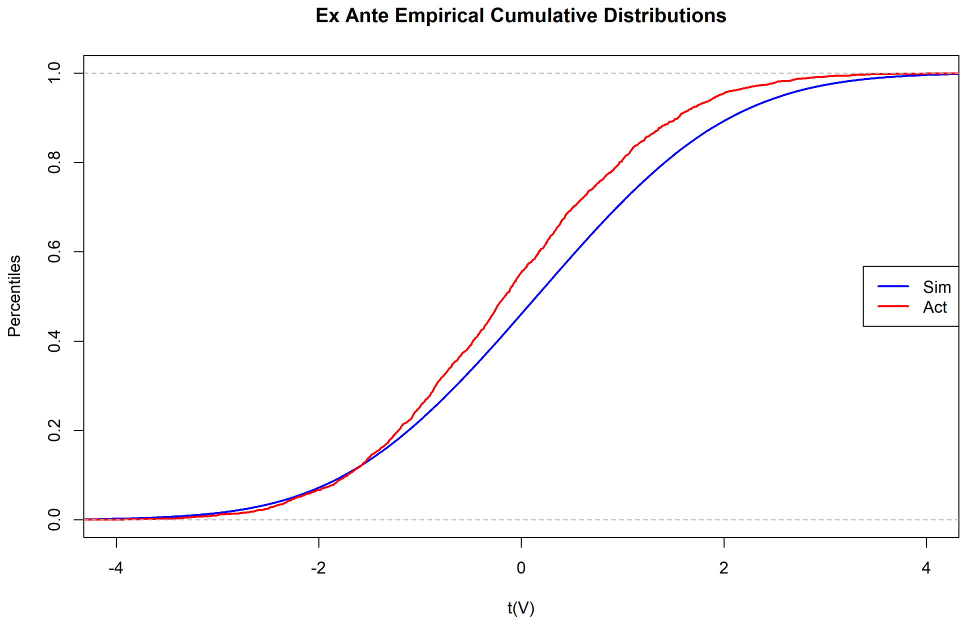

Figure 1: Simulated and Actual Ex Ante Empirical Cumulative Distributions of $t(V)$, 1993-2022.

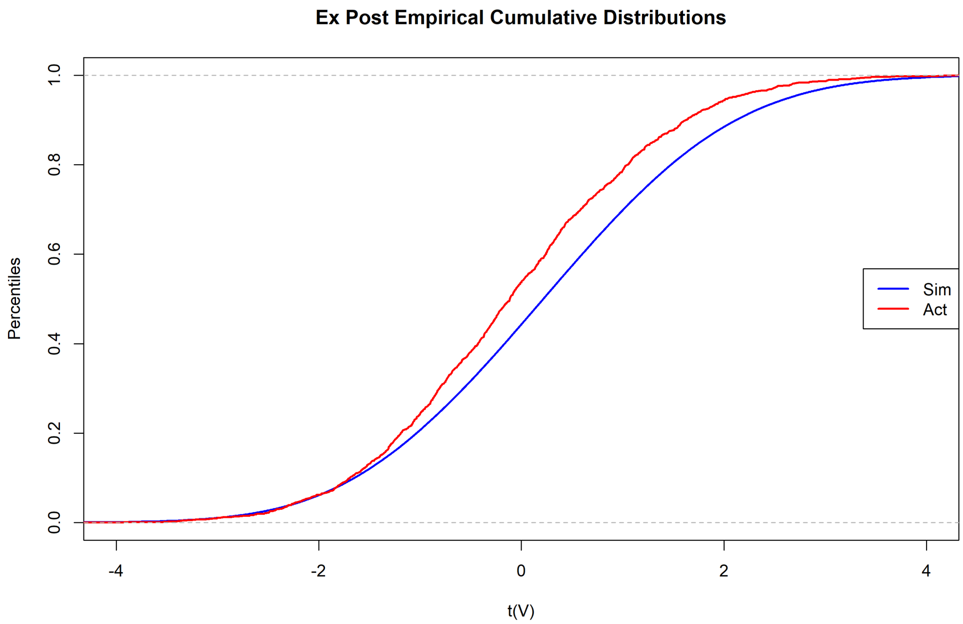

Figure 2: Simulated and Actual Ex Ante Empirical Cumulative Distributions of $t(V)$, 1993-2022.

5.2 Likelihoods of Percentiles

In this section, we focus on specific percentiles of the $t(V)$ distribution to draw more targeted inferences regarding the effect of size on the value a manager can add. Specifically, Table 3 reports, for each selected percentile, the fraction of the 1,000 simulation (likelihoods) that generate $t(V)$ values at specific percentiles greater than the corresponding value observed in actual fund data, along with the simulated and actual $t(V)$. These percentile-based comparisons offer a more granular view of how the simulated and actual distributions diverge, helping to identify specific regions where size effects are most pronounced.

Table 3 shows that the median $t(V)$ values from the simulated distributions exceed those of the actual distribution in the vast majority of simulations, with 98% of simulations in the ex ante distribution and 96% in the ex post distribution. These results provide strong evidence that for the median fund, size constrains the value a manager can create, due to increasing size diminishing returns.

Examining the tails of the distribution reveals that the constrains imposed by size, are more pronounced in the right tail of the $t(V)$ distribution—where funds that add value lie—compare to the left tail where funds that destroy value lie. For instance, at the 80th percentile, Table shows that in 100% of simulations for the ex ante case and 98% of simulations for the ex post case, the simulated $t(V)$ exceeds the corresponding actual value. In contrast, at the 20th percentile, the simulated $t(V)$ is higher than the actual in only 73% (ex ante) and 79% (ex post) of simulations. Therefore the scale effect is especially detrimental to high added value fund, while its influence on negative added value funds is weaker and less consistent. This implies a correlation between skill and the strength of the size effect.[9]

An examination of the extreme left tail of the $t(V)$ distribution provides weak evidence that for a small subset of unskilled funds size effect positive return. At the 2nd percentile in Table, only 12% of the simulated $t(V)$ values exceed the actual $t(V)$ for the ex ante distribution, and 24% for the ex post distribution. While these results are not conclusive, they suggest that, for a very small fraction of poorly performing funds (unskilled funds), larger size may be associated with improved performance.

The standard errors used in the estimation of the t-statistics presented in Table 3 do not account for potential autocorrelation and heteroskedasticity in added value. In the appendix, we address this issue by re-estimating the standard errors using the Newey and West 1987 heteroskedasticity and autocorrelation consistent (HAC) estimator with three lags. The results remain virtually unchanged, indicating that the original findings are robust to these concerns.

[9]This finding is consistent with the results of Barras, Gagliardini, and Scaillet 2022, that also find a correlation between skill and scale.

Table 3: Actual and Simulated Added Value Percentiles, Simulated Percentiles are Constructed Under the Null that size does not impose constrains on added value. The table shows values of t-statistic of added value ($t(V)$) at selected percentiles (Pct) for actual (Act) and simulated (Sim) data. The table also shows the ratio of the 1,000 simulations that produce larger values of $t(V)$ at the selected percentiles that those observed in actual fund data ($\frac{\#Sim>Act}{N_B}$). The period is from January 1993 to December 2022, is required at least 12 return observations for a fund to be included and the results are presented for the ex ante and ex post distributions.

	Ex ante distribution			Ex post distribution		
Pct	Sim	Act	$\frac{\#Sim>Act}{N_B}$	Sim	Act	$\frac{\#Sim>Act}{N_B}$
1	-3.23	-3.01	0.12	-3.02	-3.01	0.47
2	-2.84	-2.63	0.12	-2.67	-2.54	0.24
5	-2.26	-2.22	0.41	-2.13	-2.16	0.56
10	-1.75	-1.72	0.43	-1.65	-1.69	0.6
20	-1.12	-1.22	0.73	-1.04	-1.19	0.79
30	-0.65	-0.85	0.90	-0.57	-0.81	0.9
40	-0.24	-0.48	0.93	-0.17	-0.43	0.92
45	-0.04	-0.31	0.96	0.02	-0.28	0.95
47	0.03	-0.26	0.97	0.1	-0.22	0.96
48	0.07	-0.24	0.97	0.14	-0.19	0.96
50	0.15	-0.17	0.98	0.22	-0.11	0.96
52	0.22	-0.11	0.98	0.29	-0.06	0.97
53	0.26	-0.08	0.98	0.33	-0.03	0.98
55	0.34	-0.01	0.99	0.41	0.04	0.98
60	0.53	0.17	0.99	0.6	0.24	0.98
70	0.94	0.51	1.00	1.01	0.58	0.99
80	1.41	0.97	1.00	1.48	1.04	0.99
90	2.05	1.54	1.00	2.12	1.62	0.99
95	2.58	1.94	1.00	2.65	2.05	1.00
98	3.16	2.51	1.00	3.24	2.67	0.99
99	3.54	2.84	1.00	3.63	3.13	0.96

5.3 Distribution Tests

This section presents formal distributional tests to compare the empirical and simulated distributions of the t-statistic of added value, $t(V)$. The simulated distributions are constructed as described in the previous subsection.

Table 4 reports results from two tests: the Anderson-Darling two-sided (AD two sided) that tests for equality of distributions and the first-order stochastic dominance test proposed by Barrett and Donald 2003 (hereafter BD SD1). These tests are applied to compare the actual distribution of $t(V)$ with distributions simulated under different null hypotheses regarding the size effect.

The Anderson-Darling test strongly rejects the null of distributional equivalence, with p-values below 0.1%, providing strong statistical evidence that size effect value creation and return performance, in the mutual fund industry. The Barrett and Donald 2003 test fails to reject the null hypothesis that the simulated distribution stochastically dominates the empirical distribution, with p-values at 8% and 45% for the ex ante and ex post distributions, respectively, size impose constrains on the value a fund manager can add, due to increasing size diminishing performance.

Table 4: This table presents the p-values of the Anderson-Darling two-sided (AD two sided) and the first order stochastic dominance of Barrett and Donald 2003 (BD SD1 test) for comparisons of the actual and the simulated distribution of added value.

Ex ante distributions			Ex post distributions		
Tests	H_0	p-value	Tests	H_0	p-value
AD two sided	$F_{\text{Act}} = F_{\text{Sim}}$	0.00	AD two sided	$F_{\text{Act}} = F_{\text{Sim}}$	0.00
BD SD1 test	$F_{\text{Act}} > F_{\text{Sim}}$	0.08	BD SD1 test	$F_{\text{Act}} > F_{\text{Sim}}$	0.446

6 Robustness

In this section, we provide some robustness tests. First, we examine the effect of varying the selection criterion, of a fund to be included in the analysis based on the minimum number of available return observations. The baseline requirement is 12 monthly observations, which we make progressively stricter, increasing it up to 84 months. Next, we explore two alternative methods for simulating fund size. In the first approach, we assume that the time series of size growth rates follows an AR(1) process. In the second, we simulate size using log-growth rates instead of simple growth rates.

6.1 Selection Criterion: Minimum Number of Return Observations

I vary the selection criterion for the minimum number of return observations a fund must have to be included in the analysis, using thresholds of 12, 24, 36, 48, 60, and 84 months. Table 5 reports the likelihood that the simulated t-statistics exceed the actual t-statistics at selected percentiles across 1,000 simulations, under each threshold, based on the ex ante distribution. Table 6 presents analogous results for the ex post distribution.

Overall, the results in Tables 5 and 6 remain qualitatively invariant. For most funds, the value a manager can add is constrained by size, as increasing size leads to diminishing returns. The likelihoods at the mid and upper percentiles remain largely stable across the different selection criteria. However, at the extreme lower percentiles, the likelihoods increase as the selection criterion becomes stricter, effectively eliminating the weak positive effects observed in those percentiles under more lenient inclusion thresholds.

Table 5: The table shows the percent of the 1,000 simulations that produce larger values of $t(V)$ at the selected percentiles that those observed in actual fund data ($\frac{\#Sim>Act}{N_B}$), for various selection criterion on the number of return observations required to for a fund to be included in the analysis. The period is from January 1993 to December 2022.

	Ex Ante Distributions					
	Minimum Number of Monthly Return Observations					
	12 obs	24 obs	36 obs	48 obs	60 obs	84 obs
Pct	$\frac{\#Sim>Act}{N_B}$	$\frac{\#Sim>Act}{N_B}$	$\frac{\#Sim>Act}{N_B}$	$\frac{\#Sim>Act}{N_B}$	$\frac{\#Sim>Act}{N_B}$	$\frac{\#Sim>Act}{N_B}$
1	0.12	0.13	0.19	0.29	0.35	0.4
2	0.12	0.12	0.18	0.23	0.28	0.35
5	0.41	0.44	0.51	0.5	0.52	0.56
10	0.43	0.44	0.48	0.49	0.52	0.56
20	0.73	0.74	0.76	0.77	0.77	0.78
30	0.9	0.89	0.9	0.91	0.9	0.9
40	0.93	0.94	0.94	0.94	0.94	0.94
45	0.96	0.96	0.96	0.96	0.96	0.96
47	0.97	0.97	0.97	0.97	0.97	0.97
48	0.97	0.97	0.98	0.98	0.98	0.98
50	0.98	0.98	0.98	0.98	0.98	0.98
52	0.98	0.98	0.98	0.98	0.98	0.98
53	0.98	0.99	0.98	0.98	0.98	0.98
55	0.99	0.99	0.99	0.99	0.99	0.99
60	0.99	0.99	0.99	0.99	0.99	0.98
70	1.00	1.00	1.00	1.00	0.99	0.99
80	1.00	0.99	0.99	0.99	0.99	0.99
90	1.00	1.00	1.00	1.00	0.99	1.00
95	1.00	1.00	1.00	1.00	1.00	1.00
98	1.00	1.00	1.00	1.00	1.00	1.00
99	1.00	1.00	1.00	0.99	0.99	0.99

Table 6: The table shows the percent of the 1,000 simulations that produce larger values of $t(V)$ at the selected percentiles that those observed in actual fund data ($\frac{\#Sim>Act}{N_B}$), for various selection criterion on the number of return observations required to for a fund to be included in the analysis. The period is from January 1993 to December 2022.

	Ex Post Distributions					
	Minimum Number of Monthly Return Observations					
	12 obs	24 obs	36 obs	48 obs	60 obs	84 obs
Pct	$\frac{\#Sim>Act}{N_B}$	$\frac{\#Sim>Act}{N_B}$	$\frac{\#Sim>Act}{N_B}$	$\frac{\#Sim>Act}{N_B}$	$\frac{\#Sim>Act}{N_B}$	$\frac{\#Sim>Act}{N_B}$
1	0.47	0.47	0.49	0.51	0.53	0.53
2	0.24	0.25	0.26	0.28	0.29	0.3
5	0.56	0.56	0.57	0.58	0.58	0.59
10	0.6	0.6	0.61	0.61	0.62	0.63
20	0.79	0.79	0.79	0.79	0.8	0.79
30	0.9	0.9	0.9	0.9	0.9	0.9
40	0.92	0.92	0.93	0.93	0.93	0.92
45	0.95	0.95	0.95	0.95	0.95	0.95
47	0.96	0.96	0.96	0.96	0.96	0.96
48	0.96	0.96	0.96	0.96	0.96	0.96
50	0.96	0.96	0.96	0.96	0.96	0.96
52	0.97	0.97	0.98	0.98	0.97	0.97
53	0.98	0.98	0.98	0.98	0.98	0.97
55	0.98	0.98	0.98	0.98	0.98	0.97
60	0.98	0.98	0.98	0.98	0.98	0.98
70	0.99	0.99	0.99	0.99	0.99	0.99
80	0.99	0.99	0.99	0.99	0.99	0.99
90	0.99	0.99	0.99	0.99	0.99	0.99
95	1.00	1.00	1.00	1.00	0.99	0.99
98	0.99	0.99	0.99	0.99	0.98	0.99
99	0.96	0.96	0.96	0.96	0.96	0.96

6.2 Autocorrelation in the Fund Growth Ratios

Evidence presented in the appendix indicates weak first-order autocorrelation in the growth ratios of a subset of funds. To accurately reflect this characteristic in my simulations, we have modified the resampling procedure used to construct the simulated distribution. This modification assumes that fund growth ratios adhere to an AR(1) process, thereby changing how the time series of fund sizes is recursively generated. In this subsection, we briefly describe the revised approach and then proceed to discuss the corresponding results.

I start by summarizing the main steps of the revised resampling procedure below:

- For each fund $i \in \{1, 2..., n\}$ (where n is the number of funds in the sample), we estimate the growth rates ($(fgrow_{it})_{t\in\tau}$ where $\tau = \{1, 2, 3, ..., T\}$ is the set of the time indexes of the full sample) with equation (10) and keep the starting size (q_{i0}) of each fund in the sample.

- we estimate the following regression for each fund $i \in \{1, 2..., n\}$

$$fgrow_{i,t} = \phi_{i0} + \phi_{i1} fgrow_{i,t-1} + e_{i,t} \tag{17}$$

  and keep the $\hat{\phi}_{i0}$, $\hat{\phi}_{i1}$, mean fund growth $\overline{frgow_i}$ and the residuals of the AR(1) regressions $(e_{i,t})_t$ for each fund

- Estimate the excess returns for each fund $i \in \{1, 2..., n\}$ $((r_{it})_{t\in\tau})$ and the excess benchmark returns $((r_{B_i t})_{t\in\tau}\ )$.

- For each bootstrap iteration $b \in \{1, 2, ..., N_B\}$, (where N_B the number of bootstrap iterations, set to 1,000) :

  1. we re-sample with replacement from the time-index set τ and create the bootstrapped set of indexes τ^b, which leads to the bootstrapped samples of mutual fund excess returns $(r^b_{it})_{t\in\tau^b_i}$, fund growth rates residuals $(e^b_{it})_{t\in\tau^b_i}$ and benchmark index excess returns $(r^b_{B_i t})_{t\in\tau^b_i}$.
  2. Construct the resample fund growth ratios:

$$fgrow^b_t = \hat{\phi}_{i0} + \hat{\phi}_{i1} fgrow^b_{i,t-1} + e^b_{it} \tag{18}$$

$$fgrow^b_{i,1} = \overline{fgrow_i} \tag{19}$$

  3. Construct the resampled fund size process $(q^b_{it})_{t\in\tau^b_i}$ for each fund we recursively, using equation (12) with the resampled growth rate process $(fgrow^b_{it})_{t\in\tau^b_i}$.
  4. For each fund $i \in \{1, 2..., n\}$, we estimate the regression (5) with the simulated excess returns, to estimate the gross alpha of each fund $a^b_{it} = r^b_{it} - \hat{\beta}^b r^b_{B_i t}$ for $t \in \tau^b_i$.
  5. Estimate added value for each fund i, $V^b_t = a^b_{it} q^b_{it}$ and the relative t-statistic $t^b_i(V)$
  6. we drop t-statistics from $(t^b_i(V))_{i\in\{1,2...,n\}}$ that are fall outside fund-specific bandwidths, similar to the procedure discussed in the subsection 1.3.3.

Table 7 presents the results of the percentile comparisons between the actual distribution of t-statistics of added value and the simulated distribution constructed using the revised procedure. Overall, accounting for the weak autocorrelation in growth ratios yields conclusions that are nearly identical to those from the baseline approach. For most funds, the value a manager can add appears to be constrained by size. The right tail of the added value distribution provides strong evidence of this effect, whereas the effect diminishes toward the left tail. This pattern suggests a positive correlation between managerial skill and the strength of the size constraint. Notably, the weak evidence of a positive size effect in the extreme left tail observed under the baseline approach disappears when accounting for autocorrelation in the revised procedure

Table 7: Actual and Simulated Added Value Percentiles, Simulated Percentiles are Constructed Under the Null that size does not impose constrains on added value. The table shows values of t-statistic of added value ($t(V)$) at selected percentiles (Pct) for actual (Act) and simulated (Sim) data. The table also shows the ratio of the 1,000 simulations that produce larger values of $t(V)$ at the selected percentiles that those observed in actual fund data ($\frac{\#Sim>Act}{N_B}$). The period is from January 1993 to December 2022, is required at least 12 return observations for a fund to be included and the results are presented for the ex ante and ex post distributions.

	Ex Ante distribution			Ex Post distribution		
Pct	Sim	Act	$\frac{\#Sim>Act}{N_B}$	Sim	Act	$\frac{\#Sim>Act}{N_B}$
1	-3.16	-3.01	0.21	-2.97	-3.01	0.57
2	-2.78	-2.63	0.19	-2.62	-2.54	0.32
5	-2.21	-2.22	0.52	-2.1	-2.16	0.62
10	-1.72	-1.72	0.52	-1.63	-1.69	0.64
20	-1.1	-1.22	0.78	-1.03	-1.19	0.8
30	-0.64	-0.85	0.91	-0.58	-0.81	0.89
40	-0.24	-0.48	0.94	-0.18	-0.43	0.92
45	-0.05	-0.31	0.96	0.01	-0.28	0.95
47	0.02	-0.26	0.97	0.08	-0.22	0.96
48	0.06	-0.24	0.98	0.12	-0.19	0.96
50	0.14	-0.17	0.98	0.19	-0.11	0.96
52	0.21	-0.11	0.99	0.27	-0.06	0.97
53	0.25	-0.08	0.99	0.31	-0.03	0.97
55	0.33	-0.01	0.99	0.38	0.04	0.98
60	0.52	0.17	0.99	0.58	0.24	0.97
70	0.92	0.51	1.00	0.98	0.58	0.99
80	1.39	0.97	1.00	1.45	1.04	0.99
90	2.04	1.54	1.00	2.1	1.62	0.99
95	2.57	1.94	1.00	2.64	2.05	1.00
98	3.16	2.51	1.00	3.24	2.67	0.99
99	3.55	2.84	1.00	3.65	3.13	0.97

6.3 Logarithm Fund Growth

As a final robustness test, in the resampling process we use log-growth ratios instead of simple (arithmetic) growth ratios as in baseline approach. The log-growth rates are estimated by the following equation:

$$logfgrowth_t = log(\frac{q_t}{q_{t-1}}) \tag{20}$$

and given a simulated sequence of log-growth rates $logfgrowth^b_1, logfgrowth^b_2, \ldots$, the simulated fund size is generated recursively as:

$$q_0^b = q_0 \tag{21}$$

$$q_t^b = q_{t-1}^b e^{logfgrow_t^b}, \quad \forall t \geq 1 \tag{22}$$

where q_0 is the starting fund size from the original sample.

Table 8 presents the results of the percentile comparisons between the actual distribution of t-statistics of added value and the simulated distribution constructed using log-growth rates. The estimation of t-statistics and likelihoods in Table 8 is identical to that in Table 3, which is based on simple (arithmetic) growth rates. This indicates that the choice between using growth rates or log-growth rates has no material impact on the results.

Table 8: Actual and Simulated Added Value Percentiles, Simulated Percentiles are Constructed Under the Null that size does not impose constrains on added value. The table shows values of t-statistic of added value ($t(V)$) at selected percentiles (Pct) for actual (Act) and simulated (Sim) data. The table also shows the ratio of the 1,000 simulations that produce larger values of $t(V)$ at the selected percentiles that those observed in actual fund data ($\frac{\#Sim>Act}{N_B}$). The period is from January 1993 to December 2022, is required at least 12 return observations for a fund to be included and the results are presented for the ex ante and ex post distributions.

	Ex ante distribution			Ex post distribution		
Pct	Sim	Act	$\frac{\#Sim>Act}{N_B}$	Sim	Act	$\frac{\#Sim>Act}{N_B}$
1	-3.23	-3.01	0.12	-3.01	-3.01	0.48
2	-2.83	-2.63	0.12	-2.66	-2.54	0.25
5	-2.26	-2.22	0.41	-2.13	-2.16	0.56
10	-1.75	-1.72	0.44	-1.65	-1.69	0.6
20	-1.12	-1.22	0.74	-1.04	-1.19	0.79
30	-0.65	-0.85	0.9	-0.57	-0.81	0.9
40	-0.24	-0.48	0.93	-0.17	-0.43	0.93
45	-0.04	-0.31	0.95	0.02	-0.28	0.95
47	0.03	-0.26	0.97	0.1	-0.22	0.96
48	0.07	-0.24	0.97	0.14	-0.19	0.96
50	0.15	-0.17	0.98	0.22	-0.11	0.96
52	0.22	-0.11	0.98	0.29	-0.06	0.98
53	0.26	-0.08	0.98	0.33	-0.03	0.98
55	0.34	-0.01	0.99	0.41	0.04	0.98
60	0.53	0.17	0.99	0.6	0.24	0.98
70	0.94	0.51	1.00	1.01	0.58	0.99
80	1.41	0.97	1.00	1.48	1.04	0.99
90	2.05	1.54	1.00	2.12	1.62	0.99
95	2.58	1.94	1.00	2.65	2.05	1.00
98	3.16	2.51	1.00	3.23	2.67	0.99
99	3.54	2.84	1.00	3.63	3.13	0.96

7 Conclusion

In this paper, we re-examine the role of diseconomies of scale in the active mutual fund management industry. Previous studies that investigate the presence of scale effects typically focus on the relationship between fund size and traditional return-based performance measures, such as alpha. However, this approach often yields inconclusive results, as theoretical models predict that expected alpha is driven to zero in equilibrium, making it inherently difficult to detect a relationship between size and performance. In contrast, this paper investigates diseconomies of scale by examining the effect of size on the value a manager can extract from capital markets. If diseconomies of scale are present, increasing fund size should constrain a manager's ability to generate added value, as performance deteriorates with growth.

Using a novel simulation-based methodology, adapted from the luck-versus-skill framework of Fama and French 2010. The empirical findings provide strong evidence in support of diseconomies of scale. We show that the ability of skilled managers to generate added value diminishes as fund size increases, consistent with the theoretical predictions of Berk and Green 2004. Importantly, this size-related constraint is most pronounced among managers who exhibit high levels of skill, while unskilled managers—who typically destroy value—appear unaffected. These results suggest a positive correlation between skill and the strength of scale effects.

This study contributes to the literature in two significant ways. First, it offers new insights into the long-standing scale effect puzzle by focusing on value creation rather than return performance. Second, it introduces a methodological innovation by extending and modifying the Fama and French 2010 approach to test for diseconomies of scale.

Taken together, the results underscore the importance of accounting for scale effects when evaluating fund performance and designing compensation or incentive schemes. They also highlight the limitations of alpha as a measure of managerial skill in environments where capital flows adjust in response to performance. Future research may extend this analysis by examining how different sources of fund flows (e.g., retail vs institutional) or fund characteristics (e.g., investment style, liquidity constraints) interact with scale effects to shape value creation in the active management industry.

8 Appendix

8.1 Proof of Proposition 1

The value a manager adds as a function of the amount of money they choose to actively manage is given by:

$$V(q) = a^g(q)q = \phi q - cq^2 \tag{23}$$

I have assumed that managers optimize value, therefore skill V^{opt}, is the maximum value of the equation (23). The first order condition gives the optimal amount of money the manager choose to actively manage, q^{opt}:

$$\frac{dV}{dq}(q) = 0 \Rightarrow q^{opt} = \frac{\phi}{2c} \tag{24}$$

so the skill of the manager is

$$V^{opt} = \phi q^{opt} - c(q^{opt})^2 = \frac{\phi^2}{4c} \tag{25}$$

At the optimum the alpha the manager makes on the actively managed part of his/her portfolio can be calculated by substituting equation (24) to the equation $a^g(q) = \phi - cq$:

$$a_i^{opt} = \phi - c(\frac{\phi}{2c}) = \frac{\phi}{2} \tag{26}$$

The gross alpha of the fund, given total assets under management q_i, is the weighted average of the alpha earned on the actively managed portion and the alpha on the passively managed portion of the portfolio:

$$a^g = (\frac{q^{opt}}{q})a^{opt} + (\frac{q - q^{opt}}{q})0 = (\frac{q^{opt}}{q})a^{opt} \tag{27}$$

Now by using equation (26) and (24), we can simplify equation (27):

$$a^g = \frac{\phi^2}{4qc} \tag{28}$$

For gross alpha (a^g) to serve as a measure of managerial skill, it must equal to the optimal value (V^{opt}). This equality holds only when all managers manage exactly \$1 in assets, that is, when $q = 1$ for all funds. More generally, for gross alpha to act as a proxy for skill, it must be linearly related to V. This linear relationship holds only under the condition that fund sizes are constant across funds—i.e., $q =$ constant for all funds. Those arguments prove the first part of the Proposition 1.

To prove the second part of the Proposition 1, we use (28), to estimate the manager's added value:

$$a^g q = q\frac{\phi^2}{4qc} = \frac{\phi^2}{4c} \tag{29}$$

The the value added of the manager is equal to the optimal value.

8.2 Size Persistence

In this subsection, we report summary statistics on the persistence of mutual fund sizes over time, where size is measured as inflation-adjusted total net assets (real size).

I estimate an AR(1) model via OLS for the real size of each fund in the sample with at least 12 monthly observations. The average estimated autoregressive coefficient is 0.95, with a standard deviation of 0.11. This indicates a high degree of persistence in fund size, consistent with near unit root behavior.

In addition, Figure 3 shows the evolution of the distribution of the logarithm of real fund size, along with the logarithm of the total number of mutual funds over time. The plot reveals that the median real fund size has remained relatively stable throughout the sample period, alleviating concerns about non-stationarity in size. The well-documented growth of the mutual fund industry appears to be driven primarily by a steady increase in the number of funds, as indicated by the upward trend in the black line in Figure 3.

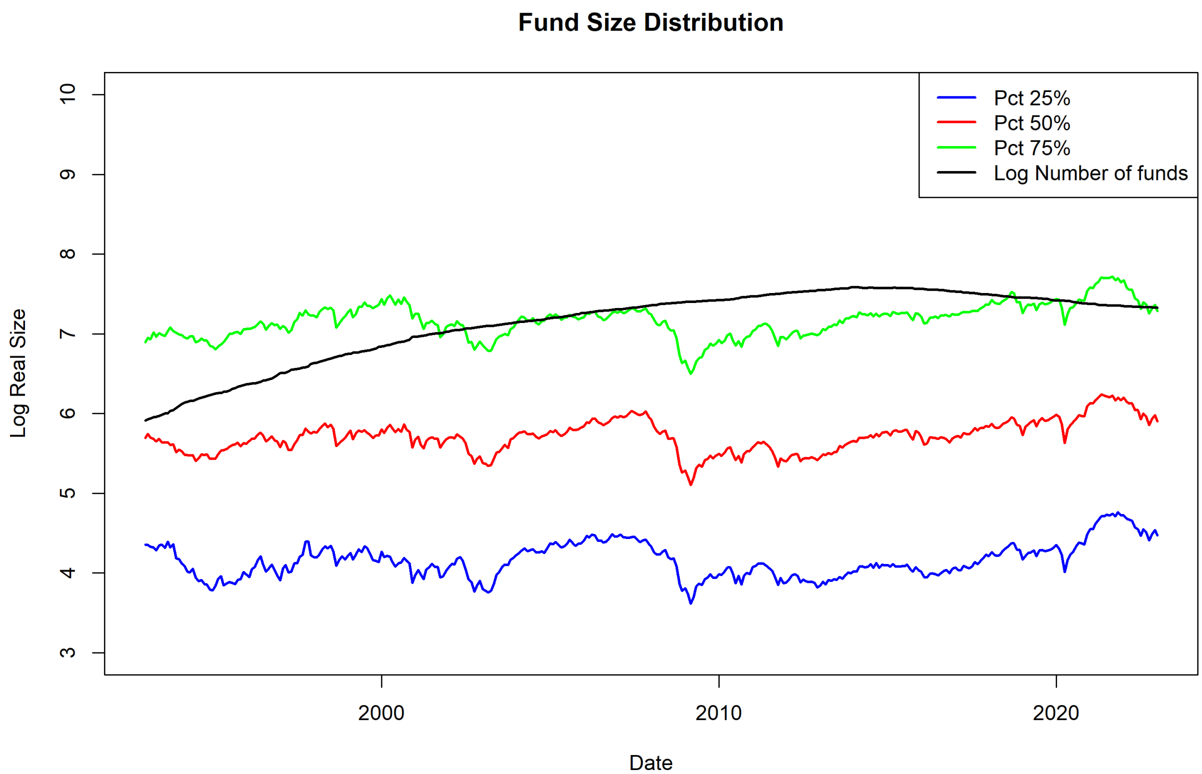

Figure 3: This figure plots the 25th, 50th, and 75th percentiles of the distribution of the logarithm of inflation-adjusted total net assets (in December 2022 dollars) over time, along with the logarithm of the total number of mutual funds (black line).

8.3 Autocorrelation Tests in Returns and Fund Growth Rates

I examine whether there is evidence of autocorrelation in mutual fund returns, mutual fund growth rates, and index returns.

For each mutual fund in the sample with at least 12 return observations, we estimate the first- and second-order autocorrelation coefficients for both fund returns and fund growth rates. Table 9 reports summary statistics for these estimates.

For mutual fund returns, there is no evidence of first- or second-order autocorrelation. The distributions of both coefficients are approximately symmetric around zero, with median values of 0.03 and -0.04, respectively.

The results for fund growth rates are slightly different. The distributions of the first- and second-order autocorrelation coefficients are right-skewed, with median values of 0.02 and 0.01, respectively. A closer look at the right tail of the distribution reveals that a subset of funds exhibits weak positive autocorrelation in growth rates. In particular, the 75th percentile of the first-order autocorrelation coefficient is 0.17, indicating mild persistence in the growth ratios for some funds.

Table 9: Autocorrelation of fund returns and fund growth rates

	Fund Returns Autocorrelation		Fund Growth Autocorrelation	
Stats	Lag 1	Lag 2	Lag 1	Lag 2
Pct10	-0.14	-0.11	-0.01	-0.01
Pct25	-0.07	-0.07	0.00	0.00
Pct50	0.03	-0.04	0.02	0.01
Mean	0.00	-0.04	0.1	0.07
Pct75	0.07	-0.01	0.17	0.1
Pct90	0.12	0.03	0.34	0.23

Next, for each benchmark index in my sample weestimate the first and second order coefficient and Table 10 reports those estimations. The the values of the first and second order autocorrelation coefficients range between $[0.02, 0.06]$ and $[-0.08, -0.03]$, respectively suggesting that there is no autocorrelation in index returns.

Table 10: Autocorrelation of market indexes

	Market Indexes Autocorrelation	
Indexes	Lag1	Lag2
Russell 1000 TR USD	0.02	-0.06
Russell 1000 Growth TR	0.02	-0.03
Russell 1000 Value TR	0.03	-0.07
S&P Mid Cap 400 TR	0.03	-0.08
Russell MidCap Growth TR	0.06	-0.03
Russell MidCap Value TR	0.06	-0.07
Russell 3000 TR	0.02	-0.06
Russell 3000 Growth TR	0.03	-0.03
Russell 3000 Value TR	0.03	-0.07
Russell 2000 (EOD) TR	0.03	-0.04
Russell 2000 Growth TR	0.04	-0.03
Russell 2000 Value TR	0.04	-0.03

8.4 HAC Standard Errors in the Estimation of added value t-statistic

Table 11: Actual and Simulated Added Value Percentiles, Simulated Percentiles are Constructed Under the Null that size does not impose constrains on added value. The table shows values of t-statistic of added value ($t(V)$) at selected percentiles (Pct) for actual (Act) and simulated (Sim) data. The standard errors in the t-statistics are estimated with the Newey and West 1987 heteroskedasticity and autocorrelation consistent (HAC) estimator with three lags. The table also shows the ratio of the 1,000 simulations that produce larger values of $t(V)$ at the selected percentiles that those observed in actual fund data ($\frac{\#Sim>Act}{N_B}$). The period is from January 1993 to December 2022, is required at least 12 return observations for a fund to be included and the results are presented for the ex ante and ex post distributions.

	Ex ante distribution			Ex post distribution		
Pct	Sim	Act	$\frac{\#Sim>Act}{N_B}$	Sim	Act	$\frac{\#Sim>Act}{N_B}$
1	-3.3	-2.89	0.01	-3.04	-2.89	0.2
2	-2.89	-2.64	0.07	-2.68	-2.47	0.13
5	-2.3	-2.1	0.11	-2.15	-2.01	0.23
10	-1.78	-1.67	0.26	-1.67	-1.62	0.42
20	-1.14	-1.19	0.61	-1.05	-1.13	0.67
30	-0.67	-0.82	0.81	-0.58	-0.77	0.83
40	-0.24	-0.47	0.91	-0.17	-0.43	0.92
45	-0.04	-0.31	0.95	0.03	-0.27	0.94
47	0.04	-0.26	0.96	0.11	-0.22	0.96
48	0.08	-0.23	0.97	0.15	-0.19	0.96
50	0.16	-0.17	0.98	0.22	-0.12	0.96
52	0.24	-0.11	0.98	0.3	-0.06	0.98
53	0.28	-0.08	0.98	0.34	-0.03	0.98
55	0.36	-0.01	0.99	0.42	0.04	0.98
60	0.56	0.17	0.99	0.62	0.23	0.98
70	0.97	0.53	1.00	1.03	0.59	0.99
80	1.44	0.97	1.00	1.49	1.03	0.99
90	2.06	1.55	1.00	2.12	1.61	0.99
95	2.56	1.98	1.00	2.62	2.04	1.00
98	3.13	2.57	1.00	3.18	2.65	0.99
99	3.51	2.86	1.00	3.56	3.01	0.98

References

Adams, John C, Darren K Hayunga, and Sattar Mansi (2018). "Diseconomies of scale in the actively-managed mutual fund industry: what do the outliers in the data tell us?" In: *Critical Finance Review, Forthcoming*.

Agyei-Ampomah, Sam et al. (2015). "On luck versus skill when performance benchmarks are style-consistent". In: *Journal of Banking & Finance* 59, pp. 127–145.

Barras, Laurent, Patrick Gagliardini, and Olivier Scaillet (2022). "Skill, scale, and value creation in the mutual fund industry". In: *The Journal of Finance* 77.1, pp. 601–638.

Barras, Laurent, Olivier Scaillet, and Russ Wermers (2010). "False discoveries in mutual fund performance: Measuring luck in estimated alphas". In: *The journal of finance* 65.1, pp. 179–216.

Barrett, Garry F and Stephen G Donald (2003). "Consistent tests for stochastic dominance". In: *Econometrica* 71.1, pp. 71–104.

Ben-David, Itzhak et al. (2022). "What do mutual fund investors really care about?" In: *The Review of Financial Studies* 35.4, pp. 1723–1774.

Berk, Jonathan B and Richard C Green (2004). "Mutual fund flows and performance in rational markets". In: *Journal of political economy* 112.6, pp. 1269–1295.

Berk, Jonathan B and Jules H Van Binsbergen (2016). "Assessing asset pricing models using revealed preference". In: *Journal of Financial Economics* 119.1, pp. 1–23.

— (2015). "Measuring skill in the mutual fund industry". In: *Journal of financial economics* 118.1, pp. 1–20.

Bhojraj, Sanjeev, YOUNG JUN CHO, and Nir Yehuda (2012). "Mutual fund family size and mutual fund performance: The role of regulatory changes". In: *Journal of Accounting Research* 50.3, pp. 647–684.

Binsbergen, Jules Van et al. (2024). "A Horizon-Based Decomposition of Mutual Fund Value Added Using Transactions". In: *The Journal of Finance* 79.3, pp. 1831–1882.

Busse, Jeffrey A et al. (2021). "Transaction costs, portfolio characteristics, and mutual fund performance". In: *Management Science* 67.2, pp. 1227–1248.

Carhart, Mark M (1997). "On persistence in mutual fund performance". In: *The Journal of finance* 52.1, pp. 57–82.

Chen, Joseph et al. (2004). "Does fund size erode mutual fund performance? The role of liquidity and organization". In: *American Economic Review* 94.5, pp. 1276–1302.

Chevalier, Judith and Glenn Ellison (1997). "Risk taking by mutual funds as a response to incentives". In: *Journal of political economy* 105.6, pp. 1167–1200.

Cuthbertson, Keith, Dirk Nitzsche, and Niall O'Sullivan (2022). "Mutual fund performance persistence: Factor models and portfolio size". In: *International Review of Financial Analysis* 81, p. 102133.

Del Guercio, Diane and Paula A Tkac (2002). "The determinants of the flow of funds of managed portfolios: Mutual funds vs. pension funds". In: *Journal of financial and quantitative analysis* 37.4, pp. 523–557.

Elton, Edwin J, Martin J Gruber, and Christopher R Blake (2012). "Does mutual fund size matter? The relationship between size and performance". In: *The Review of Asset Pricing Studies* 2.1, pp. 31–55.

Fama, Eugene F and Kenneth R French (2010). "Luck versus skill in the cross-section of mutual fund returns". In: *The journal of finance* 65.5, pp. 1915–1947.

Ferreira, Miguel A et al. (2013). "The determinants of mutual fund performance: A cross-country study". In: *Review of Finance* 17.2, pp. 483–525.

Harvey, Campbell R and Yan Liu (2022). "Luck versus skill in the cross section of mutual fund returns: Reexamining the evidence". In: *The Journal of Finance* 77.3, pp. 1921–1966.

Hendricks, Darryll, Jayendu Patel, and Richard Zeckhauser (1993). "Hot hands in mutual funds: Short-run persistence of relative performance, 1974–1988". In: *The Journal of finance* 48.1, pp. 93–130.

Hoberg, Gerard, Nitin Kumar, and Nagpurnanand Prabhala (2018). "Mutual fund competition, managerial skill, and alpha persistence". In: *The Review of Financial Studies* 31.5, pp. 1896–1929.

Huang, Haitao et al. (2023). "Bootstrap analysis of mutual fund performance". In: *Journal of econometrics* 235.1, pp. 239–255.

Jensen, Michael C (1968). "The performance of mutual funds in the period 1945-1964". In: *The Journal of finance* 23.2, pp. 389–416.

Kosowski, Robert et al. (2006). "Can mutual fund "stars" really pick stocks? New evidence from a bootstrap analysis". In: *The Journal of finance* 61.6, pp. 2551–2595.

Leippold, Markus and Roger Rueegg (2020). "How rational and competitive is the market for mutual funds?" In: *Review of Finance* 24.3, pp. 579–613.

Ma, Linlin, Yuehua Tang, and Juan-Pedro Gomez (2019). "Portfolio manager compensation in the US mutual fund industry". In: *The Journal of Finance* 74.2, pp. 587–638.

Malkiel, Burton G (1995). "Returns from investing in equity mutual funds 1971 to 1991". In: *The Journal of finance* 50.2, pp. 549–572.

McLemore, Ping (2019). "Do mutual funds have decreasing returns to scale? Evidence from fund mergers". In: *Journal of Financial and Quantitative Analysis* 54.4, pp. 1683–1711.

Newey, Whitney K. and Kenneth D. West (1987). "A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix". In: *Econometrica* 55.3, pp. 703–708. DOI: 10.2307/1913610.

Pástor, L'uboš, Robert F Stambaugh, and Lucian A Taylor (2020). "Fund tradeoffs". In: *Journal of Financial Economics* 138.3, pp. 614–634.

— (2015). "Scale and skill in active management". In: *Journal of financial economics* 116.1, pp. 23–45.

Phillips, Blake, Kuntara Pukthuanthong, and P Raghavendra Rau (2018). "Size does not matter: Diseconomies of scale in the mutual fund industry revisited". In: *Journal of Banking & Finance* 88, pp. 357–365.

Reuter, Jonathan and Eric Zitzewitz (2021). "How much does size erode mutual fund performance? A regression discontinuity approach". In: *Review of Finance* 25.5, pp. 1395–1432.

Roussanov, Nikolai, Hongxun Ruan, and Yanhao Wei (2021). "Marketing mutual funds". In: *The Review of Financial Studies* 34.6, pp. 3045–3094.

Sirri, Erik R and Peter Tufano (1998). "Costly search and mutual fund flows". In: *The journal of finance* 53.5, pp. 1589–1622.

Xu, Wenhao and Taoqin Chen (2024). "Mutual fund value creation: Insights from the residual income model". In: *Finance Research Letters* 62, p. 105254.

Yan, Xuemin Sterling (2008). "Liquidity, investment style, and the relation between fund size and fund performance". In: *Journal of Financial and Quantitative Analysis* 43.3, pp. 741–767.

Zhu, Min (2018). "Informative fund size, managerial skill, and investor rationality". In: *Journal of Financial Economics* 130.1, pp. 114–134.