DISCUSSION PAPER SERIES

DP15548

EXPECTATIONS OF ACTIVE MUTUAL FUND PERFORMANCE

Magnus Dahlquist, Markus Ibert and Felix Wilke

FINANCIAL ECONOMICS

ISSN 0265-8003

EXPECTATIONS OF ACTIVE MUTUAL FUND PERFORMANCE

Magnus Dahlquist, Markus Ibert and Felix Wilke

Discussion Paper DP15548
Published 11 December 2020
Submitted 11 December 2020

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

This Discussion Paper is issued under the auspices of the Centre's research programmes:

- Financial Economics

EXPECTATIONS OF ACTIVE MUTUAL FUND PERFORMANCE

Abstract

We recover a forward-looking distribution of expected abnormal returns (alphas) for active equity mutual funds from analyst ratings. Professional analysts believe that alphas are dispersed, that the average fund will underperform, and that the largest funds will outperform. We estimate a rational expectations learning model of fund performance and confront the model-implied expectations based on fund size, perceived skill, and fees with analysts' expectations. Analysts and the rational learner respond similarly to changes in perceived skill and fees, but in contrast to the rational learner, analysts do not believe in a negative impact of fund size on fund returns. The absence of such decreasing returns to scale in analysts' expectations and the presence thereof in actual fund returns make it difficult to reconcile analysts' expectations with rational expectations, but can help explain the size of the industry together with its poor performance.

JEL Classification: G11, G12, G14, G23

Keywords: Alpha, Expectation Formation, Mutual funds

Magnus Dahlquist - magnus.dahlquist@hhs.se
Stockholm School of Economics and CEPR

Markus Ibert - markus.f.ibert@frb.gov
Federal Reserve Board of Governors

Felix Wilke - felix.wilke@phdstudent.hhs.se
Stockholm School of Economics, Swedish House of Finance

Expectations of Active Mutual Fund Performance

Magnus Dahlquist, Markus Ibert, and Felix Wilke*

November 25, 2020

Abstract

We recover a forward-looking distribution of expected abnormal returns (alphas) for active equity mutual funds from analyst ratings. Professional analysts believe that alphas are dispersed, that the average fund will underperform, and that the largest funds will outperform. We estimate a rational expectations learning model of fund performance and confront the model-implied expectations based on fund size, perceived skill, and fees with analysts' expectations. Analysts and the rational learner respond similarly to changes in perceived skill and fees, but in contrast to the rational learner, analysts do not believe in a negative impact of fund size on fund returns. The absence of such decreasing returns to scale in analysts' expectations and the presence thereof in actual fund returns make it difficult to reconcile analysts' expectations with rational expectations, but can help explain the size of the industry together with its poor performance.

JEL: G11, G12, G14, G23.

Keywords: Alpha, expectation formation, mutual funds.

*Dahlquist: Stockholm School of Economics and CEPR; e-mail: magnus.dahlquist@hhs.se. Ibert: Board of Governors of the Federal Reserve System; e-mail: markus.f.ibert@frb.gov. Wilke: Stockholm School of Economics; e-mail: felix.wilke@phdstudent.hhs.se. We thank Jose Vicente Martinez, Richard Evans, and seminar participants at the Stockholm School of Economics for comments and suggestions.

1 Introduction

Conventional wisdom suggests that active funds do not add value and that the USD 11 trillion currently allocated to actively managed open-end equity mutual funds would be better invested in passive benchmarks (see, e.g., Jensen, 1968; French, 2008). Rational expectations equilibrium models of active management challenge this conventional wisdom and argue that active funds' inability to outperform passive benchmarks net of fees is an equilibrium outcome (see, e.g., Berk and Green, 2004; Pástor and Stambaugh, 2012).

Expectations of future performance are paramount to understand capital allocations in such an equilibrium. The (noisy) rational expectations equilibrium prescription for expectation formation is as follows (Berk and Green, 2004; Berk and van Binsbergen, 2017): Investors, who are uncertain about some of the parameters of the economy (e.g., managerial skill), update their beliefs from observed fund returns, which decrease with fund size, and allocate capital to funds such that expected abnormal fund returns (that is, fund returns in excess of the benchmark) are zero.[1]

To date, no study has investigated whether actual expectations are formed as the rational expectations paradigm prescribes. In this paper, we recover forward-looking expected net-of-fee abnormal returns (henceforth "alphas") for active equity mutual funds from analyst ratings provided by Morningstar, a large financial services firm in the mutual fund industry. Our main contribution is to provide explicit expectations of future fund performance for essentially the entire universe of equity funds and to show that these expectations deviate from rational expectations. The key deviation is that actual fund returns decrease with fund size but analysts' expectations do not (if anything, analysts believe that returns *increase* with size).[2]

[1] The models in Dangl, Wu, and Zechner (2008), Glode and Green (2011), Pástor and Stambaugh (2012), and Stambaugh (2014) share similar features, that is, learning about some parameters, returns that decrease with size, and rational provision of capital.

[2] While earlier evidence is mixed (see, e.g., Chen, Hong, Huang, and Kubik, 2004; Reuter and Zitzewitz,

The micro-foundation for such decreasing returns to scale in actual fund returns is liquidity constraints. Being larger erodes performance as larger trades are associated with a larger price impact and higher execution costs (see, e.g., Kyle, 1985). The concept of decreasing returns to scale is important for at least two reasons. First, it leads researchers to conclude that the commonly reported negative realized net alpha (see, e.g., Fama and French, 2010) implies that most funds manage too much capital and that most funds would perform better if they managed less capital (see, e.g., Roussanov et al., 2018; Leippold and Rueeg, 2020; Song, 2020; Cooper, Halling, and Yang, 2020). Second, it leads researchers to conclude that the before-fee alpha adjusted for size (as opposed to simply the before-fee alpha) is the appropriate measure of investment skill (see, e.g., Berk and van Binsbergen, 2015).

As actual fund returns do decrease with size, the absence of decreasing returns to scale in analysts' expectations does not challenge these conclusions. Instead, it can help explain why many funds are too large. When analysts expect returns to increase with size, they and the capital that follows their recommendations implicitly believe that every fund is always too small. Naturally, then, many funds are large and—since actual fund returns decrease with size—underperform.[3]

Morningstar introduced its "Morningstar Analyst Rating" in 2011, but overhauled its methodology in October 2019 and only then provided a detailed description of how the ratings are constructed. Morningstar analysts assign the ratings on a five-tier scale with three positive ratings of Gold, Silver, and Bronze, as well as a a Neutral rating and a Negative rating. Under the new methodology, Morningstar constructs a distribution of forward-looking alphas and then groups alphas (which are not reported in the database)

2015; Pástor, Stambaugh, and Taylor, 2015), more recent evidence consistently finds evidence for a negative causal impact of size on returns (Zhu, 2018; Roussanov, Ruan, and Wei, 2018; McLemore, 2019; Busse, Chordia, Jiang, and Tang, 2019; Dyakov, Jiang, and Verbeek, 2020; Pástor, Stambaugh, and Taylor, 2020).

[3] Consistent with our results, Choi and Robertson (2019) provide survey evidence indicating that individual investors do not believe that active mutual funds suffer from decreasing returns to scale.

to arrive at the final Analyst Ratings (which are reported in the database). We replicate Morningstar's methodology to recover the alphas that Morningstar analysts use. When we translate our alphas into ratings we can replicate around 90% of Morningstar's ratings.

Figure 1 illustrates the cross section of analyst alphas as of October 2020 (blue, solid bars). In contrast to the equilibrium prediction of zero alphas, analyst alphas are markedly dispersed. While professional analysts expect most funds to underperform, they expect the largest funds to outperform. This is readily apparent when comparing the value-weighted (0.50%) mean with the equal-weighted mean (−1.29%) and median (−1.15%). Higher expectations for larger funds do not necessarily imply a belief in increasing returns to scale as these differences may be driven by omitted characteristics. Specifically, larger funds could simply be perceived as having more skill. However, we argue below that omitted characteristics—in particular, perceived managerial skill—only drive part of the positive relationship between fund size and analyst alphas.

We contrast analyst alphas to alphas obtained from estimating the Berk and Green (2004) rational expectations learning model, but since zero alphas are trivially counterfactual, we relax the equilibrium implication of zero alphas. The model-implied alphas then reflect the expectations of an agent who has rational expectations and learns about managerial skill given return dynamics as prescribed by Berk and Green (2004), but who is agnostic to the equilibrium concept. Figure 1 also overlays the model-implied distribution of alphas (red, transparent bars).

To systematically investigate differences between the rational learner and analysts, we relate alphas to the fund characteristics in the rational expectations learning model: perceived skill, size, and fees. Consistent with learning under rational expectations, analyst alphas decrease with fees and increase with perceived skill. Note that in the model, perceived skill is a sufficient statistic for past performance adjusted for the impact of decreasing returns to scale. While the model's predictions related to perceived skill and fees are broadly consistent with the data, the prediction related to size is not. In one of our main specifications, controlling for perceived skill and fees, we find that a one-unit increase in log size is associated with a 0.08-percentage-point *larger* alpha. This stands in stark contrast to the model's prediction of a 0.19-percentage-point *smaller* alpha per one-unit increase in log size.

When we control for additional manager and fund characteristics, the coefficient estimates on size remain statistically different from the model-implied decreasing returns to scale parameter. Among the additional characteristics that matter for analysts' expectations are managers' personal investments in their funds ("skin in the game") and manager tenure. Consistently, the academic literature has shown a positive relationship between fund performance and both personal investments (Khorana, Servaes, and Wedge, 2007; Evans, 2008; Ibert, 2019) and experience (Greenwood and Nagel, 2009). The inclusion of fund family fixed effects increases R^2 values from 30% to 60%, suggesting that fund family characteristics are important to the formation of analysts' expectations as well. For instance, fund manager compensation practices are likely important and have been shown to systematically differ across fund families (Ibert, Kaniel, Van Nieuwerburgh, and Vestman, 2018; Ma, Tang, and Gómez, 2019). When we add lags of past returns to our main specifications, we find that analysts generally overreact to past returns up to the last six years of returns, consistent with a large literature on excess return chasing among mutual funds (see, e.g., Chevalier and Ellison, 1997; Sirri and Tufano, 1998).

A conservative interpretation of our results is that we can reject the hypothesis that analysts form their expectations according to the rational expectations learning model implied by Berk and Green (2004). Rejecting the model does not necessarily imply rejecting rational expectations in general as the model may be misspecified. In particular, it could miss important characteristics, and once these characteristics are controlled for, the coefficient estimate on size would switch signs. However, since we control for a large number of characteristics, it is unclear to us which crucial characteristics the model could be missing.

Alternatively, the model could falsely detect decreasing returns to scale in actual fund returns, while in truth there are none. Indeed, earlier research questions whether the impact of fund size on fund performance is economically and statistically significant (Reuter and Zitzewitz, 2015; Pástor et al., 2015). However, we also find robust evidence for decreasing returns to scale in actual fund returns when we consider the recursive demeaning estimator in Zhu (2018). In sum, the robust absence of decreasing returns to scale in analysts' expectations combined with the robust presence thereof in actual fund returns leads us to claim that analysts do not have rational expectations in general.

Does the absence of rational expectations in analysts' expectations imply that regular investors do not have rational expectations? In a sample from 2011 to 2015, Armstrong, Genc, and Verbeek (2019) find that fund flows respond to Analyst Ratings, even when the popular backward-looking Star Rating is controlled for. We extend their study of flows to international funds and up to 2020, and find similar results, suggesting that analysts' expectations are at least partly representative of the broader set of investors.[4]

We contribute to several strands of research. First, we contribute to the literature on expectations in macroeconomics and finance (see, e.g., Fuster, Laibson, and Mendel, 2010; Coibion and Gorodnichenko, 2012, 2015; Gennaioli, Ma, and Shleifer, 2016). Mutual funds provide an excellent laboratory in which to study expectation formation for at least three reasons: 1. there are relevant databases similar in quality to equity price databases that provide returns, fees, and assets under management (AUM); 2. rational models of mutual fund performance make clear predictions of how expectations are formed (Berk and Green, 2004; Pástor and Stambaugh, 2012), thereby providing invaluable benchmarks; and 3. the representativeness of analysts' expectations for the broader set of investors can be tested

[4]The results concerning fund flows are in line with a large literature showing that investors respond to Morningstar Star Ratings (Del Guercio and Tkac, 2008; Reuter and Zitzewitz, 2015; Evans and Sun, 2018; Ben-David, Li, Rossi, and Song, 2019) and other easy-to-follow recommendations (e.g., Reuter and Zitzewitz, 2006; Kaniel and Parham, 2018). Huang, Li, and Weng (2020) provide a theoretical justification for the positive impact of ratings on fund flows.

by examining fund flows. The latter is typically not the case for other analyst forecasts or surveys, such as earnings forecasts (see, e.g., La Porta, 1996; Bouchaud, Krüger, Landier, and Thesmar, 2019; Bordalo, Gennaioli, La Porta, and Shleifer, 2019) or inflation forecasts (see, e.g., Malmendier and Nagel, 2016).[5] Our results concerning overreaction in response to past performance in excess of the Berk and Green (2004) benchmark are reminiscent of a large literature on extrapolation and overreaction (see, e.g., Lakonishok, Shleifer, and Vishny, 1994; Bordalo, Gennaioli, Ma, and Shleifer, 2020). Landier, Ma, and Thesmar (2020) show that overreaction is particularly pronounced for less persistent processes, which is consistent with overreaction to past fund performance—a process known to show little persistence (see, e.g., Carhart, 1997). Coincidentally, early evidence on extrapolation comes from the mutual fund flow–performance literature (Ippolito, 1992; Chevalier and Ellison, 1997; Sirri and Tufano, 1998).

Second, we contribute to the literature on models of active management. Rational expectations models can explain the absence of performance persistence and a convex flow–performance relationship (Lynch and Musto, 2003; Berk and Green, 2004; Huang, Wei, and Yan, 2007). The commonly reported negative realized after-fee alpha is a challenge for the Berk and Green (2004) model, but not for rational expectations in general, as models with costly search can rationalize negative average after-fee alphas as well as dispersion in alphas (Garleanu and Pedersen, 2018; Roussanov et al., 2018). Pástor and Stambaugh (2012) relax the assumption that the agent is certain about the decreasing returns to scale parameter in order to rationalize the size of the active fund industry. We are the first to confront rational expectations models with data on actual expectations.[6] Rational expectations models of

[5] One exception is surveys of expected stock market returns, which typically evaluate representativeness by examining aggregate flows into equity mutual funds (see, e.g., Greenwood and Shleifer, 2014).

[6] Some researchers infer investors' risk models using fund flows (Berk and van Binsbergen, 2016; Barber, Huang, and Odean, 2016), while others investigate what prior beliefs can rationalize the large capital allocations to active funds (see, e.g., Baks, Metrick, and Wachter, 2001). However, neither of these approaches yields actual subjective expectations of fund performance.

active management either feature agents who believe in decreasing returns to scale, which makes them inconsistent with analysts' expectations, or they do not feature fund-level decreasing returns to scale, which makes them inconsistent with the fund data. Gennaioli, Shleifer, and Vishny (2015) and Spiegler (2020) allow for deviations from rational expectations. In one version of the model in Gennaioli et al. (2015), investors extrapolate returns and active managers have incentives to pander to investors' beliefs. In Spiegler (2020), rational and extrapolative investors co-exist in equilibrium. The absence of a belief in decreasing returns to scale provides a source of extrapolative expectations: when good past performance increases current fund size and fund returns decrease with size but investors do not believe so, then investors believe that past fund performance is sustainable even though it is not.

Third, we contribute to the literature on the expected future performance of asset managers. Jones and Martinez (2017) study expectations regarding the future performance of U.S. plan sponsors who rank their asset managers on a scale from one to five. We, however, are the first to provide explicit expectations. Jenkinson, Jones, and Martinez (2016) highlight the importance of soft factors for U.S. investment consultants' fund recommendations and Cookson, Jenkinson, Jones, and Martinez (2019) study U.K. investment platforms' fund recommendations. Armstrong et al. (2019) examine the predictive ability of Analyst Ratings for fund performance from 2011 to 2015 and find some evidence that the highest-rated funds outperform. However, they cannot recover alphas, as Morningstar's ratings before October 2019 are informative about the expected performance of an active fund relative to other active funds, but it is unclear to what extent they are informative about the expected performance of an active fund relative to a passive benchmark, which is what alpha measures.

The paper proceeds as follows. Section 2 describes how alphas are part of the construction of Analyst Ratings. Section 3 describes the data. Section 4 outlines and estimates the model. Section 5 presents the main empirical results. Section 6 discusses additional issues. Section 7 discusses implications and provides guidance for future research.

2 Forward-looking Morningstar Ratings

2.1 Old and new ratings

Morningstar has provided Analyst Ratings for a selected number of funds since 2011. Unlike the backward-looking Morningstar Rating (often referred to as the "Star Rating"), the Analyst Rating is the summary expression of Morningstar's forward looking analysis of a fund. Morningstar analysts assign the ratings on a five-tier scale with three positive ratings of Gold, Silver, and Bronze, as well as a Neutral rating and a Negative rating. Morningstar is an independent research firm and does not receive external compensation for constructing its ratings.

Up to October 2019, the Analyst Rating was based on the analyst's conviction of the fund's ability to outperform its peer group *and/or* relevant benchmark on a risk-adjusted basis over the long term. In October 2019, Morningstar overhauled its Analyst Rating system. The most important changes were a greater emphasis on fees and a share-class-specific rating in contrast to a fund-level rating.[7] Under the new rating scheme, a fund is expected to beat both its peer group *and* a relevant benchmark on a risk-adjusted basis to earn a medalist rating (that is, a Bronze, Silver, or Gold rating). The new rating system is therefore informative about alpha, as alpha measures the performance relative to a passive benchmark. In addition, in an effort to increase transparency, Morningstar for the first time also published a document detailing how the Analyst Ratings are constructed under the new methodology. Morningstar constructs alphas by combining a strategy's overall potential with pillar ratings for a fund's "Parent," "People," and "Process." Morningstar then groups the resulting alphas (which are not published in their database) into the aforementioned ratings (which are published in their database).

[7] A fund may have several share classes belonging to the same fund. Share classes of the same fund generally earn the same return before fees, but fees differ across share classes.

Since 2017 Morningstar has also provided forward-looking Quantitative Ratings; these are similar to Analyst Ratings, but are based on a machine-learning algorithm that attempts to mimic a human analyst's decision-making process. We also include funds with a Quantitative Rating in most of our analyses. Table 1 provides a summary of the different Morningstar ratings.

2.2 Analyst and Quantitative Ratings methodology

Morningstar's exact methodology for constructing the ratings follows a three-step process. First, for each fund, Morningstar estimates rolling-window factor regressions starting in January 2000:

$$R_{i,t} - R_{f,t} = \alpha_i + \beta_i(R_{b,i,t} - R_{f,t}) + \zeta_{i,t}, \tag{1}$$

where t runs over a rolling 36-month window, $R_{i,t}$ is the gross (i.e., before-fee) return of fund i, $R_{f,t}$ is a risk-free rate proxy, and $R_{b,i,t}$ is a fund-specific benchmark return. The factor regressions are estimated on the fund level, not the share-class level. The estimated gross alphas are grouped by fund strategy (e.g., U.S. equity large cap blend) to form a distribution of realized alphas. Morningstar then calculates the semi-interquartile range (SIQR) of the distribution (that is, the 75th percentile minus the 25th percentile divided by 2). The SIQR measures the historical alpha dispersion and summarizes Morningstar's assessment of the potential of a given strategy.

Second, Morningstar analysts score a fund based on the three individual pillars "People," "Parent," and "Process." Under the new methodology the scores range from –2 to +2.[8] The Analyst Rating scores are assigned based on an in-depth analysis, must be approved by a ratings committee, and are explained in detail in a written report for each rated fund. We

[8]The labels of the scores –2, –1, 0, +1, and +2 are "Low," "Below Average," "Average," "Above Average," and "High," respectively, and written as such in Morningstar products.

include an anonymized example of such a report in Appendix 7. The Quantitative Rating scores are assigned using a machine-learning algorithm that attempts to mimic a human analyst's decision-making process. The SIQR and the pillar scores are then combined to give an estimate of the before-fee expected abnormal return of a fund:

$$\mathrm{E}_t^s[r_{i,t+1} + f_{i,t}] = \mathrm{SIQR}_{k,i,t} \times \left(0.10 \times \mathrm{Parent}_{i,t} + 0.45 \times \mathrm{People}_{i,t} + 0.45 \times \mathrm{Process}_{i,t}\right), \quad (2)$$

where E_t^s is the analyst's subjective expectation, $r_{i,t+1}$ is the fund's net abnormal return, and $f_{i,t}$ is the fund's fee. The SIQR depends on the type of strategy, k, and acts as a scaling factor. The pillar ratings determine whether a share class receives a positive or negative before-fee alpha.

Third, Morningstar subtracts the share-class-specific fee to arrive at a net alpha for each share class, j, of fund i, that is, $\mathrm{E}_t^s[r_{i,j,t+1}]$. Conditional on a positive net alpha within a particular Morningstar Category, the top 15% of share classes receive a Gold rating, the next 35% receive a Silver rating, and the bottom 50% receive a Bronze rating. Conditional on a negative net alpha within a particular category, the top 70% of share classes receive a Neutral rating and the bottom 30% receive a Negative rating.

Morningstar groups the funds in closely related Morningstar Categories for the first step, but is not explicit about the grouping, nor about the benchmark return or the risk-free rate. We group funds according to their Global Category (a Morningstar variable that groups Morningstar Categories from different domiciled funds), use a fund's Morningstar Category Index as the benchmark, and use the three-month Treasury bill rate as the risk-free rate.

All funds with a Quantitative Rating are rated under the new methodology as of October 2019, but 80 funds with Analyst Ratings have not yet been assigned new Analyst Ratings as of October 2020. In the Online Appendix, we describe how we impute new Analyst Ratings to these funds. The Online Appendix also contains additional details about our replication and the data.

2.3 Replication

We replicate Morningstar's methodology to arrive at the net alphas before they are binned into the final ratings. Table 2 shows that we can replicate the vast majority of Morningstar's Analyst and Quantitative Ratings, suggesting that we indeed recovered the alphas that Morningstar uses to construct the ratings. Panel A shows that for the 8619 share classes with an Analyst Rating under the new methodology, Morningstar assigns a Neutral rating to 3215 share classes. In this case, we assign a Neutral rating in 3061 cases, yielding a replication rate of 95%. Our overall replication rate for the new Analyst Ratings is 88%. Panel B shows our replication of the Morningstar Quantitative Ratings. Our overall replication rate for Quantitative Ratings is 93%.

3 Data

We obtain gross returns, AUM, ratings, and fees for active open-end equity mutual funds from Morningstar Direct. We include all funds in the database to correctly replicate Morningstar's methodology. The sample contains both U.S.-domiciled and non-U.S.-domiciled funds. Although Morningstar only uses data as of January 2000 to construct the Analyst Ratings, we use the entire time series available in Morningstar to estimate the rational expectations model of fund performance. The monthly sample starts in January 1979, the first month for which Morningstar provides benchmark returns, and ends in October 2020. We convert all returns and assets to USD. As is common in the literature, we aggregate share-class-level variables (e.g., fees, returns and analyst alphas) to the fund level by taking an AUM-weighted average.

Figure 2 plots the AUM of funds with an Analyst Rating or Quantitative Rating under the new methodology, funds with an Analyst Rating or Quantitative Rating under the old methodology, and funds with no rating. Table 3 presents summary statistics as of October 2020. The number of funds with a Quantitative Rating is large but the assets of these funds are much smaller on average. Moreover, the table shows that funds with an Analyst Rating have much larger analyst alphas and larger perceived skill (a sufficient statistic for past performance adjusted for decreasing returns to scale, which is introduced below). Put differently, Morningstar assigns Analyst Ratings as opposed to Quantitative Ratings to funds that are larger and have performed better in the past.

4 Rational expectations model

One advantage of working with mutual fund data is the existence of well-established rational benchmarks. We benchmark analyst alphas against alphas as implied by rational learning in a standard model of fund performance. Similar to Berk and Green (2004), we model the abnormal return for fund i as

$$r_{i,t+1} + f_{i,t} = a_{i,t} - c(\mathrm{AUM}_{i,t}) + \epsilon_{i,t+1}, \tag{3}$$

where $\epsilon_{i,t+1} \sim N(0, \sigma_\epsilon^2)$, $r_{i,t+1}$ is the fund's net abnormal return, $a_{i,t}$ is unobservable managerial skill, $f_{i,t}$ is fees, and $c(\mathrm{AUM}_{i,t})$ captures decreasing returns to scale.

Following Roussanov et al. (2018), we generalize Berk and Green (2004) to allow for time-varying skill:

$$a_{i,t} = (1-\rho)a_0 + \rho a_{i,t-1} + \sqrt{1-\rho^2} \cdot \nu_{i,t}, \tag{4}$$

where $\rho \in [0,1]$, the shock is distributed as $v_{i,t} \sim N(0, \sigma_{a,0}^2)$, and skill when a fund is born is distributed as $N(a_0, \sigma_{a,0}^2)$. A rational learner updates her beliefs about managerial skill, $a_{i,t}$

(the only parameter she is uncertain about), from past returns. Allowing for time-varying skill allows the learner to rationally place a greater weight on more recent past performance. A simple Kalman filter argument implies that beliefs at each point in time are given by:

$$\widehat{a}_{i,t+1} = \rho\left(\widehat{a}_{i,t} + \frac{\widehat{\sigma}^2_{a,t}}{\widehat{\sigma}^2_{a,t} + \sigma^2_\epsilon}(r_{i,t+1} - \widehat{a}_{i,t} + c(\mathrm{AUM}_{i,t}) + f_{i,t})\right) + (1-\rho)a_0, \tag{5}$$

$$\widehat{\sigma}^2_{a,t+1} = \rho^2\widehat{\sigma}^2_{a,t}\left(1 - \frac{\widehat{\sigma}^2_{a,t}}{\widehat{\sigma}^2_{a,t} + \sigma^2_\epsilon}\right) + (1-\rho^2)\sigma^2_{a,0}, \tag{6}$$

where $\widehat{\sigma}^2_{a,t}$ describes the uncertainty concerning the perceived skill, $\widehat{a}_{i,t}$, given initial conditions a_0 and $\sigma^2_{a,0}$.[9] Following Roussanov et al. (2018), we assume a log specification for the decreasing returns to scale technology; that is, $c(\mathrm{AUM}) = \eta\log(\mathrm{AUM})$, where η is a parameter capturing the fund returns' sensitivity to an increase in AUM. We use maximum likelihood to estimate the model on the fund level (using gross fund returns and fund size).[10] We run a factor regression as in Equation (1), but over the entire life of a fund using the same benchmark that analysts use and then form $r_{i,t+1} + f_{i,t} = \widehat{\alpha}_i + \zeta_{i,t+1}$, where $\widehat{\alpha}_i$ is the sample average of realized gross abnormal returns. We then annualize the monthly abnormal returns to form the annual abnormal returns. The AUM is measured at the end of the previous year in millions of 2019 USD.

Table 4 presents the parameter estimates and their standard errors. Our parameter estimates are similar to those of Roussanov et al. (2018). Note that their sample differs from ours, as they focus on U.S.-domiciled funds, whereas we also include funds from other domiciles to be consistent with Morningstar's methodology. The estimated prior mean of managerial skill is 1.77% per year, the prior standard deviation is 1.63%, the residual volatility is 7.66%, and the persistence parameter is 0.96. With a standard deviation of log size of 1.87, the de-

[9]Equation (3) is the measurement equation and Equation (4) is the state transition equation in the Kalman filtering.

[10]The model assumes that the residuals are uncorrelated across observations. The assumption is more likely to hold for fund returns instead of share class returns, as the share class returns of a given fund are highly correlated.

creasing returns to scale parameter estimate of 0.19% implies that a one-standard-deviation increase in log size leads to a decrease of 0.36 percentage points in returns.

The model laid out so far is a simple filtering problem, independent of the equilibrium argument in Berk and Green (2004). The Berk and Green (2004) equilibrium implication is that alphas are zero at any point in time. Otherwise, the money of risk-neutral investors would flow into and out of funds, affecting alphas through decreasing returns to scale, and ultimately competing away any alphas. In contrast, a rational learner who is agnostic to the equilibrium concept expects the abnormal return to be

$$\mathrm{E}_t[r_{i,t+1}] = \widehat{a}_{i,t} - \eta \log(\mathrm{AUM}_{i,t}) - f_{i,t}, \tag{7}$$

which may or may not be equal to zero. If the rational learner also has rational expectations, she uses the true parameter values of a_0, $\sigma_{a,0}$, η, σ_ϵ, and ρ, which are approximated by our estimates, to form her expectations. We assume rational expectations to form the alphas at the end of our sample period, October 2020, for every fund according to Equation (7).[11]

In our empirical implementation of the model, the forecast horizon is one year. Morningstar states, for example, that the medalist ratings indicate an expected outperformance "over the long term, meaning a period of at least five years." To compare analyst alphas with those of our model, we assume that analysts' five-year forecasts equal their unobserved one-year forecasts. An alternative would be to iterate Equation (3) forward using a law of motion for AUM and the expected path of fees. However, modeling the path of fees and a law of motion for AUM would significantly complicate the model; it would require additional assumptions as to the fee-setting behavior of the fund over time and as to how investors' money flows into and out of funds in response to past performance.[12]

[11] Since we estimate the model using annual data, we use return data up to December 2019 to estimate a fund's perceived skill.

[12] Similarly, we do not model the possibility that the rational learner could send a signal about the quality of the fund to other investors (as analysts can). If she could send a signal, she would take into account that

5 Main results

5.1 Distribution of alphas

Figure 1 shows the distribution of net-of-fee analyst alphas together with the model-implied alphas in % per year for all funds in the sample as of October 2020. We include funds with a new Analyst Rating, an old Analyst Rating using our imputations, and a Quantitative Rating (henceforth "all ratings"). Analyst alphas are markedly dispersed and far away from the Berk and Green (2004) equilibrium prediction of a zero alpha for every fund.

Figure 3 shows the distribution of analyst alphas together with the historically realized net alphas for the same set of funds. The equal-weighted realized alpha for funds in the sample in October 2020 is -0.65% per year.[13] Analyst alphas are as dispersed as historically realized alphas, which to some extent is due to the way the ratings are constructed (see Section 2). Consistent with analysts' expectations, the literature has long recognized cross-sectional heterogeneity in realized alphas (see, e.g., Kosowski, Timmermann, Wermers, and White, 2006; Fama and French, 2010). Nevertheless, heterogeneous ex ante expectations cannot be taken for granted as many funds exhibit realized alphas different from zero by luck, or lack thereof, alone. For example, Barras, Scaillet, and Wermers (2010) report that 75% of the funds in their sample have a true alpha of zero.

her recommendation could affect flows, and hence the fund's size and in turn the alpha she signals.

[13]The average realized net alpha for all funds ever in the Morningstar data, dead or alive, is negative (-1.01%). There is some debate in the literature as to whether the negative average net alpha arises because of using uninvestable benchmarks (Cremers, Petajisto, and Zitzewitz, 2012; Berk and van Binsbergen, 2015). The Morningstar Category benchmark we use is dictated by Morningstar's methodology and has previously been used in the literature (see, e.g., Pástor et al., 2015).

5.2 Relationship between alphas and perceived skill, size, and fees

How do analysts form their expectations? Although Morningstar details how analysts arrive at their alpha estimates, the key inputs—the individual pillar ratings for "Process," "Parent," and "People"—are a black box to the economist. For instance, analysts do not explicitly rely on past returns other than for calculating the scaling factor, whereas the previous section demonstrates that past returns are one of the key inputs to expectation formation in a standard rational model.

According to the rational expectations learning model, three factors determine alphas: perceived skill, fund size, and fees. We start by investigating the univariate relationships between alphas, "net skill" (perceived skill less fees), and size. We first sort funds into deciles according to their net skill at the end of the sample period and then compute average alphas across deciles for both analysts and the rational learner.

Panel (a) of Figure 4 shows the results for the sample of funds with a new Analyst Rating and Panel (b) shows the results for all ratings. The general pattern across both panels is the same: analyst alphas and the rational learner's alphas increase with net skill. Analysts are more optimistic about funds with a new Analyst Rating compared with all ratings. This is because analysts are the least optimistic about funds with a Quantitative Rating, and these funds constitute most of the all-ratings sample. In Panel (b), both analysts and the rational learner expect the majority of funds to earn negative alphas. This result is reminiscent of the results in Roussanov et al. (2018), that funds up to the 9th decile of net skill have earned negative alphas historically.

We next sort funds into deciles based on fund size. Figure 5 shows a divergence between the rational learner and analysts. Analysts' expectations increase with size, whereas the rational learner's expectations are either unrelated to size as in Panel (a) or increase only slightly with size as in Panel (b). In Panel (b) the mismatch is particularly evident: analysts expect the largest funds to earn abnormal returns of 0.28% per year, whereas the rational learner expects abnormal returns of -0.27%. However, a belief that larger funds perform better does not necessarily imply a belief in increasing returns to scale. Perhaps, larger funds are simply perceived to be more skilled.

For this reason, we formally evaluate the rational expectations learning model in a multivariate regression. Equation (7), together with rational expectations, has clear predictions for a regression of analyst alphas on size (log AUM), perceived skill, and fees: the coefficient estimates should be $-\eta$, 1, and –1, respectively.[14] Table 5 presents two cross-sectional regressions. Specification (1) uses the sample of funds with Analyst Ratings under the new methodology; specification (2) uses the sample of all ratings. In brackets, we report p-values for the null hypothesis that the coefficients are equal to the values predicted by the model.

Fund size. The estimate on log AUM is statistically positive in both columns and has the opposite sign to that of the model's prediction, which leads us to forcefully reject the rational expectations model. For instance, in specification (2) the coefficient estimate on size is 0.08 as opposed to -0.19.

Perceived skill. As the rational expectations model predicts, greater perceived skill is associated with a larger analyst alpha. The coefficient estimate on perceived skill is statistically different from one in specification (1), but not statistically different from one in specification (2).

Fees. As the rational expectations model predicts, an increase in fees is associated with a decrease in analyst alpha. The coefficient estimate on fees is not statistically different from minus one in specification (1), but is statistically different from minus one in specification (2).

The positive coefficient estimates on size suggest that analysts believe that an increase

[14]Moreover, in theory the constant should be zero and the R^2 should be 100%.

in fund size increases returns. Alternatively, the positive coefficient estimate on size could mean that the rational expectations learning model misses important characteristics and that, once these characteristics are controlled for, the coefficient estimate on size would switch signs and be in line with actual fund returns (which decrease with size). We add additional characteristics to our empirical specifications in the next subsection, but find little evidence for this alternative hypothesis. Ultimately, we claim that the divergence between the estimate on size using actual fund returns and the estimate on size using analyst alphas lets us reject not only the specific rational expectations model tested in this subsection, but also rational expectations in general.

That said, evidence regarding missing characteristics also comes from the coefficient estimates on fees. The impact of fund size on fund returns is perhaps hard to grasp given the sophistication required to actually detect decreasing returns to scale in the data and the mixed empirical evidence in previous studies, but common sense suggests that, all else being equal, a one-percentage-point increase in fees should decrease return forecasts by one percentage point. That it does not in specification (2) suggests that other characteristics besides perceived skill, size, and fees are important to analysts.

5.3 Additional determinants of expectations

Morningstar's methodology suggests that the rational expectations model omits variables relevant to analysts' expectation formation. We are guided by Morningstar's methodology in choosing additional variables to explain analysts' expectations. We group variables corresponding to the three pillars "People," "Process," and "Parent." Most of our variables can be obtained directly from Morningstar Direct, which ensures that they are available to analysts.

For "People," we include manager tenure (the longest tenure, in months, of the managers at a fund), manager ownership (the average dollar amount managers at a fund personally invest in the fund), managerial multitasking (the average number of additional funds that the managers of a fund manage), and a dummy for whether a fund is team managed.[15] Manager ownership has been shown to predict fund performance in the U.S. and Sweden (see, e.g., Khorana et al., 2007; Ibert, 2019). However, since ownership information is only publicly available for U.S. funds our sample is restricted.

For "Process," we include a fund's top 10 assets (the percentage of AUM in the ten largest positions), a fund's tracking error (the standard deviation of returns in excess of the benchmark over the life of the fund), and fund turnover (as reported to the SEC).[16] Top 10 assets and tracking error serve as measures of diversification and activeness, respectively.

For "Parent," we include fund family fixed effects. The literature on the role of the fund family has highlighted the fund family's impact on individual fund performance (see, e.g., Massa, 2003; Gaspar, Massa, and Matos, 2006; Ferreira, Matos, and Pires, 2018).

Table 6 shows six specifications. The first three are for the sample of U.S. funds with new Analyst Ratings and the latter three are for the sample of all rated U.S. funds. Specifications (1) and (4) replicate the specifications in Table 5 for the restricted sample of U.S. funds, and show similar results. Specifications (2) and (5) include "People" and "Process" variables, and (3) and (6) additionally include Morningstar Category and fund family fixed effects. We standardize "People" and "Process" variables to mean zero and unit standard deviation, but leave perceived skill, size, and fees unstandardized for comparison to previous tables.

As expected, other characteristics besides perceived skill, size, and fees are important to analysts' expectation formation. In all specifications, manager tenure and ownership are

[15] As of 2005, the SEC requires that mutual fund managers publicly file personal investments in their own funds. Managers have to report whether their dollar ownership in their funds falls into one of the following ranges: \$0, \$1–\$10,000, \$10,001–\$50,000, \$50,001–\$100,000, \$100,001–\$500,000, \$500,001–\$1,000,000, or above \$1,000,000. As in Khorana et al. (2007), we use midpoints of the disclosed ownership ranges to calculate manager ownership, except for the maximum range, "\$1,000,001 and above," for which we us the bottom of the range.

[16] We winsorize fund turnover at the 1st and 99th percentiles as in Pástor, Stambaugh, and Taylor (2017) and do the same with top 10 assets.

positively associated with analysts' expectations. In (5), a one-standard-deviation increase in tenure and ownership is associated with 0.29- and 0.35-percentage-point larger analyst alphas, respectively.

The point estimates on fund size remain positive in all columns, albeit statistically indistinguishable from zero in (2), (3), and (5). Most importantly, the point estimates are still far from the -0.19 point estimate implied by the rational expectations model.

The coefficient estimates on fees in (3) and (6) are not statistically different from minus one, suggesting that the specifications with fund family and category fixed effects satisfy the basic principle of common sense that a one-percentage-point increase in fees decreases analyst alphas by one percentage point. These specifications give us confidence that we have not overlooked other important characteristics that could, once included, lead to a negative coefficient estimate on size. In fact, R^2 values of around 60% suggest that specifications (3) and (6) capture analyst alphas reasonably well.[17]

The presence of decreasing returns to scale in the data and the absence thereof in analysts' expectations lead us to conclude that analysts do not have rational expectations. We believe that this departure from rational expectations says as much about the magnitude of the rational expectations assumption as about the sophistication of analysts. In light of the econometric challenges of actually detecting decreasing returns to scale in the data (see, e.g., Pástor et al., 2015; Zhu, 2018) and the frequent emphasis of the financial press on the benefits of larger scale, the assumption that analysts—let alone regular mutual fund investors—are certain about the decreasing returns to scale parameter is of course conceptually demanding.[18]

[17]The increases are driven by the inclusion of fund family fixed effects as opposed to the inclusion of category fixed effects. We do not take a stance on the fund family variables relevant to analysts, but hypothesize that governance and incentives play a large role (see Evans, Prado, and Zambrana, 2020, for a recent contribution to this literature).

[18]For instance, see a *Financial Times* article entitled "Bigger is better in asset management world" from 6 October, 2019.

One potential avenue to rationalize our results would be to introduce uncertainty and learning about the decreasing returns to scale parameter (as in Pástor and Stambaugh, 2012), but even this avenue appears daunting, as the rational expectations assumption prescribes a prior mean that implies decreasing returns to scale. Given realizations that imply decreasing returns to scale and a prior mean that implies decreasing returns to scale, it appears difficult to arrive at analyst expectations that are mostly indicative of increasing returns to scale.

5.4 Response to past performance

Starting with the early contributions of Ippolito (1992), Chevalier and Ellison (1997), and Sirri and Tufano (1998), a large literature has documented that mutual fund investors chase past returns. Such return chasing is in principle rationalized in Berk and Green (2004), but it may still be excessive. In this subsection, we show that analysts overreact to past performance in excess of what the rational expectations learning model implies.

For the sample of all ratings, we were unable to reject the hypothesis that the coefficient of perceived skill differs from one. We now expose the model to a more stringent test by including several lags of past abnormal returns as additional regressors. The model makes clear predictions of how past returns should affect current expectations: they should have no effect at all once perceived skill, size, and fees are controlled for; hence, under the null hypothesis, the coefficients of past returns are zero. We consider three, six, and nine lags of net abnormal returns, respectively; for completeness we also report results for the sample of new Analyst Ratings.

Table 7 shows the results. We find that analysts overreact to past returns except for the most recent returns, for which we find mixed evidence. Analysts overreact to returns up to at least six years. For instance, in specification (6) the coefficient estimates up to the fifth lag are significantly positive, whereas the sixth lag is insignificant. This result is consistent with a large literature showing that expectations are excessively influenced by more recent news (see, e.g., Bordalo et al., 2020).

6 Additional issues

6.1 Representativeness

One advantage of working with mutual fund data is that the representativeness of analysts' expectations for the broader set of investors can be tested. Armstrong et al. (2019) show that Analyst Ratings positively correlate with fund flows from 2011 to 2015, suggesting that analysts' expectations are to some degree representative of investors' expectations. In the Online Appendix, we extend their analysis to international funds and up to 2020, and find similar results. As we have emphasized before, the caveat to this analysis is that it is unclear to what extent Analyst Ratings measure performance relative to a passive benchmark (that is, alpha) before October 2019.

6.2 Decreasing returns to scale robustness

Our conclusion that analysts do not have rational expectations relies on the presence of decreasing returns to scale in actual fund returns. We find evidence for decreasing returns to scale using our maximum likelihood estimator. In this subsection and the Online Appendix, we show that an alternative estimator detects decreasing returns to scale as well.

Pástor et al. (2015) emphasize that the causal effect of fund size on fund returns should not be identified from cross-sectional (see, e.g., Chen et al., 2004; Yan, 2008; Ferreira, Keswani, Miguel, and Ramos, 2013), but rather from within-fund variation since the matching of manager ability to size is likely not random. As the fund fixed effects estimator is downward biased in samples with a small time series, they propose a recursive demeaning estimator and find a negative but insignificant effect of size on returns. Zhu (2018) argues that the estimator in Pástor et al. (2015) lacks power, proposes an alternative recursive demeaning estimator better suited to the fund size process, and finds a highly significant negative effect of size on returns.

In the Online Appendix, we estimate the relationship between size and returns using the OLS estimator, the fund fixed effects estimator, the recursive demeaning estimator in Pástor et al. (2015) (RD1), and the recursive demeaning estimator in Zhu (2018) (RD2). We find a highly significant effect of fund size on fund returns in all specifications using the RD2 estimator. We estimate the relationship for a commonly used sample of U.S. funds or the entire universe of funds, using a log or a level functional form for fund size and using monthly or annual data. These results are in line with recent research that consistently finds evidence for a negative causal effect of fund size on fund returns (McLemore, 2019; Busse et al., 2019; Dyakov et al., 2020; Pástor et al., 2020).[19]

6.3 An alternative concept of scale

So far, a fund's scale has been given by its size, but active funds differ in how actively they deploy their assets (see, e.g., Cremers and Petajisto, 2009). Intuitively, if two funds manage the same amount of assets but the first fund deviates more from the passive benchmark than does the second fund, the first fund should be subject to steeper decreasing returns to scale as it leaves a bigger footprint in the market (Pástor et al., 2020). In forming her expectations of managerial skill, and ultimately expected abnormal returns, a rational learner should take such differences in scale into account. In the Online Appendix, we show that our results are robust to using activeness times size as opposed to simply size as the measure of scale. We find a significantly positive effect of active fund size on analyst alphas in all specifications.

[19]Most researchers agree that investment opportunities are not arbitrarily scalable in any case. The question in the decreasing returns to scale literature is not whether decreasing returns to scale exist, but whether they are economically large enough to matter (Reuter and Zitzewitz, 2015).

6.4 Conflicts of interest

A longstanding literature documents that analysts' corporate earnings forecasts are excessively optimistic because of conflicts of interest (see, e.g., Hong and Kubik, 2003). Do conflicts of interest drive the awarding of Morningstar Analyst and Quantitative Ratings? We do not believe so for two main reasons.

First, Morningstar's ratings are overly pessimistic relative to the rational expectations benchmark for most funds and only overly optimistic for the largest funds (see Figure 1).

Second, Morningstar claims that its research activities are independent of its commercial activities. In contrast to credit-rating issuers, Morningstar does not receive a fee from fund issuers for its fund analysis, and the coverage of funds is at the discretion of Morningstar's research team and driven by client demand. Moreover, Morningstar's primary business model does not entail acting as a seller of mutual funds, so it is likely not subject to the conflicts of interest that have been shown to affect broker-sold funds (see, e.g., Bergstresser, Chalmers, and Tufano, 2009). In line with these arguments, Cookson et al. (2019) use Morningstar Analyst Ratings as a benchmark for independent analysis when studying investment platforms' recommendations of mutual funds.

7 Concluding remarks and implications

A conservative interpretation of our results is that professional mutual fund analysts do not form their expectations according to the model in Berk and Green (2004). We provide arguments that, more generally, analysts do not have rational expectations. The absence of expectations of decreasing returns to scale and the presence thereof in actual fund returns are robust features of the data and difficult to reconcile with the rational expectations paradigm, under which actual expectations can be recovered empirically by large-sample distributions of the underlying moments.

Ultimately, we want to derive the implications of the divergence from rational expectations for the size and performance of the active fund industry. To do so, we need to translate analysts' expectations into expectations of optimal fund sizes.

If analysts expect returns to increase with size, as we provide evidence for, every fund is always too small. Unlimited amounts of capital should flow into all funds. Similar to the survey evidence regarding stock returns in Greenwood and Shleifer (2014), such expectations are hard to reconcile with any equilibrium in which analysts' expectations are the expectations of a representative agent. Ironically, analysts do believe that most funds will earn negative alphas at their current sizes and analysts do not recommend investing in these funds. However, with a belief in increasing returns to scale, these recommendations would change if negative-alpha funds were to gain additional capital.

This misunderstanding of returns to scale in active management can help explain the enormous amount of capital investors allocate to active funds despite active funds' inability to outperform passive benchmarks, something that has perplexed economists for decades (see, e.g., Cochrane, 2013; Greenwood and Scharfstein, 2013). When actual fund returns decrease with size but analysts—and the capital that follows their recommendations—expect them to increase with size, then naturally many funds are too large and underperform.

Finally, we hope that future research can use the expectations we have uncovered to shed further light on the rationality of mutual fund investors. For example, future research could use a longer time series to test whether analyst forecast errors are predictable. Predictable forecast errors would be further evidence for analyst irrationality. Mutual fund investors' capital allocations in response to analysts' forecast errors could in turn help us examine mutual fund investors' rationality.

Appendix

Morningstar analyst report

Below, we present an anonymized example of an analyst report. The report is for a fund rated under the new methodology and is entitled "Patient process and seasoned managers."

Summary. *The fund's* experienced team and well-defined approach earn Morningstar Analyst Ratings ranging from Silver to Neutral depending on share class fees. The team invests in dividend-paying stocks for total return, not yield. The fund typically boasts a higher yield than the Russell 1000 Value Index and the S&P 500, but that's not its main objective. *The lead manager* looks for companies with business models and management teams capable of generating enough free cash flow to support and grow dividends, and tries to buy shares when they are undervalued relative to their cash flow. *She/he* buys when *she/he* sees at least 35% upside. The team is well equipped for their task. *The lead manager* started *her/his* career in fixed income and *her/his* experience evaluating company cash flows and liabilities has helped this strategy, which *she/he* started managing in 2002. Three comanagers—*manager A*, *manager B*, and *manager C*—averaging 22 years of industry experience and at least a decade with the team, support *her/him.* A senior analyst with five years' experience rounds out the squad. *The lead manager* and *her/his* team have posted a good risk/return profile. The fund's A shares have captured about three fourths of the Russell 1000 Value's and average large-value Morningstar Category peer's downsides since *the lead manager*'s 2002 start through October 2019. Its annualized return matched the index over that period, but its muted volatility led to superior risk-adjusted performance. The portfolio is not without risk. It has some of the largest sector bets in its category. At the end of September 2019, utilities accounted for 19.3% of the portfolio and consumer defensive stocks made up 27.0%. That's 12.3 and 17.3 percentage points, respectively, above the Russell 1000 Value's stakes. Both positions rank in the top 10 of all large-value peers. The portfolio's average debt-to-capital has also steadily increased over the previous five years. But, its average return on equity and return on invested capital have been consistently above the benchmark's. *The lead manager*, however, has managed those risks over more than one market cycle.

Process. This strategy's well-defined approach earns an Above Average Process rating. Management attempts to balance income, capital appreciation, and capital preservation. *The lead manager* and *her/his* team focus on stocks with steady and increasing dividends, but they look beyond the dividend. Each team member conducts research to project a company's total-return potential during the next two to three years, focusing on companies with strong free cash flows and management teams. *The lead manager* and *her/his* comanagers seek capital appreciation by buying stocks that they determined have at least 35% upside from their current price based on cash flow and dividend discount models and other valuation measures. The team aims to preserve capital by modeling a "bear" case for each stock. They consider the market and company factors that could negatively affect the stock's price and require at least a 3-to-1 upside from the bear case to invest. If a stock's price falls more than 15% from its cost basis, a second analyst reviews the stock to provide a "devil's advocate" point of view. This approach produces a portfolio of 70-85 stocks that covers all sectors, though weightings deviate from the Russell 1000 Value Index. The fund may hold up to 25% of its assets in international stocks, and it has held double-digit cash allocations under *the lead manager*'s tenure. Though it has historically provided protection in tough conditions, the current portfolio is not without risks. First, it's heavily concentrated in two sectors: Utilities accounted for 19.3% and consumer defensive stocks 27.0% of the portfolio at the end of September 2019. That's 12.3 and 17.3 percentage points above the Russell 1000 Value Index's stakes, respectively. The heavy helping of consumer defensive stocks is not new, but the bet on utilities relative to the benchmark has risen steadily over the last five years. Its debt-to-capital ratio has also increased over that span and reached 48% in September 2019—10.0 percentage points above its 2014 level and 6.1 percentage points above the benchmark's ratio at the same period. But the companies in the portfolio have been generating solid returns. The portfolio's average return on equity and return on invested capital are both regularly above the benchmark's—the 19.3% ROIC over the last trailing 12 months through September 2019 was nearly 4.8 percentage points above the benchmark's. It has also kept its yield above the Russell 1000 Value and S&P 500. But *the lead manager* and *her/his* team are also looking for companies with at least 35% upside, such as wide moat brewer Anheuser Busch InBev ABI, which has a low ROE and ROIC but has been acquiring growing brands to increase distribution and hopes to increase margins through cost-cutting.

People. Stable leadership earns this strategy an Above Average People rating. *The lead manager* started on the team in 2002 and took over the fund one year after its inception. *She/he* joined *the fund family* in 1991 as a fixed-income trader and managed bond portfolios before shifting to equities in 1998. *The lead manager* has promoted comanagers from analyst positions, such as April 2016 when *she/he* advanced *manager C*, an analyst since early

2009. *Manager A* and *manager B* became comanagers in early 2014, a few months before then-portfolio manager *manager D* left the firm. *Manager A* and *manager B* had 10- and eight-years' experience as analysts on the strategy, respectively, before their promotions. In 2014 *the lead manager* hired experienced *analyst A*, who worked closely with *the fund family* veteran *manager E* before *she/he* retired in 2016. Though the team works collaboratively, each member has sector responsibilities. *The lead manager*, for instance, covers financials and industrials. *She/he* also rotates sector responsibilities and tries to give each team member a mix of cyclical and non-cyclical assignments to keep fresh perspectives on companies. *The lead manager* invests more than $1 million in the fund. *His/her* comanagers have smaller investments (between $100,000 and $500,000). Part of the managers' and analysts' deferred compensation is invested in restricted shares of the fund.

Parent. *The fund family* is a vast conglomerate that is growing further by acquiring *fund family B*. Acquisitions are a way of life for *the fund family*: Among them have been *fund family C* in the 1990s, *fund family D* and *fund family E* in the early 2000s, *fund family F* in 2006, *fund family G* in 2010, and the exchange-traded fund business of *fund family H* more recently. The firm's many areas—whether acquired or homegrown—present a mixed picture. In the United States, areas of strength include small-cap U.S. growth funds, dividend-focused funds, and the international funds run by the *specialized* team. The corporate-bond and quantitative equity teams in Europe also stand out. But many U.S.-focused active stock funds have suffered from poor performance and/or manager turnover. Manager turnover has also been an issue with some Hong Kong-based offerings. Various fixed-income teams in the U.S. are well-staffed, but performance has been so-so. Meanwhile, *the fund family*'s passive side has grown nicely, but there are few truly compelling choices. As for *fund family B*, that firm brings some strong international funds with substantial assets, and the *fund family B* addition also allows for cost-cutting. *The fund family CEO A* has plenty of experience in integrations. All told, along with the bright spots there remain many average or underperforming funds and uncertainty how the *fund family B* merger will play out. *The fund family* thus retains its Neutral Parent rating.

References

Armstrong, Will J., Egemen Genc, and Marno Verbeek, 2019, Going for Gold: An Analysis of Morningstar Analyst Ratings, *Management Science* 65, 2310–2327.

Baks, Klaas P., Andrew Metrick, and Jessica Wachter, 2001, Should Investors Avoid All Actively Managed Mutual Funds? A Study in Bayesian Performance Evaluation, *Journal of Finance* 56, 45–85.

Barber, Brad M., Xing Huang, and Terrance Odean, 2016, Which Factors Matter to Investors? Evidence from Mutual Fund Flows, *Review of Financial Studies* 29, 2600–2642.

Barras, Laurent, Olivier Scaillet, and Russ Wermers, 2010, False Discoveries in Mutual Fund Performance: Measuring Luck in Estimated Alphas, *Journal of Finance* 65, 179–216.

Ben-David, Itzhak, Jiacui Li, Andrea Rossi, and Yang Song, 2019, What Do Mutual Fund Investors Really Care About?, *Working Paper* .

Bergstresser, Daniel, John M.R. Chalmers, and Peter Tufano, 2009, Assessing the Costs and Benefits of Brokers in the Mutual Fund Industry, *Review of Financial Studies* 22, 4129–4156.

Berk, Jonathan B., and Richard C. Green, 2004, Mutual Fund Flows and Performance in Rational Markets, *Journal of Political Economy* 112, 1269–1295.

Berk, Jonathan B., and Jules H. van Binsbergen, 2015, Measuring Skill in the Mutual Fund Industry, *Journal of Financial Economics* 118, 1–20.

Berk, Jonathan B., and Jules H. van Binsbergen, 2016, Assessing Asset Pricing Models Using Revealed Preference, *Journal of Financial Economics* 119, 1–23.

Berk, Jonathan B., and Jules H. van Binsbergen, 2017, Mutual Funds in Equilibrium, *Annual Review of Financial Economics* 9, 147–167.

Blake, Christopher R., and Matthew R. Morey, 2000, Morningstar Ratings and Mutual Fund Performance, *Journal of Financial and Quantitative Analysis* 35, 451–483.

Bordalo, Pedro, Nicola Gennaioli, Rafael La Porta, and Andrei Shleifer, 2019, Diagnostic Expectations and Stock Returns, *Journal of Finance* 74, 2839–2874.

Bordalo, Pedro, Nicola Gennaioli, Yueran Ma, and Andrei Shleifer, 2020, Overreaction in Macroeconomic Expectations, *American Economic Review* Forthcoming.

Bouchaud, Jean Philippe, Philipp Krüger, Augustin Landier, and David Thesmar, 2019, Sticky Expectations and the Profitability Anomaly, *Journal of Finance* 74, 639–674.

Busse, Jeffrey A., Tarun Chordia, Lei Jiang, and Yuehua Tang, 2019, Transaction Costs, Portfolio Characteristics, and Mutual Fund Performance, *Management Science* Forthcoming.

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

Chen, Joseph, Harrison Hong, Ming Huang, and Jeffrey D. Kubik, 2004, Does Fund Size Erode Mutual Fund Performance? The Role of Liquidity and Organization, *American Economic Review* 94, 1276–1302.

Chevalier, Judith, and Glenn Ellison, 1997, Risk Taking by Mutual Funds as a Respone to Incentives, *Journal of Political Economy* 105, 1167–1200.

Choi, James J., and Adriana Z. Robertson, 2019, What Matters to Individual Investors? Evidence from the Horse's Mouth, *Journal of Finance* Forthcoming.

Cochrane, John H., 2013, Finance: Function Matters, Not Size, *Journal of Economic Perspectives* 27, 29–50.

Coibion, Olivier, and Yuriy Gorodnichenko, 2012, What Can Survey Forecasts Tell Us about Information Rigidities?, *Journal of Political Economy* 120, 116–159.

Coibion, Olivier, and Yuriy Gorodnichenko, 2015, Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts, *American Economic Review* 105, 2644–2678.

Cookson, Gordon, Tim Jenkinson, Howard Jones, and Jose Vicente Martinez, 2019, Best Buys and Own Brands: Investment Platforms' Recommendations of Mutual Funds, *Review of Financial Studies* Forthcoming.

Cooper, Michael, Michael Halling, and Wenhao Yang, 2020, The Persistence of Fee Dispersion among Mutual Funds, *Review of Finance* Forthcoming.

Cremers, Martijn, Antti Petajisto, and Eric Zitzewitz, 2012, Should Benchmark Indices Have Alpha? Revisiting Performance Evaluation, *Critical Finance Review* 2, 1–48.

Cremers, Martijn K.J., and Antti Petajisto, 2009, How Active Is Your Fund Manager? A New Measure that Predicts Performance, *Review of Financial Studies* 22, 3329–3365.

Dangl, Thomas, Youchang Wu, and Josef Zechner, 2008, Market Discipline and Internal Governance in the Mutual Fund Industry, *Review of Financial Studies* 21, 2307–2343.

Del Guercio, Diane, and Paula A. Tkac, 2008, Star Power: The Effect of Morningstar Ratings on Mutual Fund Flow, *Journal of Financial and Quantitative Analysis* 43, 907–936.

Dyakov, Teodor, Hao Jiang, and Marno Verbeek, 2020, Trade Less and Exit Overcrowded Markets: Lessons from International Mutual Funds, *Review of Finance* 24, 677–731.

Evans, Allison L., 2008, Portfolio Manager Ownership and Mutual Fund Performance, *Financial Management* 37, 513–534.

Evans, Richard B., Melissa P. Prado, and Rafael Zambrana, 2020, Competition and Cooperation in Mutual Fund Families, *Journal of Financial Economics* 136, 168–188.

Evans, Richard B., and Yang Sun, 2018, Models or Stars: The Role of Asset Pricing Models and Heuristics in Investor Risk Adjustment, *Review of Financial Studies* Forthcoming.

Fama, Eugene F., and Kenneth R. French, 2010, Luck versus Skill in the Cross-Section of Mutual Fund Returns, *Journal of Finance* 65, 1915–1947.

Ferreira, Miguel A., Aneel Keswani, António F. Miguel, and Sofia B. Ramos, 2013, The Determinants of Mutual Fund Performance: A Cross-Country Study, *Review of Finance* 17, 483–525.

Ferreira, Miguel A., Pedro Matos, and Pedro Pires, 2018, Asset Management within Commercial Banking Groups: International Evidence, *Journal of Finance* 73, 2181–2227.

French, Kenneth R., 2008, Presidential Address: The Cost of Active Investing, *Journal of Finance* 63, 1537–1573.

Fuster, Andreas, David Laibson, and Brock Mendel, 2010, Natural Expectations and Macroeconomic Fluctuations, *Journal of Economic Perspectives* 24, 67–84.

Garleanu, Nicolae, and Lasse Heje Pedersen, 2018, Efficiently Inefficient Markets for Assets and Asset Management, *Journal of Finance* 73, 1663–1712.

Gaspar, José Miguel, Massimo Massa, and Pedro Matos, 2006, Favoritism in Mutual Fund Families? Evidence On Strategic Cross-Fund Subsidization, *Journal of Finance* 61, 73–104.

Gennaioli, Nicola, Yueran Ma, and Andrei Shleifer, 2016, Expectations and Investment, *NBER Macroeconomics Annual* 30, 379–442.

Gennaioli, Nicola, Andrei Shleifer, and Robert Vishny, 2015, Money Doctors, *Journal of Finance* 70, 91–114.

Glode, Vincent, and Richard C. Green, 2011, Information Spillovers and Performance Persistence for Hedge Funds, *Journal of Financial Economics* 101, 1–17.

Greenwood, Robin, and Stefan Nagel, 2009, Inexperienced Investors and Bubbles, *Journal of Financial Economics* 93, 239–258.

Greenwood, Robin, and David Scharfstein, 2013, The Growth of Finance, *Journal of Economic Perspectives* 27, 3–28.

Greenwood, Robin, and Andrei Shleifer, 2014, Expectations of Returns and Expected Returns, *Review of Financial Studies* 27, 714–746.

Hartzmark, Samuel M., and Abigail B. Sussman, 2019, Do Investors Value Sustainability? A Natural Experiment Examining Ranking and Fund Flows, *Journal of Finance* 74, 2789–2837.

Hong, Harrison, and Jeffrey D Kubik, 2003, Analyzing the Analysts: Career Concerns and Biased Earnings Forecasts, *Journal of Finance* 58, 313–351.

Huang, Chong, Fei Li, and Xi Weng, 2020, Star Ratings and the Incentives of Mutual Funds, *Journal of Finance* 75, 1715–1765.

Huang, Jennifer, Kelsey D. Wei, and Hong Yan, 2007, Participation Costs and the Sensitivity of Fund Flows to Past Performance, *Journal of Finance* 62, 1273–1311.

Ibert, Markus, 2019, What Do Mutual Fund Managers' Private Portfolios Tell Us About Their Skills?, *Working Paper* .

Ibert, Markus, Ron Kaniel, Stijn Van Nieuwerburgh, and Roine Vestman, 2018, Are Mutual Fund Managers Paid for Investment Skill?, *Review of Financial Studies* 31, 715–772.

Ippolito, Richard A., 1992, Consumer Reaction to Measures of Poor Quality: Evidence from the Mutual Fund Industry, *Journal of Law and Economics* 35, 45–70.

Jenkinson, Tim, Howard Jones, and Jose Vicente Martinez, 2016, Picking Winners? Investment Consultants' Recommendations of Fund Managers, *Journal of Finance* 71, 2333–2370.

Jensen, Michael C., 1968, The Performance of Mutual Funds in the Period 1945-1964, *Journal of Finance* 23, 389–416.

Jones, Howard, and Jose Vicente Martinez, 2017, Institutional Investor Expectations, Manager Performance, and Fund Flows, *Journal of Financial and Quantitative Analysis* 52, 2755–2777.

Kaniel, Ron, and Robert Parham, 2018, WSJ Category Kings—The Impact of Media Attention on Consumer and Mutual Fund Investment Decisions, *Journal of Financial Economics* 123, 337–356.

Khorana, Ajay, and Edward Nelling, 1998, The Determinants and Predictive Ability of Mutual Fund Ratings, *Journal of Investing* 7, 61–66.

Khorana, Ajay, Henri Servaes, and Lei Wedge, 2007, Portfolio Manager Ownership and Fund Performance, *Journal of Financial Economics* 85, 179–204.

Kosowski, Robert, Allan Timmermann, Russ Wermers, and Hal White, 2006, Can Mutual Fund "Stars" Really Pick Stocks? New Evidence from a Bootstrap Analysis, *Journal of Finance* 61, 2551–2595.

Kyle, Albert S., 1985, Continuous Auctions and Insider Trading, *Econometrica* 53, 1315–1335.

La Porta, Rafael, 1996, Expectations and the Cross-Section of Stock Returns, *Journal of Finance* 51, 1715–1742.

Lakonishok, Josef, Andrei Shleifer, and Robert W. Vishny, 1994, Contrarian Investment, Extrapolation, and Risk, *Journal of Finance* 49, 1541–1578.

Landier, Augustin, Yueran Ma, and David Thesmar, 2020, Biases in Expectations: Experimental Evidence, *Working Paper* .

Leippold, Markus, and Roger Rueeg, 2020, How Rational and Competitive Is the Market for Mutual Funds?, *Review of Finance* 24, 579–613.

Lynch, Anthony W., and David K. Musto, 2003, How Investors Interpret Past Fund Returns, *Journal of Finance* 58, 2033–2058.

Ma, Linlin, Yuehua Tang, and Juan Pedro Gómez, 2019, Portfolio Manager Compensation in the U.S. Mutual Fund Industry, *Journal of Finance* 74, 587–638.

Malmendier, Ulrike, and Stefan Nagel, 2016, Learning from Inflation Experiences, *Quarterly Journal of Economics* 131, 53–87.

Massa, Massimo, 2003, How Do Family Strategies Affect Fund Performance? When Performance-Maximization Is Not the Only Game in Town, *Journal of Financial Economics* 67, 249–304.

McLemore, Ping, 2019, Do Mutual Funds Have Decreasing Returns to Scale? Evidence from Fund Mergers, *Journal of Financial and Quantitative Analysis* 54, 1683–1711.

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

Pástor, Luboš, Robert F. Stambaugh, and Lucian A. Taylor, 2015, Scale and Skill in Active Management, *Journal of Financial Economics* 116, 23–45.

Pástor, Luboš, Robert F. Stambaugh, and Lucian A. Taylor, 2017, Do Funds Make More When They Trade More?, *Journal of Finance* 72, 1483–1528.

Pástor, Ľuboš, Robert F. Stambaugh, and Lucian A. Taylor, 2020, Fund Tradeoffs, *Journal of Financial Economics* Forthcoming.

Reuter, Jonathan, and Eric Zitzewitz, 2006, Do Ads Influence Editors? Advertising and Bias in Financial Media, *Quarterly Journal of Economics* 121, 197–227.

Reuter, Jonathan, and Eric Zitzewitz, 2015, How Much Does Size Erode Mutual Fund Performance? A Regression Discontinuity Approach, *Working Paper* .

Roussanov, Nikolai L., Hongxun Ruan, and Yanhao Wei, 2018, Marketing Mutual Funds, *Review of Financial Studies* forthcoming.

Sharpe, William F., 1998, Morningstar's Risk-Adjusted Ratings, *Financial Analysts Journal* 54, 21–33.

Sirri, Erik R., and Peter Tufano, 1998, Costly Search and Mutual Fund Flows, *Journal of Finance* 53, 1589–1622.

Song, Yang, 2020, The Mismatch Between Mutual Fund Scale and Skill, *Journal of Finance* Forthcoming.

Spiegler, Ran, 2020, A Simple Model of a Money-Management Market with Rational and Extrapolative Investors, *European Economic Review* 127, 1–17.

Stambaugh, Robert F., 2014, Presidential Address: Investment Noise and Trends, *Journal of Finance* 69, 1415–1453.

Yan, Xuemin, 2008, Liquidity, Investment Style, and the Relation between Fund Size and Fund Performance, *Journal of Financial and Quantitative Analysis* 43, 741–767.

Zhu, Min, 2018, Informative Fund Size, Managerial Skill, and Investor Rationality, *Journal of Financial Economics* 130, 114–134.

Table 1: Overview of Morningstar's fund ratings

	Star Rating	Analyst Rating	Quantitative Rating	Sustainability Rating
Introduction	1985	2011	2017	2016
Key inputs	Historical fund returns	*New*: Three-pillar ratings (People, Process, and Parent), SIQR (dispersion of CAPM alphas of fund strategy), and share-class fees *Old*: Five-pillar ratings (People, Process, Parent, Performance, and Price)	*New*: Three-pillar ratings (People, Process, and Parent) estimated using a machine-learning algorithm, SIQR (dispersion of CAPM alphas of fund strategy), and share-class fees *Old*: Five-pillar ratings (People, Process, Parent, Performance, and Price) estimated using a machine-learning algorithm	Sustainalytics' company-level ESG Risk Rating
Backward- or forward-looking	Backward-looking	Forward-looking	Forward-looking	Forward-looking
Rating scale	***** **** *** ** *	Gold Silver Bronze Neutral Negative	Gold Silver Bronze Neutral Negative	5 globes 4 globes 3 globes 2 globes 1 globe
Rating level	Share class	*New*: Share class *Old*: Fund ("representative" share class)	Share class	Fund
Rating peer group	Morningstar Category	Morningstar Category	Morningstar Category	Morningstar Global Category

Continued on next page

Table 1 continued from previous page

	Star Rating	Analyst Rating	Quantitative Rating	Sustainability Rating
Ranking metric to award ratings	Morningstar Risk-Adjusted Return	Share-class alphas from Analyst and Quantitative Rating methodology	Share-class alphas from Analyst and Quantitative Rating methodology	Morningstar Historical Portfolio Sustainability Score
Medalist ranking (Gold, Silver, and Bronze) requirement		*New*: Beat benchmark index *and* peer group average *Old*: Beat benchmark index *and/or* peer group average	*New*: Beat benchmark index *and* peer group average *Old*: Beat benchmark index *and/or* peer group average	
Major updates	06/2002: Ratings assigned within Morningstar Categories (before broad asset classes, e.g., equity)	10/2019: Ratings assigned at share-class level based on expected net-of-fee alphas, reduction to three pillars, and higher bar for medalist ranking	10/2019: Ratings assigned at share-class level based on expected net-of-fee alphas, reduction to three pillars, and higher bar for medalist ranking	10/2019: Replacement of Sustainalytics' company ESG Rating with its ESG Risk Rating
Selected academic sources and sample periods for the analysis	Ben-David et al. (2019), 1991–2011, Blake and Morey (2000), 1992–1997, Del Guercio and Tkac (2008), 1996–1999, Evans and Sun (2018), 1999–2005, Khorana and Nelling (1998), 1992–1995, Sharpe (1998)	Armstrong et al. (2019), 2011–2015		Hartzmark and Sussman (2019), 2016–2017

The table compares key features of Morningstar fund ratings. The Morningstar Rating (commonly referred to as the Star Rating) is a purely quantitative, backward-looking measure of a fund's past performance. The Morningstar Analyst Rating is forward-looking and conveys an analyst's conviction of a fund's investment merits. The Morningstar Quantitative Rating is derived from a machine-learning model and attempts to replicate the Analyst Rating a Morningstar analyst might assign to a fund if a human analyst covered it. The Morningstar Sustainability Rating assesses the risk exposure of an investment portfolio to environmental, social, and governance (ESG) factors.

Table 2: Replication of Morningstar Analyst and Quantitative Ratings

Panel A: Morningstar Analyst Ratings

	Replicated rating					
Actual rating	Negative	Neutral	Bronze	Silver	Gold	Total
Negative	61	21	0	0	0	82
Neutral	52	3061	100	2	0	3215
Bronze	1	165	2263	239	4	2672
Silver	3	0	238	1642	103	1986
Gold	0	0	2	128	526	656
Total	117	3247	2603	2011	633	8611

Panel B: Morningstar Quantitative Ratings

	Replicated rating					
Actual rating	Negative	Neutral	Bronze	Silver	Gold	Total
Negative	12752	416	0	0	0	13168
Neutral	490	26345	547	5	1	27388
Bronze	0	1006	6614	435	3	8058
Silver	0	27	536	4547	140	5250
Gold	0	6	3	226	2305	2540
Total	13242	27800	7700	5213	2449	56404

The table shows how well Morningstar Analyst and Quantitative Ratings on the share class level under the new ratings methodology are replicated. The actual Morningstar Analyst Ratings are tabulated in rows, whereas the replicated ratings are tabulated in columns.

Table 3: Summary statistics

	N	Mean (V.W.)	Mean (E.W.)	S.D.	10%	25%	50%	75%	90%
Panel A: Assets under management									
New Analyst Rating	1420		4189	12535	122	353	1076	3390	8734
Old Analyst Rating	80		410	627	8	34	173	512	1115
Quantitative Rating	12262		352	1046	9	26	86	289	802
All ratings	13762		748	4306	10	30	109	400	1271
No rating	4038		151	1354	5	11	31	95	256
All	17800		613	3849	8	22	79	298	1011
Panel B: Fees									
New Analyst Rating	1420	0.79	1.07	0.38	0.65	0.84	1.01	1.25	1.58
Old Analyst Rating	80	1.10	1.22	0.60	0.57	0.82	1.01	1.71	2.00
Quantitative Rating	12262	1.11	1.45	0.71	0.66	0.96	1.38	1.84	2.28
All ratings	13762	0.92	1.41	0.69	0.65	0.94	1.31	1.79	2.24
No rating	4038	1.21	1.69	0.90	0.85	1.11	1.66	2.00	2.48
All	17800	0.94	1.47	0.75	0.68	0.96	1.39	1.86	2.28
Panel C: Perceived skill									
New Analyst Rating	1420	2.48	2.27	0.62	1.59	1.84	2.19	2.61	3.07
Old Analyst Rating	80	2.08	1.94	0.53	1.40	1.58	1.86	2.20	2.69
Quantitative Rating	12262	2.00	1.72	0.58	1.07	1.43	1.73	1.98	2.38
All ratings	13762	2.28	1.78	0.61	1.11	1.46	1.77	2.06	2.51
No rating	4038	2.10	1.79	0.54	1.28	1.61	1.77	1.89	2.27
All	17800	2.27	1.78	0.59	1.15	1.49	1.77	2.02	2.46
Panel D: Analyst alphas									
New Analyst Rating	1420	1.26	0.59	1.34	–1.09	–0.31	0.66	1.41	2.19
Old Analyst Rating	80	0.44	0.36	1.30	–1.61	–0.86	0.51	1.40	2.11
Quantitative Rating	12262	–0.56	–1.52	2.38	–4.56	–3.07	–1.46	0.11	1.43
All ratings	13762	0.50	–1.29	2.39	–4.40	–2.86	–1.15	0.42	1.58

The table shows value-weighted (V.W., by assets under management, AUM) and equal-weighted (E.W.) means, standard deviations, and various percentiles of AUM, fees, skill, and alphas for global active equity mutual funds in October 2020. AUM is the fund size in millions of USD. Perceived skill is managerial skill estimated from a rational model of fund performance. Analyst alphas are Morningstar analysts' expectations of future abnormal net-of-fee fund performance. Fees, perceived skill, and analyst alphas are expressed in % per year.

Table 4: Parameter estimates of the rational fund performance model

Parameter	Description	Estimate
η	Decreasing returns to scale (%)	0.187***
		(0.012)
a_0	Prior mean (%)	1.766***
		(0.058)
$\sigma_{a,0}$	Prior standard deviation (%)	1.634***
		(0.039)
σ_ϵ	Residual standard deviation (%)	7.665***
		(0.014)
ρ	Skill persistence	0.956***
		(0.007)

The table shows the parameter estimates of the rational fund performance model in % per year. Standard errors are in parentheses. The model is estimated using fund-year observations from 1979–2019. *, **, and *** denote 10%, 5%, and 1% significance levels, respectively, for the null of a zero coefficient.

Table 5: Cross-sectional regressions of alphas on characteristics

	Analyst Ratings	All ratings
	(1)	(2)
Perceived skill	0.643***	1.020***
	(0.066)	(0.037)
	[0.000]	[0.595]
Size (× 100)	0.060***	0.084***
	(0.022)	(0.011)
	[0.000]	[0.000]
Fees	−0.877***	−1.359***
	(0.120)	(0.030)
	[0.303]	[0.000]
Constant (× 100)	−0.362	−1.594***
	(0.214)	(0.085)
	[0.091]	[0.000]
N	1420	13762
Adj. R^2	0.16	0.28

The table shows regressions of Morningstar analyst fund alphas on skill as perceived by a rational learner, fund size (logarithm of assets under management in millions of USD), and fees. Specification (1) uses funds with an Analyst Rating under the new methodology. Specification (2) uses funds with a new Analyst Rating, an Analyst Rating under the old methodology, or a Quantitative Rating. Standard errors are heteroskedasticity robust and in parentheses. *, **, and *** denote 10%, 5%, and 1% significance levels, respectively, for the null of a zero coefficient. In brackets are p-values for the null hypothesis that the coefficients of skill, size, fees, and the constant are equal to the model-predicted parameters of +1, −0.187 (the estimate of η in Table 4), −1, and 0, respectively.

Table 6: Cross-sectional regressions of alphas on additional characteristics

	Analyst Ratings			All ratings		
	(1)	(2)	(3)	(4)	(5)	(6)
Rational learner						
Perceived skill	0.578***	0.639***	0.418***	1.078***	1.033***	0.658***
	(0.095)	(0.093)	(0.099)	(0.087)	(0.081)	(0.081)
Size (× 100)	0.130***	0.046	0.036	0.129***	0.028	0.093***
	(0.036)	(0.036)	(0.039)	(0.023)	(0.023)	(0.023)
Fees	−1.599***	−1.484***	−1.218***	−1.690***	−1.889***	−1.259***
	(0.193)	(0.205)	(0.249)	(0.143)	(0.141)	(0.142)
People						
Manager tenure		0.067*	0.110***		0.292***	0.278***
		(0.038)	(0.035)		(0.035)	(0.034)
Manager ownership		0.178***	0.155***		0.352***	0.208***
		(0.039)	(0.042)		(0.038)	(0.038)
Managerial multitasking		0.414***	0.774***		−0.029	0.005
		(0.092)	(0.175)		(0.075)	(0.139)
Management team		0.034	0.071		0.151***	0.169***
		(0.054)	(0.055)		(0.047)	(0.044)
Process						
Top 10 assets (%)		−0.059	0.140		0.013	0.158**
		(0.076)	(0.092)		(0.067)	(0.071)
Tracking error		−0.095	−0.099		−0.074	−0.126**
		(0.067)	(0.101)		(0.066)	(0.060)
Turnover ratio		−0.468***	−0.552***		−0.203**	−0.065
		(0.127)	(0.144)		(0.088)	(0.068)
N	691	691	642	2742	2742	2541
Adj. R^2	0.24	0.30	0.59	0.23	0.29	0.62
Morningstar Category FE	No	No	Yes	No	No	Yes
Fund Family FE	No	No	Yes	No	No	Yes

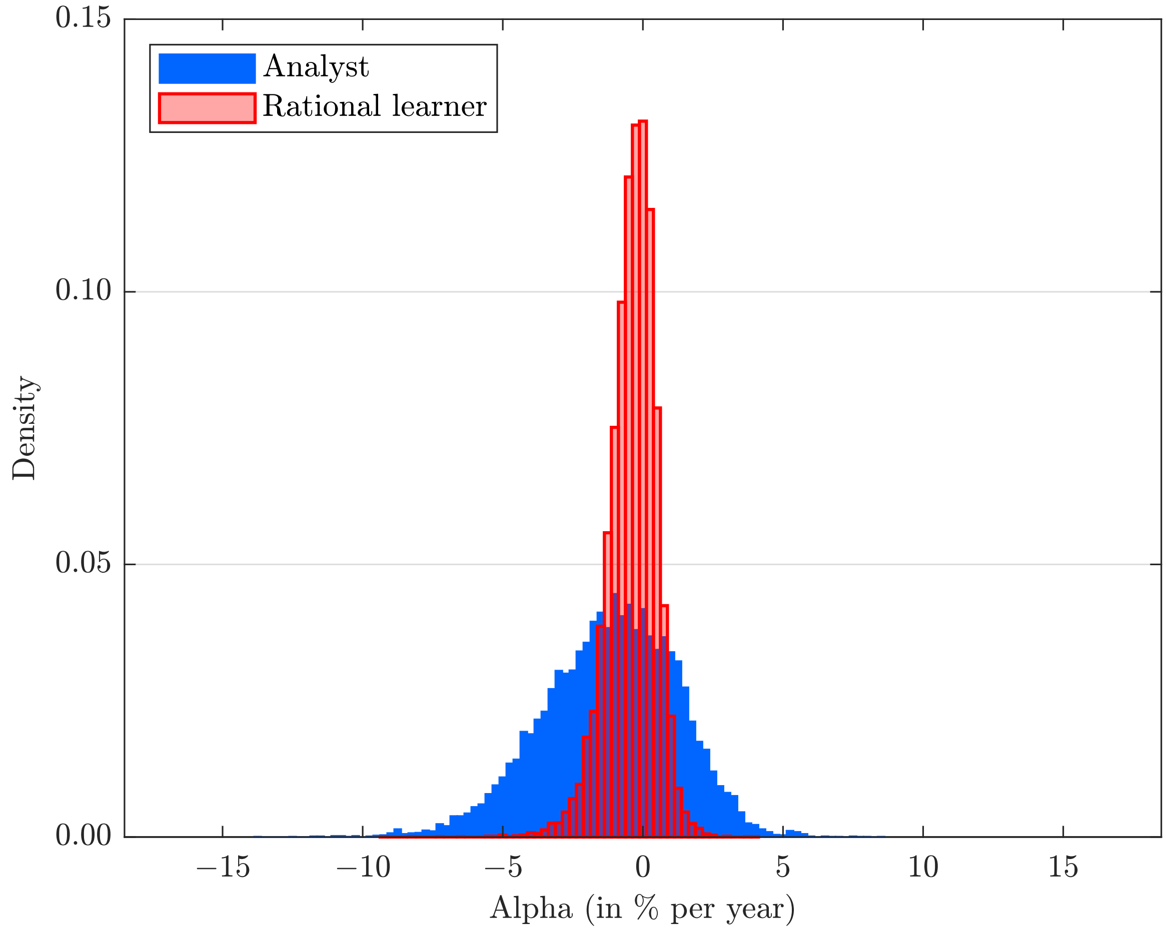
The table shows regressions of Morningstar analyst alphas on fund and manager characteristics. Specifications (1)–(3) use U.S. funds with an Analyst Rating under the new methodology. Specifications (4)–(6) use U.S. funds with a new Analyst Rating, an Analyst Rating under the old methodology, and a Quantitative Rating. Manager tenure is the maximum tenure (in months) taken over all managers, manager ownership is the average amount managers at a fund personally invest in the fund, managerial multitasking is the average number of additional funds managers of a particular fund manage, and management team is a dummy for team-managed funds. "People" and "Process" variables are standardized to zero mean and unit variance, and the coefficient estimates are multiplied by 100. Standard errors are heteroskedasticity robust and in parentheses. *, **, and *** denote 10%, 5%, and 1% significance levels, respectively, for the null of a zero coefficient.

Table 7: Cross-sectional regressions of alphas on past returns

	Analyst Ratings			All ratings		
	(1)	(2)	(3)	(4)	(5)	(6)
Perceived skill	0.639***	0.262***	0.172	0.642***	0.332***	0.231**
	(0.082)	(0.095)	(0.122)	(0.053)	(0.070)	(0.103)
Net abn. return	−0.026***	−0.017**	−0.016*	0.021***	0.031***	0.031***
	(0.008)	(0.008)	(0.009)	(0.005)	(0.005)	(0.006)
Net abn. return (t–1)	0.022***	0.032***	0.031***	0.055***	0.076***	0.078***
	(0.007)	(0.008)	(0.009)	(0.005)	(0.005)	(0.006)
Net abn. return (t–2)	0.024***	0.038***	0.042***	0.036***	0.053***	0.057***
	(0.008)	(0.009)	(0.009)	(0.004)	(0.005)	(0.006)
Net abn. return (t–3)		0.055***	0.053***		0.033***	0.034***
		(0.007)	(0.009)		(0.005)	(0.006)
Net abn. return (t–4)		0.040***	0.038***		0.023***	0.032***
		(0.008)	(0.008)		(0.005)	(0.006)
Net abn. return (t–5)		0.025***	0.025***		0.019***	0.020***
		(0.008)	(0.008)		(0.005)	(0.006)
Net abn. return (t–6)			0.003			0.001
			(0.006)			(0.005)
Net abn. return (t–7)			0.008			0.004
			(0.010)			(0.006)
Net abn. return (t–8)			0.022***			0.019***
			(0.007)			(0.005)
Size (× 100)	0.065***	0.100***	0.125***	0.080***	0.084***	0.094***
	(0.023)	(0.025)	(0.027)	(0.012)	(0.014)	(0.015)
Fees	−0.850***	−0.708***	−0.691***	−1.274***	−1.179***	−1.117***
	(0.125)	(0.121)	(0.126)	(0.035)	(0.042)	(0.045)
Constant (× 100)	−0.425*	−0.020	−0.051	−0.830***	−0.284**	−0.176
	(0.236)	(0.258)	(0.289)	(0.109)	(0.137)	(0.182)
N	1359	1271	1164	11495	9410	7551
Adj. R^2	0.18	0.23	0.24	0.31	0.32	0.32

The table shows regressions of Morningstar analyst alphas on skill as perceived by a rational learner, past net-of-fee abnormal fund returns, fund size (logarithm of assets under management in millions of USD), and fees. Specifications (1) and (2) use funds with an Analyst Rating under the new methodology. Specifications (3) and (4) use funds with a new Analyst Rating, an Analyst Rating under the old methodology, and a Quantitative Rating. Standard errors are heteroskedasticity robust and in parentheses. *, **, and *** denote 10%, 5%, and 1% significance levels, respectively, for the null of a zero coefficient.

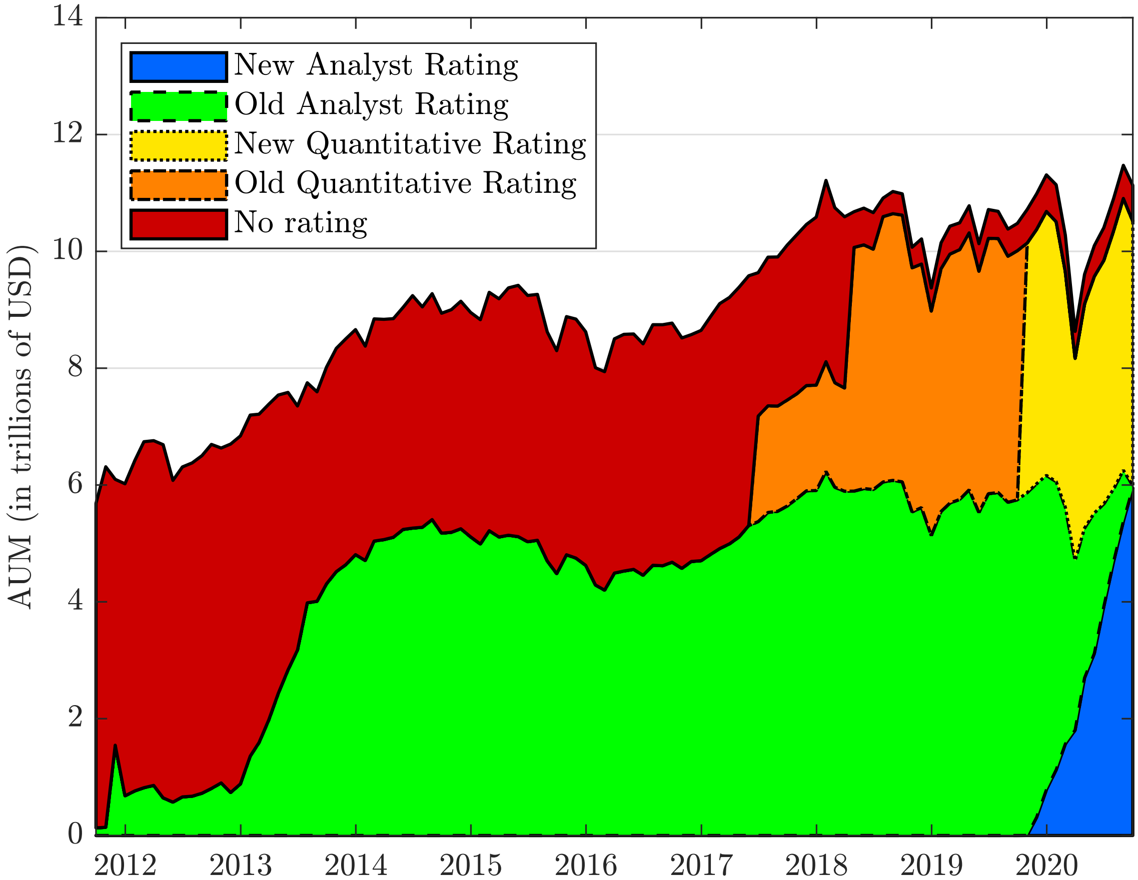
Figure 1: Histogram of analyst alphas and model-implied alphas



	Mean (V.W.)	Mean (E.W.)	S.D.	10%	25%	50%	75%	90%
Analyst	0.50	−1.29	2.39	−4.40	−2.86	−1.15	0.42	1.58
Rational learner	−0.26	−0.51	0.85	−1.56	−0.98	−0.41	0.05	0.44

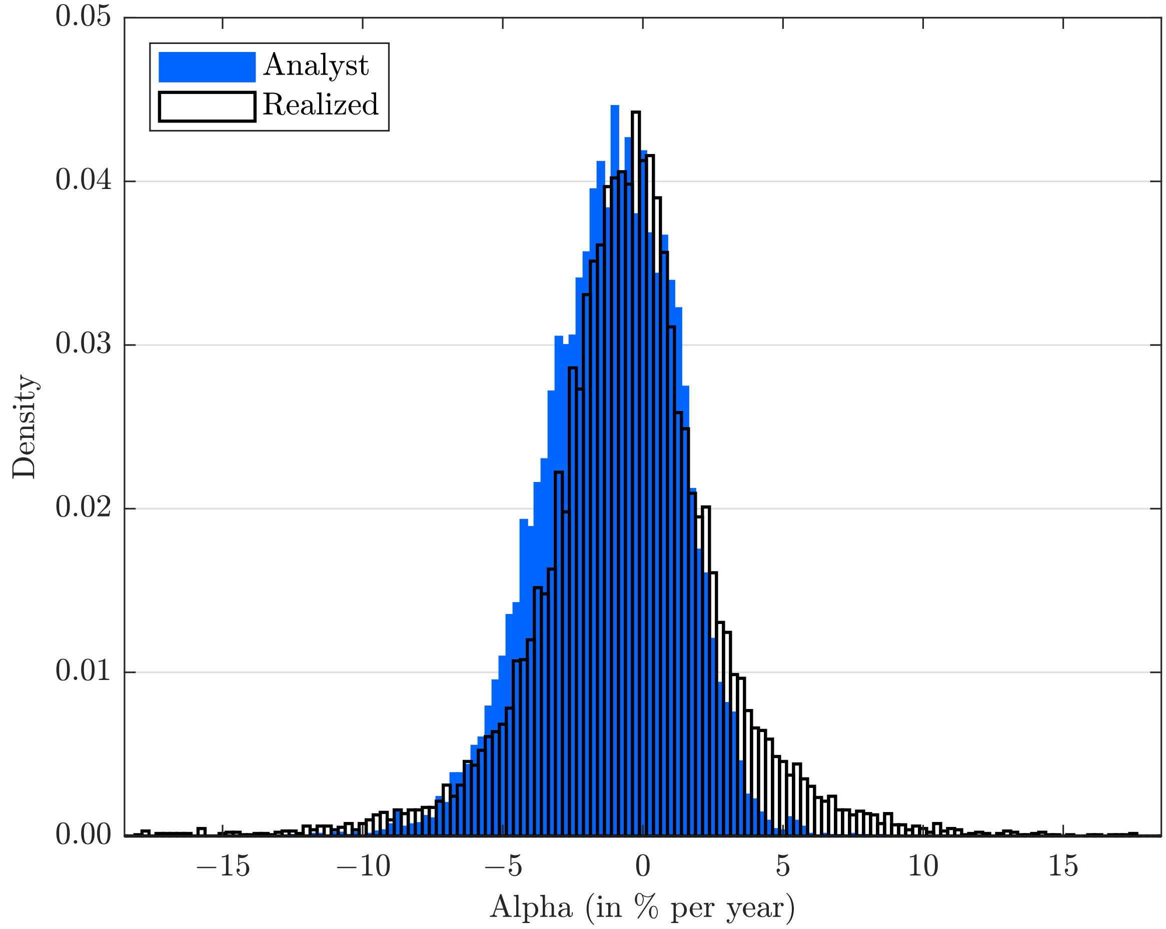
The figure shows the distribution of analysts' expected net-of-fee abnormal returns (alphas) and the distribution of expected net-of-fee abnormal returns implied by a rational model of fund performance as of October 2020. The rows below show value-weighted (by assets under management) and equal-weighted means of alphas, as well as the standard deviations and various percentiles. Alphas are estimated relative to each fund's Morningstar Category benchmark.

Figure 2: Size of active equity mutual fund industry



The figure shows the assets under management (AUM) of actively managed equity mutual funds up to October 2020. New Analyst Rating indicates funds with a Morningstar Analyst Rating according to the new methodology. Old Analyst Rating indicates funds with an Analyst Rating still under the old methodology. Quantitative Rating indicates funds with a Morningstar Quantitative Rating, all of which are rated under the new methodology.

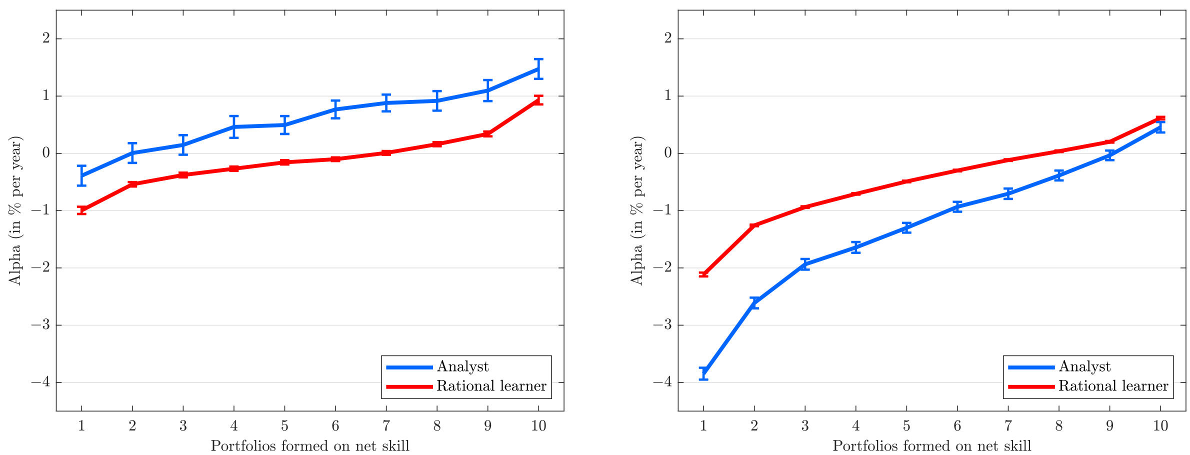
Figure 3: Histogram of analyst alphas and realized alphas



The figure shows the distribution of analysts' expected net-of-fee abnormal returns (alphas) and the distribution of historically realized average net-of-fee abnormal returns as of October 2020. Alphas are estimated relative to each fund's Morningstar Category benchmark; 0.16% of realized alphas lies outside the range shown in the figure.

Figure 4: Alphas against net skill

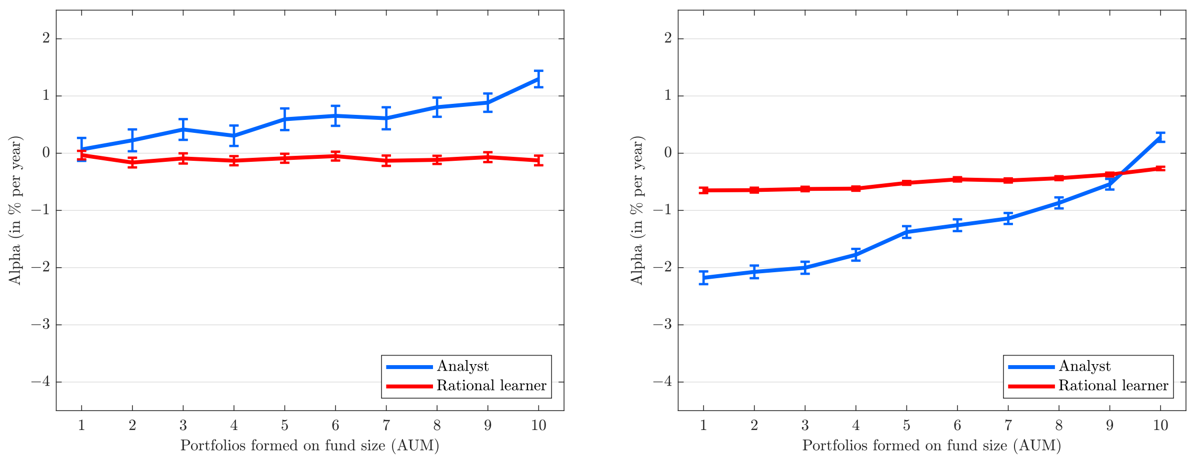
(a) Analyst Ratings **(b)** All ratings



The figure shows expected net-of-fee abnormal returns (alphas) against net skill as of October 2020 for analysts (in blue) and for a rational learner (in red). Net skill is a rational learner's posterior belief about managerial skill less fees. Panel (a) includes funds with an Analyst Rating under the new methodology. Panel (b) includes funds with a new Analyst Rating, an Analyst Rating under the old methodology, and a Quantitative Rating. The bars indicate 90% confidence bands.

Figure 5: Alphas against fund size

(a) Analyst Ratings **(b)** All ratings



The figure shows expected net-of-fee abnormal returns (alphas) against fund size (AUM) as of October 2020 for analysts (in blue) and for a rational learner (in red). Panel (a) includes funds with an Analyst Rating under the new methodology. Panel (b) includes funds with a new Analyst Rating, an Analyst Rating under the old methodology, and a Quantitative Rating. The bars indicate 90% confidence bands.

Online Appendix for

"Expectations of Active Mutual Fund Performance"

Magnus Dahlquist, Markus Ibert, and Felix Wilke

A Data appendix

A.1 Morningstar data

We retrieve the universe of worldwide open-end equity mutual funds from Morningstar Direct. The data belong to 413 Morningstar Categories, which are exclusively designated "Equity" by the Morningstar variable Global Broad Category Group and include live as well as dead funds. We effectively exclude bond funds, money market funds, target date funds as well as other non-equity funds and we ensure that all funds have a Morningstar Category.

The data contain, among other variables, Morningstar's fund and share class identifiers, the Global Category, Morningstar Category, returns, share class net assets, fund sizes, fees, and monthly Morningstar Analyst and Quantitative Ratings. We download the entire time series from January 1965 to October 2020, but benchmark returns are only available from January 1979 and onwards. In total, we collect data for 191744 share classes belonging to 58417 unique funds (as identified by FundId); 43653 funds have at least one non-missing return.

We proceed in two separate steps. First, we describe the data for the replication of Analyst Ratings on the share-class level, which aims to use the data that Morningstar uses. Second, we describe the data for estimating the rational model of fund performance, which aims to use the data that academic research has used previously. In the end, we merge the two data sets to arrive at the final sample for our regressions.

A.2 Replication of Analyst Ratings

The replication of Analyst Ratings follows the three broad steps laid out in the main text:

1. Estimate the semi-interquartile range (SIQR) as a measure of strategy potential for a given group of funds.

2. Construct the before-fee fund alpha based on the SIQR and pillar ratings assigned to individual funds by Morningstar analysts.

3. Subtract share class fees from before-fee fund returns and bin resulting net-of-fee alphas into final ratings.

A.2.1 Gross returns

To estimate historical before-fee fund alphas (Equation (1) in the main text), we use a variable for the gross returns, which is presumably what Morningstar does too, as opposed to adding fees back to net returns.[1] Morningstar uses the fee variable Representative Cost to calculate gross returns from net returns. Hence, using net fund returns and adding back the monthly representative cost should yield similar gross returns.

A.2.2 Benchmark indices

For the benchmark return in Equation (1) in the main text, we use the Morningstar Category Index return of a particular Morningstar Category. Since a fund's Morningstar Category can vary over time, we generally work with the historical Morningstar Category as opposed to the snapshot version and exclude fund-month observations with the Morningstar Category taking values other than the 413 Morningstar Categories that we download (this may happen because historically funds may have belonged to non-equity categories).[2]

A.2.3 Fund strategy potential (SIQR)

Equipped with the time series of before-fee fund returns and benchmark returns, we estimate all active funds' rolling 36-month before-fee alphas from January 2000 forward according to Equation (1) in the main text.

To calculate the SIQR for a particular type of strategy, Morningstar groups funds that invest in the same universe of stocks by aggregating Morningstar Categories from different markets around the world. However, Morningstar is not explicit about the exact mapping of Morningstar Categories into super groups. These super groups are used solely to assess the

[1] We take a value-weighted average of gross share class returns to form the gross fund return. We do this before our cleaning and imputation procedures for AUM since we do not believe analysts employ these procedures. In the data, gross share class returns for a given fund are very similar with slight divergences.

[2] The Morningstar Category is mostly unique among all share classes of a fund, with a few exceptions when a fund's share classes belong to two Morningstar Categories. In all of those cases, one of the two Morningstar Categories is either "EAA Fund Other Equity" or "EAA Fund Property—Indirect Other." Neither category has a designated Morningstar Category Index or a Morningstar Quantitative Rating, but both categories contain some share classes having Morningstar Analyst Ratings. Therefore, we believe it is likely that in Morningstar's process of awarding ratings, all share classes of those funds with a Morningstar Analyst Rating and having two Morningstar Categories are included in the other Morningstar Category we see among the share classes of the respective funds (that is, the category that is not "EAA Fund Other Equity" or "EAA Fund Property—Indirect Other"). We proceed by setting the Morningstar Category variable to equal that other Morningstar Category for all share classes of the fund to be able to correctly replicate the ratings. Picking the Morningstar Category that has most of the fund's AUM leads to the same result for 91% of the funds.

alpha opportunity of fund strategies and the remainder of the rating setting occurs within Morningstar Categories.

We group Morningstar Categories based on the Global Category variable to calculate the SIQR and assign a SIQR to every fund based on its Morningstar Category in October 2020. First, we identify the most common Global Category among all funds within each Morningstar Category. Most funds within a Morningstar Category share the same Global Category. Then, we bundle all Morningstar Categories that have the same most common Global Category. In total, we aggregate funds to 40 different strategies based on 40 Global Categories in our sample.[3] When grouping fund alphas we exclude index funds (as identified by "Index Fund") but keep smart beta funds (as identified by "Strategic Beta") following Morningstar's methodology. Finally, we calculate the semi-interquartile range of the resulting distribution of realized alphas, which reflects Morningstar's assessment of the potential of a given strategy.

A.2.4 Pillar ratings

Morningstar analysts evaluate funds based on three areas that they believe are crucial in order to predict future success: People, Process, and Parent. We noticed that pillar ratings are missing for some share classes of funds that have a Morningstar Analyst or Quantitative Rating. Since pillar ratings are awarded at the fund level, we fill in missing data from other share classes of the same fund.[4] We then calculate the forward-looking before-fee alpha according to Equation (2) in the main text.

A.2.5 Fees

Under the new methodology, Morningstar deducts share-class-specific fees from before-fee alphas to arrive at net alphas and awards Analyst Ratings for each share class. Morningstar uses the fee variable Representative Cost, which contains Morningstar's best estimate of the recurring costs charged by funds.

We notice that fees are still missing for some share classes that have a rating in October 2020. In those cases, we source fees at the end of the sample from other variables to be

[3]For example, the Morningstar Categories "US Fund Large Value" and "EAA Fund US Large-Cap Value Equity" are grouped to form the fund strategy "US Equity Large Cap Value"

[4]Filling in pillar scores allows us to calculate alphas for every share class of a rated fund and to eventually calculate a value-weighted fund-level net alpha reflecting the fee structure of all share classes. However, we do not include alphas of share classes that do not have a Morningstar Analyst or Quantitative Rating in the data when binning net-of-fee alphas into final ratings.

able to replicate as many ratings as possible. In particular, we fill in missing data using the annual report net expense ratio, ongoing cost, prospectus net expense ratio, and the semi annual report net expense ratio, in that order. We set observations smaller or equal to zero to missing in all fee variables that we consider before merging data.

A.3 Data for estimating the rational model of fund performance

Replicating the Analyst Ratings only required a historical time series of before-fee and benchmark returns. To estimate the rational model of fund performance, in addition we need historical data on fund sizes. Before estimating the model, we first clean the data in accordance with the literature (e.g., Pástor, Stambaugh, and Taylor, 2015; Berk and van Binsbergen, 2015).

We start from the monthly gross return dataset, which has 11,661,994 share class/month observations with non-missing returns. Then, we merge in other variables. We merge in only observations of the share class/month when return data exist (in month t or $t+1$). If a variable is missing, we keep the share class/month observation and record a missing value for that variable.

A.3.1 Fees

Since we use gross returns in estimating the model, we do not need additional fee data for the model estimation itself, but will use fees as a filter to drop funds that are unlikely to be actively managed. Our measure of fees is the Morningstar variable "Representative Cost," which is generally populated using a fund's net expense ratio (this can be from the annual report, semi-annual report, or another source) according to Morningstar. At the share-class level, we set fees less than or equal to zero to "missing."

Then, we fill in missing data with the annual report net expense ratio. First, we set the net expense ratio to missing if it is smaller or equal to zero. Next, we place the net expense ratio on the the fiscal year end month if available in Morningstar Direct, and otherwise assume that the fiscal year ends in December. Afterwards, we backward fill missing month ends for up to 12 months (or until the previous reported value) first and then forward fill for up to 12 months. Finally, we use this series to fill in missing monthly fee data.

A.3.2 Cleaning assets under management

Pástor et al. (2015) discover instances of extreme reversal patterns in AUM in the Morningstar data that likely reflect decimal-place mistakes. We adopt their procedure to remove these extreme reversals in monthly fund sizes as well as share class net assets. First, we create a variable for the fractional change in assets from last month to the current month,

$$\%\text{AUM}_t = \frac{\text{AUM}_t - \text{AUM}_{t-1}}{\text{AUM}_{t-1}}. \tag{A1}$$

Second, we create a reversal variable to capture the reversal pattern,

$$\text{Reversal}_t = \frac{\text{AUM}_{t+1} - \text{AUM}_t}{\text{AUM}_t - \text{AUM}_{t-1}}. \tag{A2}$$

This variable will be approximately –1 if it is a reversal (e.g., 20 million, 2 million, 20 million). Finally, if

$$\text{abs}(\%\text{AUM}_t) >= 0.5, -0.75 > \text{Reversal}_t > -1.25, \text{ and } \text{AUM}_{t-1} >= 10 \text{ million}, \tag{A3}$$

then we set assets at time t (i.e., 2 million in this example) to missing. As a result of this procedure, 0.05% of monthly fund size and 0.02% of monthly share class net asset observations are changed to missing.

We use share class net assets when aggregating variables such as returns or fees to the fund level and therefore need monthly asset information. However, there is a significant number of missing asset observations. This is in part due to funds reporting at a quarterly or annual frequency, particularly before 1993. We apply the following procedure to fill in missing monthly share class net assets and fund sizes:

1. We impute missing values in the middle of the data series by using their past values, returns, and a factor adjusted for flow rates as in Ibert, Kaniel, Van Nieuwerburgh, and Vestman (2018). Specifically, let $[t_0, t]$ and $[t+n, T]$ be periods when asset data are non-missing. The missing values are filled in as follows:

$$\text{AUM}_k = F \times \text{AUM}_{k-1}(1 + r_k), \text{ for } k \in [t+1, t+n-1], \tag{A4}$$

$$F = \left(\frac{1}{\prod_{k=t+1}^{t+n}(1 + r_k)} \frac{\text{AUM}_{t+n}}{\text{AUM}_t} \right)^{\frac{1}{n}}, \tag{A5}$$

where F is the factor adjusted for flow rate and r_k is the return. We implement this step allowing for a maximum gap of 12 months between non-missing observations at times t and $t + n$.

2. When returns are not available for all months with missing asset data between times t and $t + n$, we linearly interpolate the missing observations, again allowing for a maximum gap of 12 months.

3. If assets are missing for the last month in the sample, we forward fill the latest available data going back for a maximum of 12 months from the sample end to account for a time lag in reporting.

4. Finally, we set observations where assets are zero or negative to missing.

A.3.3 Aggregation of share-class level to fund level

We take value-weighted averages of returns and fees across share classes using lagged share class assets as weights to form fund-level variables. We take the average across all non-missing share class values and do not set values to missing at the fund level when one or more share classes have missing data. If all share classes have missing assets, we take an equal-weighted average. We treat the fund size variable as AUM on the fund level and use the sum of share class net assets if fund size is missing.

A.3.4 Benchmark indices

A mutual fund's Morningstar Category can evolve over time, for example, due to the fund experiencing style drift (e.g., from US Fund Small Cap Growth to US Fund Small Cap Blend). Therefore, we use the Morningstar Category time series to assign a benchmark return for every fund-month. We forward and backward fill the Morningstar Category for a maximum of 12 months and exclude fund-month observations for which the Morningstar Category takes values other than the 413 Morningstar Categories that we download.

A.3.5 Further sample restrictions

Following Pástor et al. (2015), we exclude fund-month observations with fees below 0.1% per year since it is unlikely that any actively managed fund charges such low fees. In addition, we drop fund-months with fees above 20% per year. Moreover, we drop observations before the fund's inflation-adjusted AUM reached USD 5 million, similar to Berk and van Binsbergen (2015) and Fama and French (2010). We keep only funds with 12 monthly observations in a given year and 12 non-missing returns. When going from fund-month to fund-year, we keep the observation in December of each year. Next, we check whether a given fund has a gap in the annual dataset. If a fund has a missing year, we delete all the fund's observations from the sample.

A.3.6 Identifying index funds

To create a dummy variable to indicate index funds, as in Pástor et al. (2015) we use a simple two-step procedure:

1. If Morningstar indicates a fund to be an index fund (identified by "Index Fund" or "Enhanced Index"), then we classify it as an index fund. Otherwise, we move to the next step.

2. If the fund name contains "Index" or "index," we classify it as an index fund.

Otherwise, we classify the fund as active. As a result of this procedure, we identify and drop 5,251 index funds out of 58,417 funds in total (9.0%).

A.3.7 Inflation adjustment

To make AUM comparable across time, we adjust for inflation using the Consumer Price Index from the Federal Reserve Economic Data provided by the St. Louis Fed (FRED). We use the Consumer Price Index for All Urban Consumers: All Items in U.S. City Average (CPIAUCSL) series and express all USD items in December 2019 USD.

A.4 Aggregation of analyst alphas from share class to fund level

The replication of Analyst Ratings is on the share-class level. We then take a value-weighted average of analyst alphas across share classes to arrive at a fund-level alpha using the cleaned AUM from above.

We take the average across all non-missing share class values and do not set values to missing at the fund level when one or more share classes have missing data. For value-weighting, we use lagged share class net assets. If all share classes have missing assets, we take an equal-weighted average.

A.5 Rational learner alpha

Using the data from Section A.3, we estimate the rational model of fund performance. Since we estimate the model using annual data, we use return data up to December 2019 to estimate a fund's perceived skill. Then, we form rational learner alphas at the end of our sample for every fund according to Equation (7) in the main text using perceived skill measured at the end of 2019 as well as fees and fund sizes measured in October 2020.

The intersection of the fund-level analyst alpha data (Section A.2 and Section A.4) and the data for the model estimation (Section A.3) is the sample for the main regressions in the paper.

A.6 Assigning new ratings to old funds

There are 80 funds with an Analyst Rating under the old methodology that are not yet rated under the new methodology. As we still would like to include them in some of our analyses, we predict what their rating will be once updated. We have all the required inputs except for the new individual pillar scores. Under the old methodology, individual pillar scores ranged from "Negative" via "Neutral" to "Positive." We assume that these three verbal expressions correspond to pillar scores of -1, 0, and $+1$, respectively. Then, for each of the three pillars, we translate the scoring scale into the new scoring scale from -2 to $+2$ by i) regressing the new pillar ratings on a set of characteristics for the sample of updated funds, and ii) using the fitted values from these regressions to predict the pillar score for a not-yet-updated fund:

$$\text{PillarScore}_i = \gamma_0 + \gamma' X_i + \psi_i, \tag{A6}$$

where the vector of characteristics, X_i, includes a fund's old pillar rating, its old Morningstar Analyst Rating, and its annual fee.[5] The adjusted R^2 values in these regressions range from 58% to 75%.

[5] This is similar to the process that Morningstar recommends for predicting the new ratings for not-yet-updated funds: "For instance, if we run a fund through the updated methodology and that fund sits in the same peer group; has similar People, Process, and Parent Pillar ratings; and sports a similar expense ratio to a fund that hasn't gone through yet, then the peer fund's Analyst Rating can offer clues into how that fund will eventually be rated under the new methodology" (Ptak, 2019).

B Fund flows and Analyst Ratings (2011–2020)

Table B1 shows regressions of monthly fund flows,

$$\text{Flow}(\%) = \frac{\text{AUM}_{i,t} - \text{AUM}_{i,t-1} \times (1 + R_{i,t})}{\text{AUM}_{i,t-1} \times (1 + R_{i,t})} \times 100, \tag{B1}$$

on Morningstar Analyst Ratings, Star Ratings, and various control variables. In specification (1), a fund with a Gold rating receives 0.445-percentage-point larger flows than does a fund with a Neutral rating in a given year-month. In (2), a five-star fund receives 1.080-percentage-point larger flows than does a fund with a three-star rating. Specification (3) shows that the effect of Analyst Ratings on flows weakens once the Star Rating is included. Nevertheless, Gold, Silver, and Bronze funds still attract significantly more flows than do funds with a Neutral Analyst Rating. A Gold fund receives 0.13-percentage-point larger flows than does a Neutral fund with the same Star Rating.

Specifications (4)–(6) repeat specifications (1)–(3) but include funds with a Quantitative Rating—which in (1)–(3) enter the "Unrated" group since these funds do not have an Analyst Rating—in the Gold, Silver, Bronze, Neutral, and Negative groups. The sample starts in 2017, which is when the Quantitative Ratings were first introduced.

C Decreasing returns to scale

Table C1 shows regressions of fund abnormal returns on fund size using the OLS estimator, the fund fixed effects estimator, the recursive demeaning estimator in Pástor et al. (2015) (RD1), and the recursive demeaning estimator in Zhu (2018) (RD2). All specifications using the preferred RD2 estimator show a significantly negative impact of fund size on fund returns.

When we restrict the sample to U.S. funds and to 1995–2014, the sample period in Zhu (2018), we obtain estimates very close to hers despite calculating abnormal returns slightly differently.[6] For example, in untabulated results in the regressions using monthly data and the logarithm of AUM, the coefficient estimates become -0.13 for the fund fixed effects estimator and -0.21 for the RD2 estimator compared with -0.15 and -0.26, respectively, in Zhu (2018).

D Rational expectations model with indexing

In an extension of their model, Berk and Green (2004) capture the idea that more active funds are subject to steeper decreasing returns to scale by allowing funds to index part of their assets, that is, to directly invest into the passive benchmark. Investors still pay the fee on this part, but since it is not actively managed it does not affect returns through the cost function c in Equation (3) in the main text.

If active funds are allowed to index part of their assets, following Equation (11) from Berk and Green (2004), the measurement equation becomes

$$r_{i,t+1} + f_{i,t} = h_{i,t}\alpha_{i,t} - c(h_{i,t}AUM_{i,t}) + h_{i,t}\epsilon_{i,t+1}, \tag{D1}$$

where $h_{i,t}$ refers to a fund's fraction of assets that are actively managed. The state transition equation is the same as before, that is, Equation (4) in the main text. The updating equations

[6] Both Zhu (2018) and Pástor et al. (2015) calculate abnormal returns as simply the difference between the fund return and the benchmark return without adjusting for different exposures to the benchmark.

Table B1: Fund flows on Analyst Ratings

	Analyst Ratings 2011–2020			All Ratings 2017–2020		
	(1)	(2)	(3)	(4)	(5)	(6)
Gold	0.445***		0.130**	0.382***		0.102**
	(0.049)		(0.055)	(0.046)		(0.045)
Silver	0.280***		0.054	0.249***		0.046
	(0.042)		(0.042)	(0.036)		(0.043)
Bronze	0.225***		0.095**	0.156***		0.037
	(0.040)		(0.039)	(0.027)		(0.025)
Neutral						
Negative	−0.156		0.019	−0.050		0.082*
	(0.143)		(0.148)	(0.038)		(0.043)
Unrated	0.037		0.001	0.012		−0.054
	(0.040)		(0.039)	(0.089)		(0.091)
Five-star		1.080***	0.668***		1.025***	0.224
		(0.029)	(0.189)		(0.059)	(0.181)
Four-star		0.334***	−0.074		0.280***	−0.517***
		(0.017)	(0.193)		(0.031)	(0.178)
Three-star			−0.404**			−0.792***
			(0.194)			(0.167)
Two-star		−0.189***	−0.593***		−0.125***	−0.926***
		(0.022)	(0.192)		(0.018)	(0.169)
One-star		−0.355***	−0.759***		−0.366***	−1.182***
		(0.032)	(0.188)		(0.047)	(0.146)
No-star		0.438***	0.032		0.551***	−0.239
		(0.048)	(0.189)		(0.072)	(0.178)
N	1331942	1331942	1331942	503428	503428	503428
Adj. R^2	0.09	0.09	0.09	0.09	0.10	0.10
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes

The table shows regressions of monthly equity mutual fund flows on Morningstar Analyst, Quantitative, and Star Ratings up to October 2020. Specifications (1)–(3) include Analyst Ratings, whereas (4)–(6) additionally include Quantitative Ratings in the Gold, Silver, Bronze, Neutral, and Negative ratings. The controls include the logarithm of AUM and fund family AUM (in millions of USD), fund age (logarithm of number of months since fund inception), fees, past 12-month fund returns, 12-month volatility of fund returns, 12-month average fund flows, and maximum manager tenure. Standard errors are calculated using the spatial estimator of Driscoll and Kraay (1998), which allows for both cross-sectional and serial correlation up to four lags in the errors as well as heteroskedasticity in the errors. ***, **, and * denote significance at 1%, 5%, and 10% levels, respectively.

Table C1: Decreasing returns to scale

	U.S. sample				All fund sample			
	(1) OLS	(2) FE	(3) RD1	(4) RD2	(5) OLS	(6) FE	(7) RD1	(8) RD2
Panel A: Monthly data, Logarithm of fund AUM								
Size (× 100)	−0.000	−0.100***	−0.444***	−0.182***	0.014***	−0.102***	−0.441**	−0.088***
	(0.002)	(0.006)	(0.140)	(0.020)	(0.002)	(0.004)	(0.204)	(0.029)
N	630587	630587	630587	630587	2968377	2968377	2968377	2968377
Panel B: Monthly data, Dollar fund AUM								
Size (× 10^6)	0.002	−0.067***	−1.473	−1.256***	0.020***	−0.089***	−0.653	−0.958***
	(0.005)	(0.009)	(2.109)	(0.223)	(0.005)	(0.008)	(1.088)	(0.371)
N	630587	630587	630587	630587	2968377	2968377	2968377	2968377
Panel C: Annual data, Logarithm of fund AUM								
Size (× 100)	−0.168***	−1.322***	−1.874	−1.735***	0.043	−1.649***	2.255	−2.664***
	(0.040)	(0.095)	(1.499)	(0.273)	(0.030)	(0.061)	(2.315)	(0.339)
N	40268	40268	40268	40268	162124	162124	162124	162124
Panel D: Annual data, Dollar fund AUM								
Size (× 10^6)	−0.085	−0.825***	0.726	−8.814***	0.104	−1.232***	−1.059	−21.686***
	(0.067)	(0.107)	(3.016)	(2.314)	(0.065)	(0.117)	(3.036)	(4.552)
N	40268	40268	40268	40268	162124	162124	162124	162124

The table shows coefficient estimates on lagged fund size in regressions of net abnormal fund returns on lagged fund size in an unbalanced panel from 1979 to 2020. FE refers to the estimator that includes fund fixed effects. RD1 refers to the recursive demeaning estimator of Pástor et al. (2015), which recursively forward-demeans all variables and instruments for forward-demeaned fund size using backward-demeaned fund size while imposing a zero intercept in the first stage. RD2 refers to the recursive demeaning estimator of Zhu (2018), which instead instruments for forward-demeaned fund size using total fund size and includes an intercept in the first-stage regression. The U.S. sample of funds includes funds from the following nine Morningstar Categories: U.S. Fund Large Growth, U.S. Fund Large Blend, U.S. Fund Large Value, U.S. Fund Small Growth, U.S. Fund Small Blend, U.S. Fund Small Value, U.S. Fund Mid-Cap Growth, U.S. Fund Mid-Cap Blend, and U.S. Fund Mid-Cap Value. Fund size is the fund's total assets under management (AUM) at the end of the previous period expressed in millions of December 2019 USD. Standard errors are presented in parentheses and clustered by Morningstar Category times year-month or Morningstar Category times year. Standard errors are additionally clustered by fund in the RD specifications. ***, **, and * denote significance at 1%, 5%, and 10% levels, respectively.

become

$$\widehat{a}_{i,t+1} = \rho \left(\widehat{a}_{i,t} + \frac{\widehat{\sigma}^2_{a,t}}{h_{i,t}(\widehat{\sigma}^2_{a,t} + \sigma^2_{\epsilon})} (r_{i,t+1} - h_{i,t}\widehat{a}_{i,t} + c(h_{i,t}\text{AUM}_{i,t}) + f_{i,t}) \right) + (1-\rho)a_0, \quad \text{(D2)}$$

$$\widehat{\sigma}^2_{a,t+1} = \rho^2 \widehat{\sigma}^2_{a,t} \left(1 - \frac{\widehat{\sigma}^2_{a,t}}{\widehat{\sigma}^2_{a,t} + \sigma^2_{\epsilon}} \right) + (1-\rho^2)\sigma^2_{a,0}. \quad \text{(D3)}$$

Our previous cost function, c, was the logarithmic function. Theoretically, a fund could index all of its assets so that the log of actively managed assets is undefined. Therefore, we choose a more flexible form for the impact of scale on returns: $c(hAUM) = \eta \frac{(hAUM)^{\gamma}-1}{\gamma}$ with $\gamma \in (0, 1]$ (as in Roussanov, Ruan, and Wei, 2020). If $\gamma = 1$, the cost function is linear in active size. As γ approaches zero, the cost function converges to the logarithmic function.

We estimate a fund's three-year rolling-window R^2 relative to the benchmark, and compute the active share as $1 - R^2$ (Amihud and Goyenko, 2013). We estimate the model using maximum likelihood, recalculate perceived skill using Equation (D2) at the end of our sample, and reproduce Tables 5 and 6 in the main text using the corresponding variables from the measurement equation, Equation (D1). We winsorize the R^2 values at the 1st and 99th percentiles to estimate the model using data from 1979–2019, and to be consistent we also winsorize the R^2 values at the end of our sample in October 2020 (used to calculate active perceived skill and active fund size) for our cross-sectional regressions.

Table D1 shows the parameter estimates. To compare these estimates to those in Table 4 in the main text, the parameter estimates for the prior mean, the prior standard deviation, and the residual standard deviation need to be multiplied by the active share. The average active share in the data is 13%. The estimate close to zero for γ suggests that the log functional form of the cost function fits the data very well. As before, we find a parameter estimate, η, that is significantly positive, indicating decreasing returns to scale in actual fund returns.

Table D2 reproduces Table 5 based on Equation (D1). If the rational expectations model was the model analysts use to form their expectations, the coefficient estimates should be 1 on active share times perceived skill, $-\eta$ on active fund size, and -1 on fees. As before, the coefficient estimates on scale, this time measured as actively managed size, are significantly positive and significantly different from the model-implied coefficient estimate of -0.11.

Table D3 reproduces Table 6. The coefficient estimates on actively managed size are significantly positive in all specifications.

Table D1: Parameter estimates of the rational fund performance model with indexing

Parameter	Description	Estimate
η	Decreasing returns to scale (DRS) (%)	0.110***
		(0.005)
γ	Shape of DRS ($\times$ 10^6)	0.001
		(0.003)
a_0	Prior mean (%)	18.268***
		(0.387)
$\sigma_{a,0}$	Prior standard deviation (%)	43.903***
		(0.643)
σ_ϵ	Residual standard deviation (%)	115.067***
		(0.286)
ρ	Skill persistence	0.836***
		(0.009)

The table shows the parameter estimates of the rational fund performance model with indexing in % per year. Standard errors are in parentheses. The model is estimated using fund-year observations from 1979–2019. *, **, and *** denote 10%, 5%, and 1% significance levels, respectively, for the null of a zero coefficient.

Table D2: Cross-sectional regressions of alphas on characteristics with indexing

	Analyst Ratings	All ratings
	(1)	(2)
Perceived skill $\times$ h	0.201***	0.136***
	(0.034)	(0.018)
	[0.000]	[0.000]
Active fund size ($\times$ 100)	0.109***	0.185***
	(0.019)	(0.011)
	[0.000]	[0.000]
Fees	−0.816***	−1.454***
	(0.109)	(0.033)
	[0.089]	[0.000]
Constant ($\times$ 100)	0.714***	0.211***
	(0.143)	(0.055)
	[0.000]	[0.000]
N	1415	13185
Adj. R^2	0.13	0.23

The table shows regressions of Morningstar analyst fund alphas on skill as perceived by a rational learner, active fund size, and fees following Equation (D1). Active fund size is measured as $\frac{(h\text{AUM})^{\gamma}-1}{\gamma}$, where AUM refers to a fund's total assets under management, h refers to a fund's active share, and γ is given in Table D1. Specification (1) uses funds with an Analyst Rating under the new methodology. Specification (2) uses funds with a new Analyst Rating, an Analyst Rating under the old methodology, and a Quantitative Rating. Standard errors are heteroskedasticity robust and in parentheses. *, **, and *** denote 10%, 5%, and 1% significance levels, respectively, for the null of a zero coefficient. In brackets are p-values for the null hypothesis that the coefficients of skill, size, fees, and the constant are equal to the model-predicted parameters of +1, −0.110 (the estimate of η in Table D1), −1, and 0, respectively.

Table D3: Cross-sectional regressions of alphas on additional characteristics with indexing

	Analyst Ratings			All ratings		
	(1)	(2)	(3)	(4)	(5)	(6)
Rational learner						
Perceived skill × h	0.057	0.156***	0.107**	0.160**	0.259**	0.086*
	(0.043)	(0.044)	(0.045)	(0.067)	(0.102)	(0.049)
Active fund size (× 100)	0.224***	0.148***	0.096***	0.273***	0.169***	0.178***
	(0.031)	(0.033)	(0.035)	(0.026)	(0.030)	(0.023)
Fees	−1.385***	−1.245***	−1.196***	−1.751***	−1.786***	−1.221***
	(0.181)	(0.201)	(0.250)	(0.172)	(0.168)	(0.147)
People						
Manager tenure		0.067*	0.129***		0.323***	0.316***
		(0.039)	(0.035)		(0.037)	(0.034)
Manager ownership		0.143***	0.138***		0.339***	0.208***
		(0.041)	(0.042)		(0.039)	(0.039)
Managerial multitasking		0.337***	0.762***		−0.115	−0.005
		(0.092)	(0.177)		(0.081)	(0.138)
Management team		0.035	0.058		0.159***	0.166***
		(0.054)	(0.056)		(0.048)	(0.046)
Process						
Top 10 assets (%)		−0.223***	0.120		−0.168**	0.140*
		(0.078)	(0.103)		(0.072)	(0.075)
Tracking error		−0.132*	−0.095		−0.209**	−0.164**
		(0.070)	(0.099)		(0.093)	(0.072)
Turnover ratio		−0.579***	−0.648***		−0.236***	−0.106
		(0.128)	(0.143)		(0.087)	(0.072)
N	691	691	642	2706	2706	2507
Adj. R^2	0.22	0.28	0.59	0.19	0.27	0.61
Morningstar Category FE	No	No	Yes	No	No	Yes
Fund Family FE	No	No	Yes	No	No	Yes

The table shows regressions of Morningstar analyst alphas on fund and manager characteristics. Specifications (1)–(3) use U.S. funds with an Analyst Rating under the new methodology. Specifications (4)–(6) use U.S. funds with a new Analyst Rating, an Analyst Rating under the old methodology, and a Quantitative Rating. Manager tenure is the maximum tenure (in months) taken over all managers, manager ownership is the average amount managers at a fund personally invest in the fund, managerial multitasking is the average number of additional funds managers of a particular fund manage, and management team is a dummy for team-managed funds. "People" and "Process" variables are standardized to zero mean and unit variance, and the coefficient estimates are multiplied by 100. Standard errors are heteroskedasticity robust and in parentheses. *, **, and *** denote 10%, 5%, and 1% significance levels, respectively, for the null of a zero coefficient.

References

Amihud, Yakov, and Ruslan Goyenko, 2013, Mutual Fund's R2 As Predictor of Performance, *Review of Financial Studies* 26, 667–694.

Berk, Jonathan B., and Richard C. Green, 2004, Mutual Fund Flows and Performance in Rational Markets, *Journal of Political Economy* 112, 1269–1295.

Berk, Jonathan B., and Jules H. van Binsbergen, 2015, Measuring Skill in the Mutual Fund Industry, *Journal of Financial Economics* 118, 1–20.

Driscoll, John, and Aart Kraay, 1998, Consistent Covariance Matrix Estimation With Spatially Dependent Panel Data, *Review of Economics and Statistics* 80, 549–560.

Fama, Eugene F., and Kenneth R. French, 2010, Luck versus Skill in the Cross-Section of Mutual Fund Returns, *Journal of Finance* 65, 1915–1947.

Ibert, Markus, Ron Kaniel, Stijn Van Nieuwerburgh, and Roine Vestman, 2018, Are Mutual Fund Managers Paid for Investment Skill?, *Review of Financial Studies* 31, 715–772.

Pástor, Luboš, Robert F. Stambaugh, and Lucian A. Taylor, 2015, Scale and Skill in Active Management, *Journal of Financial Economics* 116, 23–45.

Ptak, Jeffrey, 2019, Introducing the Enhanced Morningstar Analyst Rating for Funds.

Roussanov, Nikolai, Hongxun Ruan, and Yanhao Wei, 2020, Mutual Fund Flows and Performance in (Imperfectly) Rational Markets?, *Working Paper* .

Zhu, Min, 2018, Informative Fund Size, Managerial Skill, and Investor Rationality, *Journal of Financial Economics* 130, 114–134.