Expectations of Mutual Fund Performance

Magnus Dahlquist, Markus Ibert, and Felix Wilke*

First draft: February 1, 2020

This draft: June 4, 2020

Abstract

We use forward-looking Morningstar Analyst Ratings to infer a distribution of expected abnormal returns (alphas) for mutual funds. Professional analysts believe that alphas are dispersed and that the largest funds will perform best: the value- and equal-weighted industry alphas are 0.70% and -1.21% per year, respectively, going forward. We estimate a rational learning model of fund performance and benchmark analysts' expectations against the model-implied expectations based on fund size, perceived skill, and fees. Analysts and the rational learner respond similarly to changes in perceived skill and fees, but in contrast to analysts the rational learner believes in decreasing returns to scale. Additional manager and fund family characteristics beyond size, perceived skill, and fees are important to analysts' expectations formation.

JEL: G11, G12, G14, G23.

Keywords: Alpha, expectations 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 for comments and suggestions.

1 Introduction

Expectations are paramount to economics. Among other things, expectations determine how much agents save and consume, what fraction of their wealth they invest in the stock market, and how much capital they allocate to active mutual funds. Yet, for a long time research on how expectations are formed has been limited. Under the rational expectations paradigm, expectations are implied by the model and can be recovered empirically by large-sample distributions of the underlying moments. By now, ample evidence against the full-information rational expectations (FIRE) hypothesis exists, making the study of how expectations are formed a fruitful exercise (Coibion and Gorodnichenko, 2012, 2015).

We infer forward-looking expected net-of-fee abnormal returns, henceforth "alphas," for active equity mutual funds from Morningstar Analyst Ratings to study how expectations are formed. We have three main findings: 1. the distribution of analyst alphas is markedly dispersed; 2. analyst alphas respond to perceived managerial skill and fees but not to fund size as predicted by a rational expectations model.; and 3. beyond perceived skill, size, and fees, additional manager and fund family characteristics are important for formation of analysts' expectations.

Mutual funds provide an excellent laboratory to study expectations formation for at least three reasons. First, there exist databases similar in quality to equity price databases that provide returns, prices (fees), and quantities (assets under management, AUM). Second, rational models of mutual fund performance have clear predictions on how expectations are formed (Berk and Green, 2004; Pástor and Stambaugh, 2012), thereby providing invaluable benchmarks. These models are noisy rational expectations (NRE) equilibrium models in which agents are uncertain about some of the underlying parameters in the economy (e.g., managerial skill) and update their beliefs from observed fund returns. Third, the representativeness of analysts' expectations for the broader set of investors can be directly tested by examining fund flows. This is typically not the case for other analyst forecasts or surveys, such as earnings forecasts (e.g., La Porta, 1996; Bordalo et al., 2019) or inflation forecasts (e.g., Malmendier and Nagel, 2016).[1]

Morningstar provides Analyst Ratings since 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, a Neutral rating, and a Negative rating. Under the new methodology, Morningstar estimates a distribution of alphas and then groups alphas (which are not reported in the database) to arrive at the final Analyst Ratings (which are reported in the database). We replicate Morningstar's methodology to infer the alphas that Morningstar analysts use. When we translate our alphas into the ratings we can replicate around 90% of Morningstar's ratings.

First, Figure 1 illustrates the dispersion of analyst alphas (blue, solid bars). This dispersion is interesting because NRE equilibrium models with risk-neutral investors commonly imply that alphas are zero at all points in time (e.g., Berk and Green, 2004). Through the lens of these models, these beliefs imply that most funds do not run at their equilibrium sizes and that there are severe capital misallocations among active equity mutual funds: The funds in our sample manage around 10 trillion USD and account for approximately 10% of global equity market capitalization. Professional analysts expect the majority of funds to underperform but they expect the largest funds to outperform. This is readily apparent in Figure 1 when comparing the value-weighted (0.70%) mean to the equal-weighted mean (−1.21%) and median (−1.05%).

Second, we contrast analyst alphas with alphas in a Berk and Green (2004) rational expectations learning model, but, following Roussanov et al. (2018), we relax the equilibrium

[1] One exception is surveys of expected stock market returns, which typically evaluate representativeness by examining flows into the aggregate of equity mutual funds (e.g., Greenwood and Shleifer, 2014). Nevertheless, examining fund flows on the individual fund level allows for a refined test.

implication of zero alphas. Figure 1 also overlays the model-implied distribution of alphas (red, transparent bars). The rational learner expects abnormal returns to be much more concentrated than analysts. To systematically investigate differences between the rational learner and analysts, we relate alphas to the fund characteristics in our 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. In one of our main specifications, analyst alphas respond one-to-one to perceived skill as predicted by the model, and a one-percentage-point increase in fees is associated with a -1.36 percentage points lower alpha, which is slightly larger than the one-to-one pass through predicted by the model. When we decompose perceived skill into a prior and past returns, we find that analysts overreact to at least the last six years of returns and that they underreact to the prior. While the model's predictions related to perceived skill and fees are broadly consistent with the data, the prediction related to size is not. Controlling for perceived skill and fees, we find that a one-unit increase in log size is associated with a 0.11 percentage points *larger* alpha. This stands in stark contrast to the model's prediction of a 0.18 *lower* alpha per one-unit increase in log size, but is reminiscent of studies reporting that investors do not believe active mutual funds suffer from decreasing returns to scale (Choi and Robertson, 2019).

Third, the R^2s in our main regressions range from 14% to 32%, leaving ample room for other determinants of analysts' expectations. Once we include additional manager and fund characteristics, we find a positive impact of managers' personal investments in their funds ("skin in the game") and manager tenure on analyst alphas. Consistently, the academic literature has shown a positive relationship between both personal investments (Khorana et al., 2007; Evans, 2008; Ibert, 2019) and experience (Greenwood and Nagel, 2009), respectively, and fund performance. Nonetheless, fees and perceived skill still matter most and have the largest standardized coefficient estimates. The inclusion of fund family fixed effects results in R^2s of around 60% suggesting that fund family characteristics play an important role in analysts' expectation formation. For instance, fund manager compensation practices are likely important and have been shown to systematically differ across fund families (Ibert et al., 2018; Ma et al., 2019). More generally, the literature on the role of the fund family has highlighted a fund family's impact on individual fund performance (Massa, 2003; Gaspar et al., 2006; Pool et al., 2016; Berk et al., 2017; Ferreira et al., 2018).

Overall, some of our results are remarkably consistent with the rational expectations model. The model predicts that perceived skill, size, and fees matter exclusively, and indeed we find that fees and perceived skill matter most. At the same time, the absence of decreasing returns to scale in analysts' beliefs is an important deviation given the crucial role decreasing returns to scale play in our understanding of how mutual funds equilibriate and of how to measure managerial skill (Berk and van Binsbergen, 2015, 2017). Similar to the survey evidence on stock returns in Greenwood and Shleifer (2014), the expectations we uncover can hardly be the expectations of a representative mutual fund investor in a rational expectations equilibrium. Apart from the absence of decreasing returns to scale in beliefs, this can be easily seen in Figure 1 in which some funds have a negative alpha. At the same time, Armstrong et al. (2019) do find that fund flows respond to Analyst Ratings—even when the popular backward-looking Star Rating is controlled for—suggesting that analysts' expectations are at least partly representative for the broader set of mutual fund investors. We extend their study on flows to international funds and up to 2020, and find similar results. The results on 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 et al., 2019) and other recommendations (e.g., Reuter and Zitzewitz, 2006; Kaniel and Parham, 2018).

Our paper is related to several strands of research. By providing novel evidence from a large market of vast importance to households, institutional investors, and pension systems we contribute to the literature on expectations in macroeconomics and finance (see, e.g., Fuster et al., 2010; Coibion and Gorodnichenko, 2012; Gennaioli et al., 2015). Departures from the null of FIRE can be attributed to departures of the full-information or the rational expectations assumption. The behavioral finance arm of this literature has focused on the latter, and often documents extrapolation and excessive overreaction (see, e.g., Greenwood and Shleifer, 2014). Coincidentally, early evidence on extrapolation comes from the mutual fund literature (Chevalier and Ellison, 1997; Sirri and Tufano, 1998). Mutual fund investors' return chasing has been interpreted as irrational behavior given the inability of the average fund to outperform a passive benchmark. Berk and Green (2004) rationalize return chasing and the absence of outperformance, but we show that analysts' expectations in general respond too much to recent past performance relative to Berk and Green's benchmark. While we do not interpret our evidence as evidence for analyst irrationality, our results are reminiscent of recent evidence for overreaction in macroeconomics and finance (see, e.g., Bordalo et al., 2020). Landier et al. (2020) show that overreaction is particularly pronounced for less persistent processes, which is consistent with overreaction to past fund performance.

We also contribute to the literature on professional forecasts. The finance literature has studied professional forecasts of corporate earnings (see, e.g., Bouchaud et al., 2019), credit spreads (see, e.g., Greenwood and Hanson, 2013; Bordalo et al., 2019), short-term interest rates (Cieslak, 2018), and stock market returns (see, e.g., Ben-David et al., 2013). We hope that our new moments on professional forecasts for mutual fund alphas can guide the development of theoretical models, similar to how survey evidence on stock returns has guided the development of a new generation of asset pricing models (Barberis et al., 2015; Adam et al., 2017; Barberis et al., 2018; Nagel and Xu, 2019).

Finally, we contribute to the mutual fund literature. Despite expectations being paramount to understanding capital allocations to active mutual funds, the mutual fund literature has been silent about empirical measures of investors' expectations. Similar in spirit to Berk and Green (2004), the literature has investigated which prior beliefs in a Bayesian setting can rationalize the large capital allocations to mutual funds despite poor past performance (see, e.g., Baks et al., 2001; Pástor and Stambaugh, 2002; Busse and Irvine, 2006). Other researchers back out investors' preferences using fund flows (e.g., Berk and van Binsbergen, 2016; Barber et al., 2016). We are the first to provide explicit forecasts for individual funds and the first to relate these forecasts to an equilibrium model of fund performance. Relatedly, Jones and Martinez (2017) study the future performance expectations of U.S. plan sponsors that rank their asset managers on a scale from one to five. Consistent with our results, they find that both past performance and soft factors matter for expectations. Similarly, Jenkinson et al. (2016) highlight the importance of soft factors for U.S. investment consultants' 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 funds outperform, but they cannot recover alphas as Morningstar's methodology was presumably not publicly known when their study was published.[2]

The paper proceeds as follows. Section 2 describes how Analyst Ratings are constructed. Section 3 describes the data. Section 4 outlines and estimates the model. Section 5 contains the main empirical results. Section 6 concludes.

2 Forward-looking Morningstar Analyst Ratings

2.1 Old and new ratings

Morningstar provides Analyst Ratings for a selected number of funds since 2011. Unlike the backward-looking Morningstar Rating (often referred to as the "Star Rating"), the

[2]Moreover, the 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.

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, a Neutral rating, and a Negative rating. Morningstar is an independent research firm and its Analyst Ratings likely reflect unbiased opinions. Armstrong et al. (2019) and Cookson et al. (2019) discuss Morningstar analyst unbiasedness and offer additional details for the process behind the ratings. Up until October 2019, the Analyst Rating was based on the analyst's conviction in 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. The most important changes were a larger emphasis on fees and a share class specific rating in contrast to a fund level rating. 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 (i.e., a Bronze, Silver or Gold rating). The new rating is, hence, informative about alpha as alpha measures the performance relative to a passive benchmark. Morningstar has applied the new methodology to more than 800 active equity mutual funds as of April 2020, and it expects to update the remaining funds until October 2020. 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 available in the database) into the aforementioned ratings (which are available in the database).

Since 2017 Morningstar also provides forward-looking Quantitative Ratings; they are similar to Analyst Ratings, but 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 analysis. All funds with a Quantitative Rating are rated under the new methodology as of October 2019. Table 1 provides a summary of the different Morningstar ratings.

2.2 Analyst and Quantitative Ratings methodology

Morningstar's exact methodology to construct 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) fund return, $R_{f,t}$ 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) to form a distribution of realized alphas. Morningstar then calculates the semi-interquartile range (SIQR) of the distribution (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.[3]

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.[4] 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 include an anonymized example of such a report in Appendix A. 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 an

[3] Harvey and Liu (2019) use the interquartile range as a measure of cross-sectional return dispersion and show that fund flows react more strongly to performance in periods of low return dispersion.

[4] 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.

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 denotes analysts' subjective expectations, r is the fund's net abnormal return, and f 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 together 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 which groups Morningstar Categories from different domiciled funds), use a fund's Morningstar Category benchmark as the benchmark, and rely on Kenneth French's data for the risk-free rate. Appendix B contains additional details about our replication and the data.

2.3 Assigning new ratings to old funds

Around 600 funds with an Analyst Rating under the old methodology are not yet rated under the new methodology. As we still would like to include them in some of our analysis, 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:

$$\text{PillarScore}_i = \gamma_0 + \gamma' X_i + \psi_i, \tag{3}$$

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^2s in these regressions range from 59% to 82%. We have tried other ways to impute the pillar scores for not yet updated funds and the results appear insensitive to the specific assumptions.

2.4 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 recover the alphas that Morningstar uses to construct the ratings. Panel A shows that for the 5,461 share classes with an Analyst Rating under the new methodology, Morningstar assigns a Neutral rating to 2,197 share classes. In this case, we assign a Neutral rating in 2,036 cases yielding a replication rate of 93%. Our overall replication rate for the new Analyst Ratings is 88%. Panel B shows our replication for the Morningstar Quantitative Ratings. Our overall replication rate for Quantitative Ratings is 93%.

[5] This is similar to the process that Morningstar recommends to predict 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).

3 Data

We obtain gross returns, AUM, ratings, and fees for active open-end equity mutual funds from Morningstar Direct. We focus on global funds to correctly replicate Morningstar's methodology, that is, the sample contains both U.S. domiciled funds and non-U.S. domiciled funds. Although Morningstar only uses data as of January 2000 to construct the Analyst Ratings, we obtain the entire time-series available in Morningstar to estimate our model of mutual fund performance. The monthly sample starts in January 1979, the first month in which Morningstar provides benchmark returns, and ends in April 2020. We convert all returns and assets to USD. 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. 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 a AUM-weighted average.

Figure 2 plots the AUM of funds with an Analyst Rating under the new methodology, funds with an Analyst Rating under the old methodology, funds with a Quantitative Rating (all of them are rated under the new methodology), and funds with no rating. Table 3 shows summary statistics for the different sub-samples. The number of funds with a Quantitative Rating is large but their assets on average are much smaller. Moreover, the table shows that funds with an Analyst Rating, old or new, have much larger analyst alphas and larger perceived skill (a sufficient statistic for past performance adjusted for decreasing returns to scale and introduced below).

4 Rational benchmark 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 mutual fund performance. Similar to Berk and Green (2004) we model the abnormal return for a given fund as:

$$r_{i,t+1} + f_{i,t} = a_{i,t} - c(\mathrm{AUM}_{i,t}) + \epsilon_{i,t+1}, \tag{4}$$

where $\epsilon \sim N(0, \sigma_\epsilon^2)$, r is the fund's net abnormal return, a is unobservable managerial skill, f fees, and $c(\mathrm{AUM})$ captures decreasing returns to scale.[6]

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

$$a_{i,t} = (1-\rho)a_0 + \rho a_{i,t-1} + \sqrt{1-\rho^2} \cdot \nu_{i,t}, \tag{5}$$

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

$$\widehat{a}_{i,t+1} = \rho\left(\widehat{a}_{i,t} + \frac{\widehat{\sigma}_{a,t}^2}{\widehat{\sigma}_{a,t}^2 + \sigma_\epsilon^2}(r_{i,t+1} - \widehat{a}_{i,t} + c(\mathrm{AUM}_{i,t}) + f_{i,t})\right) + (1-\rho)a_0, \tag{6}$$

$$\widehat{\sigma}_{a,t+1}^2 = \rho^2 \widehat{\sigma}_{a,t}^2 \left(1 - \frac{\widehat{\sigma}_{a,t}^2}{\widehat{\sigma}_{a,t}^2 + \sigma_\epsilon^2}\right) + (1-\rho^2)\sigma_{a,0}^2, \tag{7}$$

where $\widehat{\sigma}_{a,t}^2$ describes the uncertainty around the perceived skill $\widehat{a}_{i,t}$ given initial conditions a_0

[6]While decreasing returns to scale are common in theoretical equilibrium models, empirical evidence for decreasing returns to scale is not ubiquitous. Pástor et al. (2015) find a negative but insignificant effect of fund size on fund returns. Zhu (2018) argues that the Pástor et al. (2015) estimator suffers from a misspecification which results in a loss of power, and finds stronger evidence of decreasing returns to scale after correcting for the misspecification. Chen et al. (2004) and Pástor et al. (2019) find additional evidence of decreasing returns to scale, whereas Reuter and Zitzewitz (2015) find little evidence of decreasing returns to scale.

and $\sigma^2_{a,0}$.[7] Following Roussanov et al. (2018), we assume a log specification for the decreasing returns to scale technology; that is, $c(\text{AUM}) = \eta \log(\text{AUM})$, where η is a parameter capturing fund returns' sensitivity to AUM. We use maximum likelihood to estimate the model on the fund level (using gross fund returns and fund size).[8] We run a factor-regression as in Equation (1) but over the entire life of a fund using the same benchmark that we believe analysts use and 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 return to form the annual abnormal return. The AUM are 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 compared to Roussanov et al. (2018). Note that their sample is different 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 estimate of the prior mean of managerial skill is 1.49%, the prior standard deviation is 1.79%, the residual volatility is 7.51%, and the persistence parameter is 0.94. With a standard deviation of log size of 1.82, the decreasing returns to scale parameter estimate of 0.18% implies that a one-standard-deviation increase in log size leads to a 0.33 percentage points decrease 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 equilibrium implication is that alphas are zero at any point in time, otherwise the money of risk-neutral investors would flow in and out of funds, thereby impacting alphas through decreasing return to scale and competing away any alphas. In contrast, a rational learner expects the abnormal return to

[7]Equation (4) is the measurement equation and Equation (5) is the state transition equation in the Kalman filtering. Mamaysky et al. (2008) use Kalman filtering to track dynamic factor loadings and to identify managers with market timing ability ex ante. Factor loadings in our model are constant.

[8]The model assumes that ϵ is uncorrelated across observations. The assumption is more likely to hold with fund returns instead of share class returns as share class returns of a given fund are highly correlated.

be

$$\mathrm{E}_t[r_{i,t+1}] = \widehat{a}_{i,t} - \eta \log(\mathrm{AUM}_{i,t}) - f_{i,t}, \tag{8}$$

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 beliefs. We assume rational expectations to form alphas at every point in time for every fund according to Equation (8).

In our empirical implementation of the model, the forecast horizon is one year. Morningstar states that, for example, the medalist ratings indicate an expected outperformance "over the long term, meaning a period of at least five years." To compare analyst alphas to our model, we assume that the analysts' five-year forecasts equal their unobserved one-year forecasts. An alternative would be to iterate Equation (4) 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 significantly complicate the model; it would require additional assumptions on the fee setting behavior by the fund over time and on how investors' money flows in and out of funds in response to past performance.

5 Results

5.1 Distribution of alphas

Figure 1 shows the distribution of net-of-fee analyst alphas in % per year together with the model-implied alphas for all funds in the sample as of April 2020. We include funds with a new Analyst Rating, an old Analyst Rating using our imputations, and funds with a Quantitative Rating (henceforth "all ratings"). Analyst alphas are much more dispersed than the rational learner's alphas, which previews our result of differences in expectations formation.

Figure 3 shows the distribution of analyst alphas together with the historically realized alphas for the same set of funds. The equal-weighted realized alpha for funds in the sample in April 2020 is -0.81% per year.[9] 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 et al., 2006; Kacperczyk et al., 2008; Fama and French, 2010). Nevertheless, heterogeneous analyst alphas cannot be taken for granted because many funds exhibit realized alphas different from zero by luck, or lack thereof, alone. For instance, Barras et al. (2010) report that 75% of the funds in their sample have a true alpha of zero.

5.2 Relation of alphas to skill, size, and fees

How do analysts form their expectations? Although Morningstar details how analysts arrive at their alpha estimates, the key inputs to the process—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 the calculation of the scaling factor, whereas the previous section demonstrates that past returns are one of the key inputs to the expectations formation process in a standard rational model.

According to the model, three factors determine alphas: perceived skill, fund size, and fees. We start by investigating the univariate relationships between alphas, "net skill" $(\widehat{a}_{i,2019} - f_{i,2019})$, and size. We first sort funds into deciles according to their net skill at the end of the sample and then compute average alphas across deciles for both analysts and

[9] Consistent with previous studies, the average realized alpha for all funds ever in the Morningstar data, dead or alive, is negative (-1.30%). There is some debate in the literature whether the negative average alpha arises because of using uninvestable benchmarks (Cremers et al., 2012; Berk and van Binsbergen, 2015). The Morningstar Category benchmark we use is dictated by Morningstar's methodology and has been used in the literature before (see, e.g., Pástor et al., 2015).

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 of funds with a new rating compared with all ratings. The reason is that analysts are least optimistic of funds with a Quantitative Rating, and these funds constitute the majority of the all ratings sample. Both analysts and the rational learner expect the majority of funds to earn negative alphas: in Panel (b) the alpha turns positive only at the 9th decile. This result is reminiscent of the results in Roussanov et al. (2018), who find 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 significantly with size, whereas the rational learner's expectations are either unrelated to size as in Panel (a) or increase much less with size as in Panel (b). In Panel (b) the mismatch is particularly evident: Analysts expect the largest funds to earn abnormal returns of 0.46% per year, whereas the rational learner expects abnormal returns of −0.37%.

Next, we formally evaluate our learning model in a multivariate regression. Equation (8) together with rational expectations has clear predictions for a regression of alphas on perceived skill, size (log AUM), and fees: the coefficient estimates should be 1, $-\eta$, and –1, respectively. Moreover, the constant should be zero. Table 5 presents the results. Specification (1) estimates the cross-sectional regression for 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.

Perceived skill. For the sample of funds with new Analyst Ratings, specification (1) shows that analysts' expectations respond significantly less to an increase in perceived skill than what the model predicts. In (1) a one-percentage-point increase in perceived skill is associated with a 0.73 percentage points larger analyst alpha as opposed to the one-to-one pass through predicted by the model. However, for the sample of all ratings in (2) the coefficient on perceived skill is close to one and not statistically different from one.

Fund size. The estimate on log AUM is close to zero in (1), positive in (2), and statistically significantly different from -0.18% ($-\widehat{\eta}$) in both columns. The positive coefficient estimate on fund size in (2) is the opposite sign of the prediction by the model. Through the lens of the model, it implies that analysts do not only not account for decreasing returns to scale when forming their expectations, but that they believe in *increasing* returns to scale. The model, however, may be misspecified. It is possible that size correlates with omitted variables which analysts use to form their expectations. Irrespective of analysts' beliefs about the causal effect of size on returns, given the crucial role decreasing returns to scale play in the rational expectations model our estimates present a significant deviation from the model. This deviation is perhaps not too surprising: In light of the econometric challenges and sophistication required to detect decreasing returns to scale in the data (Pástor et al., 2015; Zhu, 2018), the rational expectations assumption that agents know the decreasing returns to scale parameter with certainty is conceptually challenging.[10] Moreover, our results mirror survey evidence reporting that investors do not believe active mutual funds suffer from decreasing returns to scale (Choi and Robertson, 2019).

Fees. Consistent with the model, analyst alphas decrees with fees. In specification (1), we cannot reject the hypothesis that the coefficient on fees is different from minus one, whereas

[10] Pástor and Stambaugh (2012) relax this assumption to explain the size of the active fund industry. In their model the agent is uncertain about the decreasing returns to scale parameter but has priors that imply decreasing returns to scale.

the coefficient estimate in (2) is smaller than minus one.

Constant. While the constant is not statistically different from zero in (1), it is significantly negative in (2). This means that analysts are more pessimistic compared with the rational model, consistent with the univariate evidence in both Panel (b)'s of Figures 4 and 5, respectively.

In sum, the data are broadly consistent with the model's prediction for perceived skill and fees. We believe that the results on perceived skill and fees are remarkable given the simplicity of the learning model, which necessarily abstracts from the complex decision-making process that analysts use to form their expectations in reality. At the same time, the results in Table 5 point to a significant deviation of the model related to size, which implies that analysts' beliefs can hardly be the beliefs of a representative investor in a rational expectations equilibrium model. One advantage of working with mutual fund data is that the representativeness assumption can be tested. Armstrong et al. (2019) show that Analyst Ratings positively correlate with fund flows in the period 2011–2015 even when the popular Star Rating is controlled for, suggesting that analysts' expectations are to some degree representative of investors' expectations. In Appendix C 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 (i.e., alpha) before October 2019.

Finally, we do not interpret our evidence as evidence for analyst irrationality. Instead, the simple model may miss important determinants of analysts' expectations. R^2s of 14% and 32% (as opposed to the 100% predicted by the model), respectively, corroborate this hypothesis.

5.3 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. In Berk and Green (2004) and our model, chasing past returns is entirely rational and the question whether analysts (and the investors that follow them) over- or underreact to past performance relative to the rational benchmark becomes a quantitative one. Our model has clear predictions on how past returns should affect current expectations: They should not at all once perceived skill, size, and fees are controlled for.

For the sample of all ratings we were not able to reject the hypothesis that the coefficient on perceived skill is different from one. We now expose the model to a more stringent test by including several lags of past returns as additional regressors. Under the null hypothesis the coefficients on these lags 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 6 shows the results.

We find that analysts overreact to past returns except for the very most recent return for which we find mixed evidence. Analysts overreact to returns up to at least six years. For example, in specification (6) the first to the fifth lags are significantly positive, whereas the sixth lag is insignificant. We find that a constant skill specification results in an even larger overreaction to recent past returns because time-varying skill already allows the rational learner to place a larger weight on more recent past returns (untabulated results).

To understand how these results are consistent with the coefficient estimate of one on perceived skill in specification (2) of Table 5, in Appendix D we decompose perceived skill into a prior and past fees, size, and returns by iterating on Equation (6). We find that analysts underreact to the prior, which summarizes more distant past performance, but overreact to more recent returns. This is consistent with a large literature which shows that expectations are excessively influenced by more recent news.

5.4 Additional determinants of expectations

Morningstars' methodology suggests that we omit variables relevant to analysts' expectations formation. We are guided by Morningstar's methodology when 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 directly obtained 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.[11] Manager ownership has been shown to predict fund performance across different datasets (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 that are held in the ten largest positions) and a fund's tracking error (the standard deviation of returns in excess over the benchmark over the life of the fund). 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 a fund family's impact on individual fund performance (Massa, 2003; Gaspar et al., 2006; Pool et al., 2016; Berk et al., 2017; Ferreira et al., 2018).

Table 7 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 U.S. funds with new or old Analyst

[11] The SEC as of 2005 requires mutual fund managers to publicly file personal investments in their own funds. Managers have to report whether the dollar ownership in their funds falls in 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) to calculate manager ownership we use midpoints of the disclosed ownership ranges except for the maximum range "\$1,000,001 and above" for which we us the bottom of the range.

Ratings, or a Quantitative Rating. Specifications (1) and (4) replicate the specifications in Table 5 for the restricted sample of U.S. funds. 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. There are some differences for the sample of U.S. funds compared with Table 5. The coefficient estimate on fees is significantly below one and the estimate on size is significantly positive, respectively, for both the sample of new Analyst Ratings and the all ratings sample.

In all specifications, manager tenure and ownership are positively associated with analysts' expectations. In (5), a one-standard-deviation increase in tenure and ownership are associated with a 0.30 and 0.34 percentage points larger analyst alpha, respectively. Tracking error is negatively associated with analyst alphas. This stands in contrast to empirical evidence documenting that more active funds outperform their less active peers (Cremers and Petajisto, 2009; Amihud and Goyenko, 2013). Jenkinson et al. (2016) report that soft factors actually matter more than past performance for the recommendations of investment consultants. Instead, we find that—consistent with the model—perceived skill and fees still matter most. Untabulated regressions show that the standardized coefficient estimates on fees and perceived skill are 1.51 and 0.73 in (5), respectively.

Once fixed effects are included in (3) and (6), the R^2s increase by 30 percentage points. 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 the fund family variables relevant to analysts but hypothesize that governance plays a large role. For example, fund manager compensation practices are likely important and have been shown to systematically differ across fund families (Ibert et al., 2018; Ma et al., 2019).

In conclusion, the theoretical model captures the most important determinants of analysts' expectations (fees and perceived skill), but at the same time leaves out important determinants of expectations.

5.5 Industry performance and size

Equipped with alphas and fund size for each fund we compute the alpha for the active equity mutual fund industry as perceived by analysts and the rational learner. Analysts expect a value-weighted abnormal return of 0.70% going forward. In contrast, the rational learner is much more pessimistic and expects an abnormal industry return of −0.35%, which is statistically different from zero. The equal-weighted averages reveals that analysts' optimism is truly rooted in the performance of the largest funds: the average abnormal return is −1.21% for analysts, which is lower than the −0.72% average abnormal return for the rational learner. The negative value-weighted alpha for the rational learner together with her beliefs in decreasing returns to scale implies that the rational learner thinks the industry is too large in terms of AUM. The absence of decreasing returns to scale in analysts' belief, however, makes it difficult to translate analyst alphas into beliefs about the optimal size of the mutual fund industry.

Last, we compute the total value analysts expect mutual funds to add (Berk and van Binsbergen, 2015):

$$\mathrm{E}^{s}_{2019}[\text{ValueAdded}_{2020}] = \sum_{i=1}^{N} \left[\left(\mathrm{E}^{s}_{t}[r_{i,2020}] + f_{i,2019} \right) \times \text{AUM}_{i,2019} \right], \tag{9}$$

where N is the total number of funds in April 2020. Analysts expect mutual funds to extract 132 billion USD from global equity markets in 2020 of which 58 billion USD accrue to investors. In contrast, the rational learner expects mutual funds to only extract 44 billion USD from global equity markets and investors to lose 30 billion USD in value.

6 Conclusion

We use forward-looking Morningstar Analyst Ratings to infer a distribution of expected abnormal returns for mutual funds. We have three main findings. First, we find significant dispersion of alphas in contrast to expectations in an NRE equilibrium, in which all alphas are zero or close to zero. Second, we cannot consistently reject the hypothesis that analysts' expectations respond to perceived managerial skill and fees as predicted by a rational expectations model but we can reject the hypothesis that expectations respond to fund size as predicted by the model. Third, additional manager and fund family characteristics are important to analysts' expectations formation.

Ultimately, we hope that our approach can shed 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 evidence for analyst irrationality.[12] Mutual fund investors' capital allocation decisions in response to forecast errors in turn can help to examine mutual fund investors' rationality.

[12]Predictable forecast errors are evidence for irrationality but not conclusive evidence. For instance, in the NRE equilibrium model of Pástor and Stambaugh (2012) forecast errors are predictable.

Appendices

A Morningstar analyst report

Below, we include an anonymized example of an analyst report. The report is for a fund rated under the new methodology and is titled "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.

B Data appendix

B.1 Morningstar data

We retrieve data for open-end equity mutual funds belonging to 127 Morningstar categories from Morningstar Direct. The categories cover all funds with Morningstar Analyst Ratings according to the new methodology as of April 2020. The data contain Morningstar's fund and share class identifiers, the Morningstar category, returns, share class net assets, total fund net assets, fees, and monthly Morningstar Analyst and Quantitative Ratings. We downloaded the entire time-series from January 1965 to April 2020, but benchmark returns are only available from January 1979 and onwards.

In addition, we download monthly returns from January 2000 to April 2020 for all Morningstar global categories that we assign Morningstar categories to. We need these additional returns to correctly estimate the before-fee potential of every fund strategy.

B.2 Replication of Analyst Ratings

B.2.1 Gross returns

To estimate historical before-fee fund alphas we use a variable for the gross return, which is presumably what Morningstar does too.[13] Morningstar uses the fee variable "representative cost" to calculate gross returns. Hence, using net fund returns and adding back the monthly representative cost should yield a similar gross return.

B.2.2 Benchmark indices

Morningstar evaluates the before-fee alpha potential of a given strategy using factor regressions of fund returns on Morningstar Category specific benchmark returns. We use the Morningstar Category Index as the benchmark because we believe that Morningstar analysts rely on this index. If this benchmark is missing for a Morningstar Category, we use the categories' FTSE/Russel benchmark index or S&P Dow Jones benchmark index, whichever index has the longest available data history.

Equipped with the time-series of before-fee fund returns and benchmark returns, we estimate historical before-fee alphas to calculate the semi-interquartile range as described in

[13] 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.

the main text.

B.2.3 Pillar ratings

Morningstar analysts evaluate funds on three areas that they believe are crucial to predict future success: People, Process, and Parent. We noticed that pillar ratings are missing for some share classes that hold 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. We then calculate the forward-looking before-fee alpha as described in the main text.

B.2.4 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 on the re-occurring costs charged by funds. Since we aim to replicate the Analyst Ratings we also use monthly representative cost as a measure of share class specific fees.

We notice that fees are still missing for some share classes that hold an Analyst Rating at the end of the sample. In those cases, we source fees for the last month in the sample from other variables to be able to replicate as many ratings as possible. In particular, we fill missing data with the annual report net expense ratio, ongoing cost, and the prospectus net expense ratio.

B.2.5 Aggregation from share class to fund level

We take a value-weighted average of share class level alphas to form fund level alphas. 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 (cleaning described below). In case all share classes have missing assets, we take an equal-weighted average.

B.3 Data for estimation of rational model of fund performance

The replication of Analyst Ratings only required a historical time-series of before-fee and benchmark returns. For the estimation of the rational model of fund performance, in addition we need historical data on yearly fund sizes. Before estimating the model, we first clean the data in accordance with the literature (e.g., Pástor et al., 2015; Berk and van Binsbergen, 2015).

B.3.1 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. First, we create a variable for the fraction change in assets from last month to the current month,

$$\%\text{AUM}_t = \frac{\text{AUM}_t - \text{AUM}_{t-1}}{\text{AUM}_{t-1}}. \tag{B1}$$

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

$$\text{Reversal}_t = \frac{\text{AUM}_{t+1} - \text{AUM}_t}{\text{AUM}_t - \text{AUM}_{t-1}}. \tag{B2}$$

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

$$\text{abs}(\%\text{AUM}_t) >= 0.5, -0.75 > \text{Reversal}_t > -1.25, \text{ and } \text{AUM}_{t-1} >= 10 \text{ million}, \tag{B3}$$

then we set assets at time t (i.e., 2 million in this example) to missing. As a result of this procedure, 0.06% of monthly total fund net asset 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 are a significant number of missing asset observations. This is in part due to funds reporting at a quarterly or annual frequency, in particular before 1993. We apply the following procedure to fill in missing monthly share class net asset and total fund net asset data:

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 et al. (2018). Specifically, let $[t_0, t]$ and $[t + n, T]$ be periods when asset data is non-missing. The missing values are filled as follows:

$$\text{AUM}_k = F \times \text{AUM}_{k-1}(1 + r_k), \text{ for } k \in [t + 1, t + n - 1], \tag{B4}$$

$$F = \left(\frac{1}{\prod_{k=t+1}^{t+n} (1 + r_k)} \frac{\text{AUM}_{t+n}}{\text{AUM}_t} \right)^{\frac{1}{n}}, \tag{B5}$$

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 twelve months between non-missing observations at time t and $t + n$.

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

B.3.2 Aggregation of share class level to fund level

We take value-weighted averages of returns and fees across share classes with 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. In case all share classes have missing assets, we take an equal-weighted average.

B.4 Fees

At the share class level we set fees smaller or equal to zero to missing and fill in missing data with the annual report next expense ratio. We use the fiscal year end month if available in Morningstar Direct and assume the fiscal year ends in December otherwise. Following Pástor et al. (2015) we also exclude fund-month observations with fees below 0.1% per year since it is unlikely that any actively managed fund charges such low fees.

B.5 Identifying index funds

In order 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, then we classify it as an index fund. Otherwise, we move to step 2.

2. If the fund name contains "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 2,506 index funds out of 35,010 funds in total (7.2%).

B.6 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.

B.7 Further sample restrictions

We drop observations before the fund's (inflation adjusted) AUM reached 5 million USD similar to Berk and van Binsbergen (2015) and Fama and French (2010). We keep only funds with twelve monthly observations in a given year and twelve 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.

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

C Fund flows and Analyst Ratings (2011–2020)

Table C1 shows regressions of monthly fund flows,

$$\text{Flow}(\%) = \frac{\text{AUM}_{i,t} - \text{AUM}_{i,t-1} \times (1 + R_{i,t})}{\text{AUM}_{i,t-1} \times (1 + R_{i,t})} \times 100, \tag{B6}$$

on Morningstar Analyst Ratings, Star Ratings, and various control variables. In specification (1) a fund with a Gold rating receives 0.530 percentage points larger flows than a fund with a Neutral rating in a given year-month. In (2) a five-star fund receives 1.155 percentage points larger flows than 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 and Bronze funds still significantly attract more flows than funds with a Neutral Analyst Rating. A Gold fund receives 0.195 percentage points larger flows than 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.

Table C1: Fund flows on Analyst Ratings

	Analyst Ratings 2011–2020			All Ratings 2017–2020		
	(1)	(2)	(3)	(4)	(5)	(6)
Gold	0.530***		0.195***	0.449***		0.137**
	(0.055)		(0.059)	(0.061)		(0.061)
Silver	0.311***		0.077	0.262***		0.039
	(0.053)		(0.051)	(0.048)		(0.050)
Bronze	0.237***		0.097**	0.174***		0.040
	(0.043)		(0.040)	(0.030)		(0.026)
Neutral						
Negative	−0.109		0.057	−0.050		0.077
	(0.159)		(0.162)	(0.042)		(0.047)
Unrated	0.031		−0.001	−0.046		−0.090
	(0.044)		(0.043)	(0.079)		(0.077)
Five-star		1.155***	1.208***		1.065***	0.559***
		(0.034)	(0.150)		(0.043)	(0.200)
Four-star		0.399***	0.459***		0.330***	−0.169
		(0.022)	(0.153)		(0.039)	(0.207)
Three-star			0.065			−0.491**
			(0.159)			(0.203)
Two-star		−0.219***	−0.153		−0.134***	−0.632***
		(0.027)	(0.164)		(0.034)	(0.209)
One-star		−0.347***	−0.282*		−0.303***	−0.817***
		(0.042)	(0.164)		(0.069)	(0.233)
No-star		0.462***	0.522***		0.566***	0.071
		(0.042)	(0.168)		(0.072)	(0.217)
N	920698	920698	920698	308651	308651	308651
Adj. R^2	0.09	0.09	0.09	0.10	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 April 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's inception), fees, past 12-month fund return, 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.

D Decomposing perceived skill

By iterating Equation (6) we can decompose perceived skill into a prior of perceived skill, past net abnormal returns, past sizes, and past fees. For notational simplicity, let us assume $\rho = 1$. After one iteration we have

$$\begin{aligned}
\widehat{a}_{i,t+1} = \widehat{a}_{i,t-1}&(1-K_t)(1-K_{t-1}) \\
&+ K_t r_{i,t+1} + K_{t-1}(1-K_t) r_{i,t} \\
&+ K_t \log(\text{AUM})_{i,t} + K_{t-1}(1-K_t)\log(\text{AUM})_{i,t-1} \\
&+ K_t f_{i,t} + K_{t-1}(1-K_t) f_{i,t-1},
\end{aligned}$$

where $K_t = \widehat{\sigma}^2_{a,t}/(\widehat{\sigma}^2_{a,t} + \widehat{\sigma}^2_{\epsilon})$ is the Kalman gain (signal-to-noise ratio). By iterating to date 0 we get

$$\begin{aligned}
\widehat{a}_{i,t+1} = {} & a_{i,0}(1-K_t) \times (1-K_{t-1}) \times ... \times (1-K_0) \\
& + K_t r_{i,t+1} + K_{t-1}(1-K_t) r_{i,t} + ... + K_0(1-K_t) \times ... \times (1-K_1) r_{i,1} \\
& + K_t \log(\text{AUM})_{i,t} + K_{t-1}(1-K_t)\log(\text{AUM})_{i,t-1} + ... + K_0(1-K_t) \times ... \times (1-K_1)\log(\text{AUM}_{i,0}) \\
& + K_t f_{i,t} + K_{t-1}(1-K_t) f_{i,t-1} + ... + K_0(1-K_t) \times ... \times (1-K_1) f_{i,0}.
\end{aligned}$$

For example, if we regress the rational learner's expectations on a constant, size, fees, lagged return, and lagged perceived skill, the coefficient estimate on the lagged return recovers the Kalman gain. By running the same regressions for analyst alphas, we can recover the implicit Kalman gains in analysts' expectations. In the general case with $\rho \neq 1$, the Kalman gains are adjusted for the persistence of skill.

Table D1 presents regressions of alphas on lagged returns, lagged fund sizes, lagged fees, and a lag of perceived skill for the all ratings sample. We show specifications with three, six, and nine lags, respectively. To save space, we do not show coefficient estimates on fund sizes and fees. Specifications (1)–(4) show the estimates for the rational learner, whereas (5)–(8) show the estimates for analyst alphas. For example, the Kalman gain (adjusted for skill persistence) for the rational learner is 0.041 in (2), whereas it is 0.039 for analysts in (6). As in Table 6 we find that analysts' expectations respond more to past returns than the rational learners' expectations except for the very most recent past return. This holds true up to a lag of six past returns.

In specification (2) of Table 5 analysts respond one-to-one to perceived skill. The skill decomposition in this appendix suggests that analysts must then underreact to the lag of perceived skill. Indeed, in specification (4) of Table D1 the coefficient estimate on the ninth lag of perceived skill is 0.382 for the rational learner, whereas it is only 0.143 in (8) for analysts.

Table D1: Cross-sectional regressions of alphas on components of perceived skill

	Rational learner				Analyst			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Perceived skill	1.000***				1.005***			
	(0.000)				(0.036)			
Perceived skill (t–3)		0.719***				0.626***		
		(0.001)				(0.042)		
Perceived skill (t–6)			0.524***				0.269***	
			(0.000)				(0.044)	
Perceived skill (t–9)				0.382***				0.143***
				(0.001)				(0.045)
Net abn. return		0.041***	0.040***	0.040***		0.039***	0.037***	0.031***
		(0.000)	(0.000)	(0.000)		(0.005)	(0.005)	(0.006)
Net abn. return (t-1)		0.037***	0.037***	0.036***		0.072***	0.076***	0.073***
		(0.000)	(0.000)	(0.000)		(0.005)	(0.005)	(0.006)
Net abn. return (t–2)		0.034***	0.033***	0.033***		0.036***	0.044***	0.039***
		(0.000)	(0.000)	(0.000)		(0.005)	(0.006)	(0.006)
Net abn. return (t–3)			0.030***	0.030***			0.039***	0.039***
			(0.000)	(0.000)			(0.005)	(0.006)
Net abn. return (t–4)			0.028***	0.027***			0.047***	0.050***
			(0.000)	(0.000)			(0.006)	(0.006)
Net abn. return (t–5)			0.025***	0.025***			0.030***	0.030***
			(0.000)	(0.000)			(0.005)	(0.006)
Net abn. return (t–6)				0.022***				0.015***
				(0.000)				(0.005)
Net abn. return (t–7)				0.020***				0.003
				(0.000)				(0.006)
Net abn. return (t–8)				0.018***				0.025***
				(0.000)				(0.005)
N	10763	8981	7440	6043	10763	8981	7440	6043
Adj. R^2	1.00	1.00	1.00	1.00	0.32	0.35	0.37	0.37
Fees and size	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

The table shows regressions of Morningstar analyst alphas on skill as perceived by a rational learner, fund sizes (logarithm of assets under management in millions of USD), lagged fund sizes, past net-of-fee abnormal fund returns, fees, and lagged fees. The sample contains funds with a new Analyst Rating, an Analyst Rating under the old methodology, and funds with a Quantitative Rating. The coefficient estimates on fund sizes, lagged fund sizes, fees, and lagged fees are not shown. 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

Adam, Klaus, Albert Marcet, and Johannes Beutel, 2017, Stock Price Booms and Expected Capital Gains, *American Economic Review* 107, 2352–2408.

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

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.

Barberis, Nicholas, Robin Greenwood, Lawrence Jin, and Andrei Shleifer, 2015, X-CAPM: An Extrapolative Capital Asset Pricing Model, *Journal of Financial Economics* 115, 1–24.

Barberis, Nicholas, Robin Greenwood, Lawrence Jin, and Andrei Shleifer, 2018, Extrapolation and Bubbles, *Journal of Financial Economics* 129, 203–227.

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, John R. Graham, and Campbell R. Harvey, 2013, Managerial Miscalibration, *Quarterly Journal of Economics* 128, 1547–1584.

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

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.

Berk, Jonathan B., Jules H. van Binsbergen, and Binying Liu, 2017, Matching Capital and Labor, *Journal of Finance* 72, 2467–2504.

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., and Paul J. Irvine, 2006, Bayesian Alphas and Mutual Fund Persistence, *Journal of Finance* 61, 2251–2288.

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.

Cieslak, Anna, 2018, Short-Rate Expectations and Unexpected Returns in Treasury Bonds, *Review of Financial Studies* 31, 3265–3306.

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.

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.

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.

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

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

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

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

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

Greenwood, Robin, and Samuel G. Hanson, 2013, Issuer Quality and Corporate Bond Returns, *Review of Financial Studies* 26, 1483–1525.

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

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.

Harvey, Campbell R., and Yan Liu, 2019, Cross-Sectional Alpha Dispersion and Performance Evaluation, *Journal of Financial Economics* 134, 273–296.

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.

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

Kacperczyk, Marcin, Clemens Sialm, and Lu Zheng, 2008, Unobserved Actions of Mutual Funds, *Review of Financial Studies* 21, 2379–2416.

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.

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

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

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.

Mamaysky, Harry, Matthew Spiegel, and Hong Zhang, 2008, Estimating the Dynamics of Mutual Fund Alphas and Betas, *Review of Financial Studies* 21, 233–264.

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.

Nagel, Stefan, and Zhengyang Xu, 2019, Asset Pricing with Fading Memory, *Working Paper* .

Pástor, Luboš, and Robert F. Stambaugh, 2002, Investing in Equity Mutual Funds, *Journal of Financial Economics* 63, 351–380.

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, 2019, Fund Tradeoffs, *Journal of Financial Economics* forthcoming.

Pool, Veronika K., Clemens Sialm, and Irina Stefanescu, 2016, It Pays to Set the Menu: Mutual Fund Investment Options in 401 (k) Plans, *Journal of Finance* 71, 1779–1812.

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

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.

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*: 3-pillar ratings (People, Process, Parent), SIQR (dispersion of CAPM alphas of fund strategy), share class fees *Old*: 5-pillar ratings (People, Process, Parent, Performance, Price)	*New*: 3-pillar ratings (People, Process, Parent) estimated using a machine learning algorithm, SIQR (dispersion of CAPM alphas of fund strategy), share class fees *Old*: 5-pillar ratings (People, Process, Parent, Performance, 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)	*New*: share class *Old*: fund ("representative" share class)	Fund
Ranked 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

Continued on next page

Table 1 continued from previous page

	Star Rating	Analyst Rating	Quantitative Rating	Sustainability Rating
Rating peer group	Morningstar Category	Morningstar Category	Morningstar Category	Morningstar Global Category
Medalist ranking (Gold, Silver, 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, higher bar for medalist ranking	10/2019: ratings assigned at share class level based on expected net-of-fee alphas, reduction to three pillars, higher bar for medalist ranking	10/2019: replacement of company ESG Rating with its ESG Risk Rating
Selected academic work 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 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 in 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 the fund if a human analyst covered the fund. 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	58	13	0	0	0	71
Neutral	23	2036	135	3	0	2197
Bronze	0	60	1372	148	2	1582
Silver	0	1	128	1045	77	1251
Gold	0	0	0	51	309	360
Total	81	2110	1635	1247	388	5461

Panel B: Morningstar Quantitative Ratings

	Replicated rating					
Actual rating	Negative	Neutral	Bronze	Silver	Gold	Total
Negative	8359	385	1	1	0	8746
Neutral	248	16515	624	3	0	17390
Bronze	3	293	4054	266	2	4618
Silver	0	3	218	2533	78	2832
Gold	0	0	1	171	1357	1529
Total	8610	17196	4898	2974	1437	35115

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 Rating is tabulated in rows, whereas the replicated rating is 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	867		3509	10153	127	374	993	3021	7963
Old Analyst Rating	593		3707	11430	62	240	859	2628	6772
Quantitative Rating	9303		332	943	11	28	90	285	757
All ratings	10763		774	4183	12	33	118	420	1332
No rating	821		207	1976	7	15	43	118	288
All	11584		734	4069	11	31	108	393	1249
Panel B: Fees									
New Analyst Rating	867	0.80	1.06	0.38	0.62	0.83	1.01	1.26	1.59
Old Analyst Rating	593	0.76	1.12	0.42	0.64	0.85	1.06	1.33	1.73
Quantitative Rating	9303	1.06	1.39	0.75	0.64	0.92	1.26	1.77	2.25
All ratings	10763	0.89	1.35	0.72	0.64	0.91	1.22	1.71	2.20
No rating	857	0.84	1.40	0.90	0.39	0.85	1.33	1.86	2.30
All	11620	0.89	1.36	0.73	0.63	0.90	1.23	1.72	2.20
Panel C: Perceived skill									
New Analyst Rating	867	2.27	2.02	0.67	1.30	1.55	1.93	2.36	2.85
Old Analyst Rating	593	2.29	2.01	0.69	1.26	1.51	1.90	2.43	2.94
Quantitative Rating	9303	1.71	1.41	0.63	0.69	1.07	1.42	1.71	2.16
All ratings	10763	2.07	1.49	0.67	0.74	1.12	1.48	1.83	2.32
No rating	859	1.58	1.51	0.54	1.06	1.43	1.49	1.49	1.97
All	11622	2.06	1.49	0.66	0.76	1.14	1.49	1.80	2.30
Panel D: Analyst alphas									
New analyst rating	867	1.28	0.68	1.42	–1.05	–0.28	0.71	1.59	2.47
Old analyst rating	593	1.53	0.58	1.28	–1.21	–0.52	0.76	1.52	2.20
Quantitative rating	9303	–0.48	–1.50	2.45	–4.68	–3.07	–1.40	0.14	1.53
All ratings	10763	0.70	–1.21	2.45	–4.48	–2.79	–1.05	0.55	1.75

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 April 2020. AUM are fund sizes 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 rational fund performance model

Parameter	Description	Estimate
η	Decreasing returns to scale (%)	0.18^{***}
		(0.01)
a_0	Prior mean (%)	1.49^{***}
		(0.07)
$\sigma_{a,0}$	Prior standard deviation (%)	1.79^{***}
		(0.04)
σ_ϵ	Residual standard deviation (%)	7.51^{***}
		(0.02)
ρ	Skill persistence	0.94^{***}
		(0.01)

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

	New Analyst Ratings	All ratings
	(1)	(2)
Perceived skill	0.727^{***}	1.005^{***}
	(0.084)	(0.036)
	[0.001]	[0.886]
Size ($\times 100$)	-0.012	0.112^{***}
	(0.031)	(0.012)
	[0.000]	[0.000]
Fees	-0.871^{***}	-1.358^{***}
	(0.128)	(0.034)
	[0.316]	[0.000]
Constant ($\times 100$)	0.216	-1.414^{***}
	(0.270)	(0.092)
	[0.424]	[0.000]
N	867	10763
Adj. R^2	0.14	0.32

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 funds with 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 on skill, size, fees, and the constant are equal to the model-predicted parameters of +1, −0.18 (the estimate of η in Table 4), −1, and 0, respectively.

Table 6: Cross-sectional regressions of alphas on past returns

	New Analyst Ratings			All ratings		
	(1)	(2)	(3)	(4)	(5)	(6)
Perceived skill	0.666***	0.160	−0.126	0.799***	0.396***	0.203*
	(0.117)	(0.151)	(0.189)	(0.056)	(0.077)	(0.108)
Net abn. return	−0.027**	−0.011	0.001	0.009	0.025***	0.025***
	(0.011)	(0.012)	(0.013)	(0.006)	(0.006)	(0.007)
Net abn. return (t–1)	0.022**	0.041***	0.048***	0.043***	0.061***	0.066***
	(0.010)	(0.011)	(0.012)	(0.005)	(0.006)	(0.007)
Net abn. return (t–2)	0.034***	0.055***	0.063***	0.014***	0.032***	0.039***
	(0.010)	(0.012)	(0.012)	(0.005)	(0.006)	(0.007)
Net abn. return (t–3)		0.060***	0.060***		0.033***	0.036***
		(0.011)	(0.013)		(0.005)	(0.007)
Net abn. return (t–4)		0.040***	0.043***		0.039***	0.046***
		(0.010)	(0.012)		(0.006)	(0.007)
Net abn. return (t–5)		0.027***	0.026**		0.019***	0.022***
		(0.010)	(0.011)		(0.005)	(0.006)
Net abn. return (t–6)			−0.002			0.008
			(0.009)			(0.006)
Net abn. return (t–7)			0.036***			0.006
			(0.013)			(0.007)
Net abn. return (t–8)			0.045***			0.024***
			(0.009)			(0.005)
Size (× 100)	0.002	0.046	0.097***	0.109***	0.130***	0.157***
	(0.033)	(0.036)	(0.037)	(0.014)	(0.016)	(0.018)
Fees	−0.893***	−0.712***	−0.577***	−1.303***	−1.181***	−1.115***
	(0.144)	(0.150)	(0.164)	(0.040)	(0.043)	(0.052)
Constant (× 100)	0.229	0.737**	0.724*	−1.014***	−0.506***	−0.394**
	(0.292)	(0.350)	(0.393)	(0.112)	(0.139)	(0.172)
N	834	792	729	9150	7566	6129
Adj. R^2	0.18	0.21	0.25	0.35	0.37	0.37

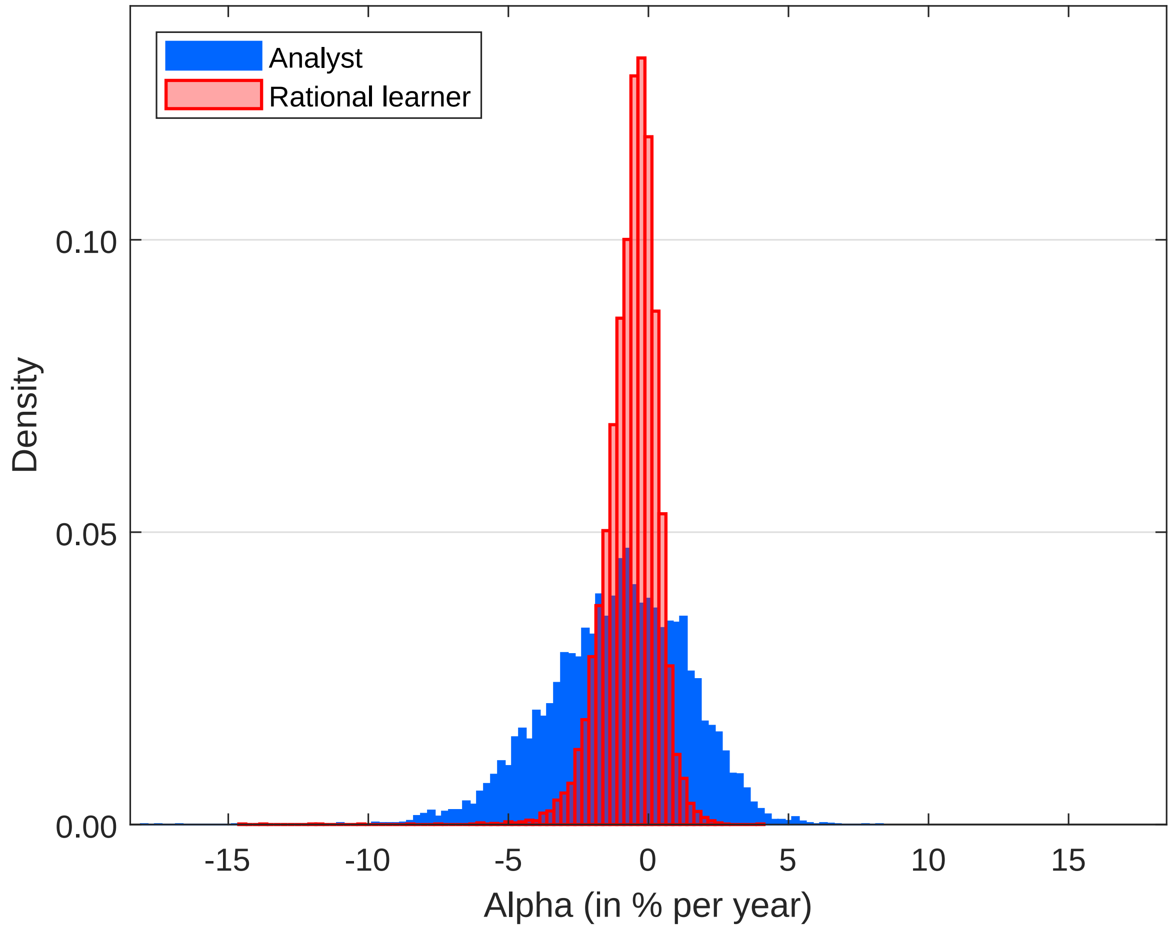
The table shows regressions of Morningstar analyst alphas on skill as perceived by a rational learner, past net-of-fee abnormal fund returns, fund size (logarithm of assets under management in millions of USD), and fees. Specifications (1) and (2) use funds with an Analyst Rating under the new methodology. Specifications (3) and (4) use funds with a new Analyst Rating, an Analyst Rating under the old methodology, and funds with 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.

Table 7: Cross-sectional regressions of alphas on additional characteristics

	New Analyst Ratings			All ratings		
	(1)	(2)	(3)	(4)	(5)	(6)
Rational learner						
Perceived skill	0.540***	0.714***	0.632***	1.117***	1.096***	0.755***
	(0.105)	(0.100)	(0.130)	(0.077)	(0.074)	(0.077)
Size (× 100)	0.116***	0.042	0.001	0.097***	0.003	0.079***
	(0.043)	(0.041)	(0.048)	(0.022)	(0.023)	(0.023)
Fees	−1.578***	−1.174***	−0.899**	−1.955***	−2.056***	−1.231***
	(0.223)	(0.241)	(0.351)	(0.145)	(0.154)	(0.142)
People						
Manager tenure		0.157***	0.216***		0.299***	0.290***
		(0.051)	(0.051)		(0.035)	(0.033)
Manager ownership		0.170***	0.214***		0.337***	0.196***
		(0.045)	(0.061)		(0.038)	(0.040)
Managerial multitasking		0.182***	0.314***		0.062**	0.009
		(0.033)	(0.109)		(0.025)	(0.048)
Management team		−0.109*	0.014		0.097**	0.145***
		(0.064)	(0.083)		(0.046)	(0.047)
Process						
Top 10 assets (%)		−0.144*	0.140		0.163***	0.212***
		(0.074)	(0.103)		(0.051)	(0.057)
Tracking error		−0.202***	−0.145		−0.136**	−0.138**
		(0.064)	(0.108)		(0.053)	(0.054)
N	456	456	411	2922	2922	2713
Adj. R^2	0.24	0.33	0.58	0.26	0.31	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 funds with 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.

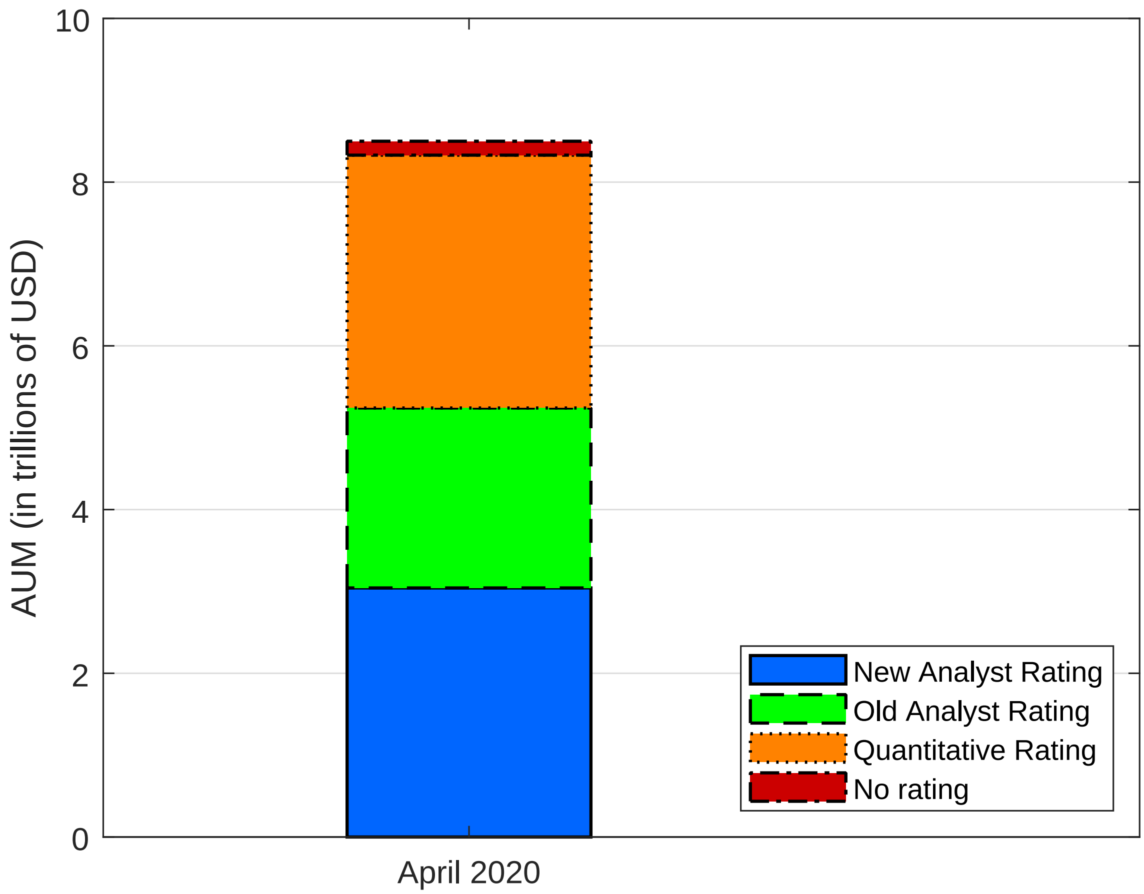
Figure 1: Histogram of analyst alphas and model-implied alphas



	Mean (V.W.)	Mean (E.W.)	S.D.	10%	25%	50%	75%	90%
Analyst	0.70	−1.21	2.45	−4.48	−2.79	−1.05	0.55	1.75
Rational learner	−0.35	−0.72	0.94	−1.88	−1.22	−0.60	−0.12	0.28

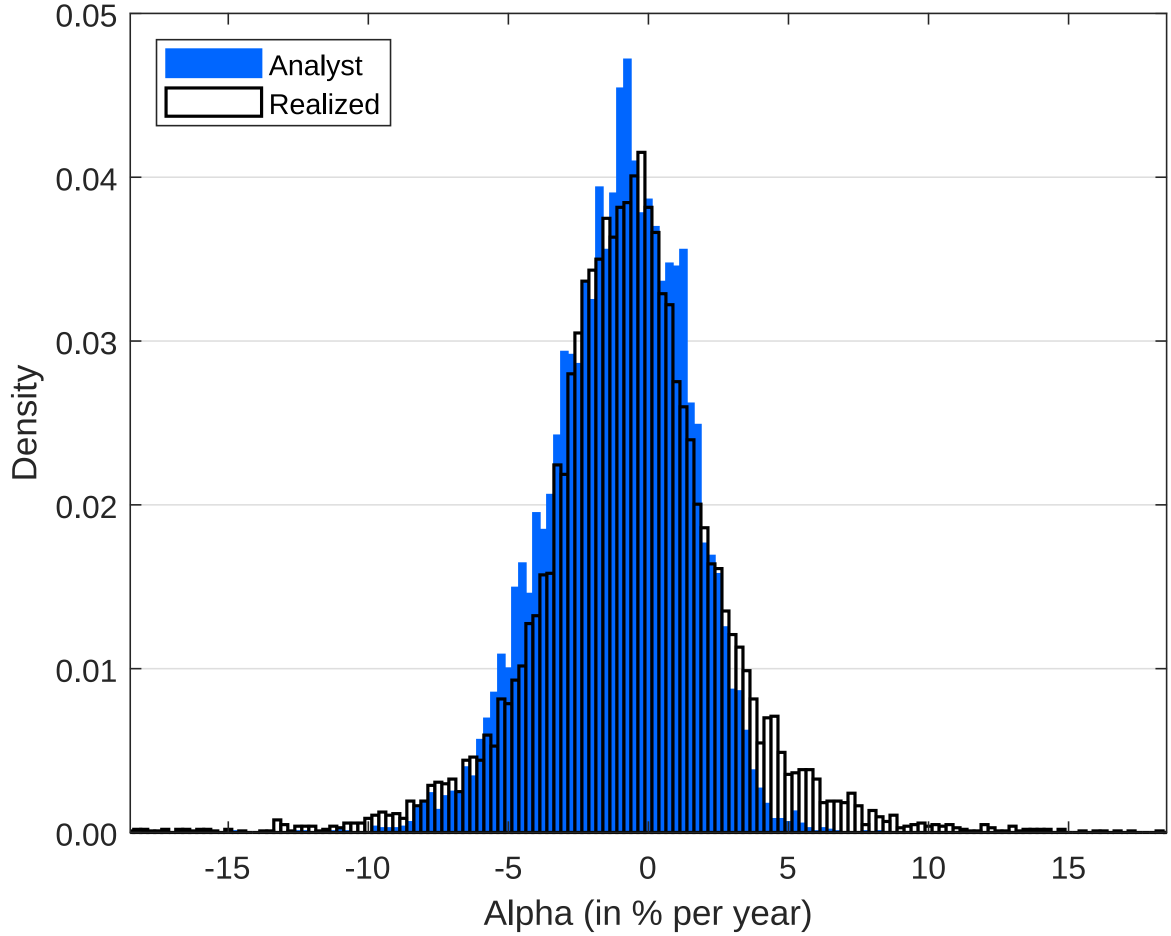
The figure shows the distribution of analysts' expected net-of-fee abnormal returns (alphas) and the distribution of expected net-of-fee abnormal returns implied by a rational model of fund performance as of April 2020. The rows below show value-weighted (by assets under management) and equal-weighted 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 as of April 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.

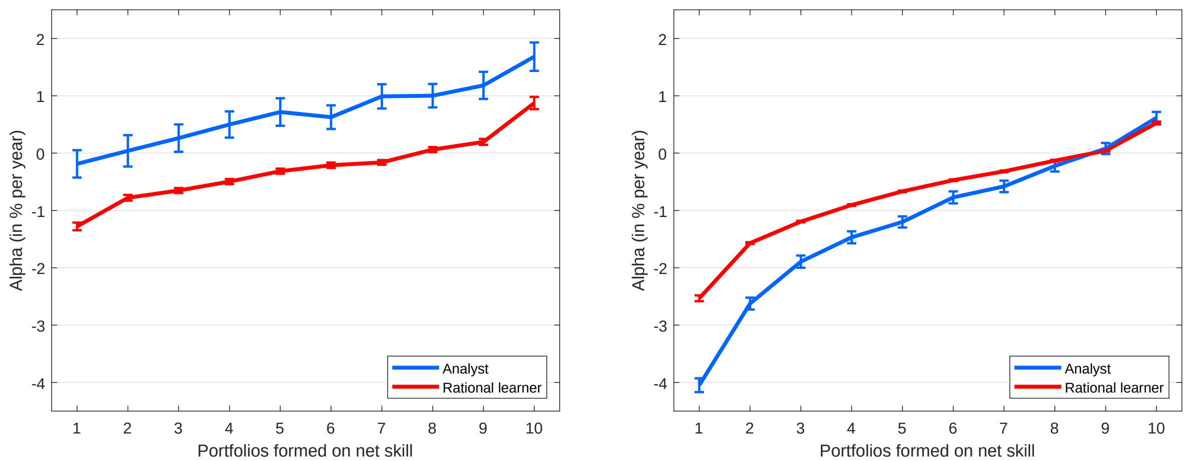
Figure 3: Histogram of analyst alphas and realized alphas



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 April 2020. Alphas are estimated relative to each fund's Morningstar Category benchmark. 0.16% of realized alphas are outside the range shown in the figure.

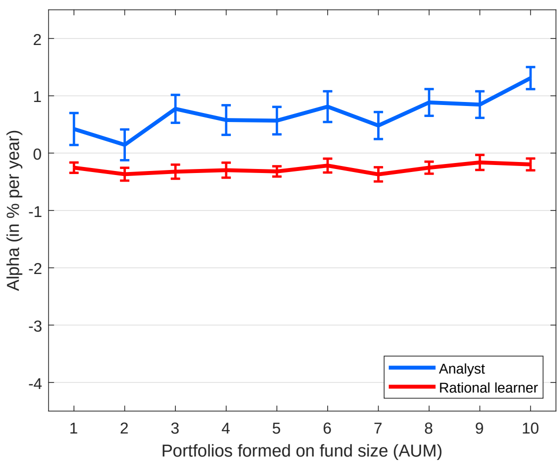
Figure 4: Alphas against net skill

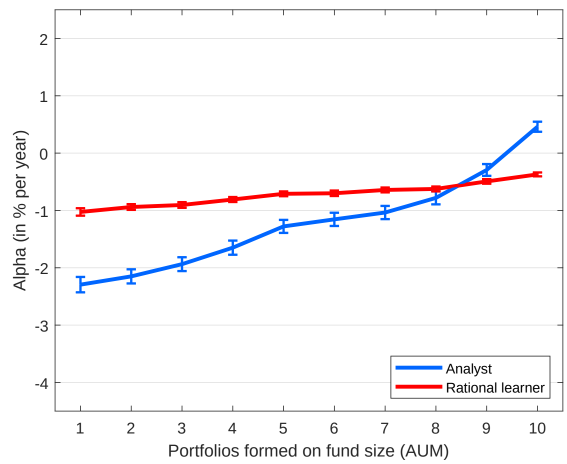
(a) New Analyst Ratings **(b)** All ratings



The figure shows expected net-of-fee abnormal returns (alphas) against net skill as of April 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 funds with a Quantitative Rating. The bars indicate 90% confidence bands.

Figure 5: Alphas against fund size





The figure shows expected net-of-fee abnormal returns (alphas) against fund size (AUM) as of April 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 funds with a Quantitative Rating. The bars indicate 90% confidence bands.