SFB 649 Discussion Paper 2017-001

Fake Alpha

Marcel Müller*
Tobias Rosenberger*
Marliese Uhrig-Homburg*

*Karlsruhe Institute of Technology (KIT), Germany

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

http://sfb649.wiwi.hu-berlin.de
ISSN 1860-5664

SFB 649, Humboldt-Universität zu Berlin
Spandauer Straße 1, D-10178 Berlin

SFB 649 ECONOMIC RISK BERLIN

Fake Alpha

Marcel Müller* Tobias Rosenberger† Marliese Uhrig-Homburg‡

January 31, 2017

Abstract

Why do investors entrust active mutual fund managers with large sums of money while receiving negative excess returns on average? Our explanation is that investors have a coarser information set than fund managers which leads them to systematically misinterpret managers' skill. When investors are unable to correctly quantify risk because they have no knowledge of factor investing on beyond-market-risk factors, *Fake Alpha* strategies based on factor investing look like skill from the investors' perspective. As running such strategies is relatively cheap for the managers, the investors' coarser information set misleads them to invest beyond the point of zero excess returns in equilibrium. We confirm our theory by analyzing the sample of US equity active managed mutual funds and find significant evidence of decreasing returns to scale at the fund level as well as negative excess returns to investors in equilibrium states.

Keywords: mutual funds, active management, managerial skill, alpha

JEL Classification Numbers: G11, G20, G23

*Chair of Financial Engineering and Derivatives, Karlsruhe Institute of Technology (KIT), Blücherstr. 17, 76185 Karlsruhe, Germany, marcel.mueller@kit.edu

†Chair of Financial Engineering and Derivatives, Karlsruhe Institute of Technology (KIT), Blücherstr. 17, 76185 Karlsruhe, Germany, tobias.rosenberger@student.kit.edu

‡Chair of Financial Engineering and Derivatives, Karlsruhe Institute of Technology (KIT), Blücherstr. 17, 76185 Karlsruhe, Germany, uhrig@kit.edu

We acknowledge financial support for this project from the Deutsche Forschungsgemeinschaft (DFG) through CRC 649 Economic Risk.

1 Introduction

Suppose the average active mutual fund investor is making a systematic error quantifying risk. How could he then be able to correctly evaluate the performance of a fund? We think he cannot, which might be the reason why e.g. Fama and French (2010) empirically observe negative risk adjusted average net returns to investors for active managed mutual equity. Extending the Berk and Green (2004) model by introducing investors who have a coarser information set compared to fund managers and are therefore unable to correctly quantify risk, we can show that negative net returns to investors in equilibrium might be simply a result of all agents behaving rational under these circumstances.

In line with the model, we empirically find that the average net alpha measured against the CAPM[1] is significantly negative in equilibrium states. In order to establish this empirical result, we introduce a new way of measuring a fund's size by standardizing the time series of size observations for each fund by its sample mean. Not only does this reduce the cross sectional variation of the size between funds, but we show that the measure facilitates at the same time the identification of equilibrium states. This allows us to jointly analyse equilibrium observations across different funds at different points in time. Using this standardized quantity, we also obtain a significant negative relationship between a fund's size and the net alpha to investors which is the second main implication of our model. Drilling deeper, we decompose the CAPM alpha in two components: First, the *Fake Alpha* component, which accounts for the beyond-market-risk premia part. We obtain it by using a multi-factor model, e.g. Fama and French (1993) (FF3) or Carhart (1997) (CH4), and then estimate the part of the CAPM alpha which is linked to loadings on factors other than the market factor. Second, the *True Alpha* component, which is equal to the difference between the CAPM alpha and the *Fake Alpha*, and henceforth represents the part of the CAPM alpha not linked to any beyond-market-factor risk premia. We provide evidence that fund managers consider

[1]The CAPM is originally developed by Sharpe (1964), Lintner (1965), and Mossin (1966).

Fake Alpha and *True Alpha* as substitutes: Using the *Industry Size* measure for market competitiveness as introduced by Pástor et al. (2015) on the one hand, and the varying risk premia sizes of the beyond-market-risk factors on the other hand, we find that whenever *Industry Size* and hence competitiveness rises it gets relatively expensive for the managers to generate *True Alpha*. In this situation, fund managers seem to increase their exposures towards the beyond-market-risk factors to generate *Fake Alpha* instead. This relationship is also confirmed the other way around, meaning whenever risk premia on beyond-market-risk factors increase and hence it gets relatively cheap to generate *Fake Alpa*, the managers excessively load on the beyond-market-risk factors. At the same time they focus less on the *True Alpha* part. This could be suggestive of managers being aware about the investors' coarser information set, but since being rewarded proportional to the managers' assets under management (AUM) their behavior is perfectly rational.

Our model is built upon two pillars: First, we assume that investors chase CAPM alphas. This is empirically backed by findings of Berk and van Binsbergen (2016) and Barber et al. (2016) who show that amongst all major asset pricing models it is most likely that investors built their investment decisions looking on CAPM alphas. Second, we assume that investors not only chase CAPM alphas, but they are also unaware of any beyond-market factor risk.[2] This enables a manager to easily generate CAPM alphas by loading on beyond-market-risk factors. Based upon those two pillars, we assume investors Bayesian-learn a manager's skill by observing past returns and provide flows to a fund up to the point where they expect zero excess returns, given their coarser information set. From an ex-post perspective although (i.e. looking at it from an outside planer's perspective) investors obtain negative excess returns in equilibrium. Thus, ultimately, investors' ignorance of factor strategies leads them to overestimate manager's skill, and eventually, to overinvest into the fund in equilibrium.

[2]There is an ongoing debate whether premia on several beyond-market factors are compensation for risk or exist due to market restrictions. Chu et al. (2016) find that e.g. the momentum premium severely decreases if short sale restrictions are relaxed. Our argument would also hold if instead of assuming that investors are unaware of beyond-market factor risk, we would assume that investors have no sense for market restrictions.

As equilibrium states are hard to identify for investors in the first place, and even from an outside perspective they are empirically rare in our sample, we neglect the possibility that investors learn about their mistake. We elaborate carefully on whether or not this no-learning assumption makes sense and find empirical support for our view.

The paper adds to the literature in several ways: First, we connect the empirical findings of Berk and van Binsbergen (2016) and Barber et al. (2016), who show that of all major assets pricing models the CAPM is the one that is most likely used by investors of equity mutual funds to determine their investment decision, with the strand of literature following Fama and French (1993), which collectively shows that various multi-factor models explain empirical equity prices more accurately than the CAPM. Second, we find strong empirical evidence for decreasing excess returns to scale on the fund level thereby supporting one basic insight of the Berk and Green (2004) model. This is new empirical evidence as recent work by Pástor et al. (2015) depicts no significant results for decreasing returns to scale. We use part of their econometric methodology to address several biases on the empirical side. However, looking at the standardized size of a fund instead of the absolute (deflated) assets under management we find a robust decreasing returns to scale pattern. Third, we show how to reconcile the implications from Berk and Green (2004) with empirical results of Fama and French (2010), who point out that net alphas are of remarkable negative size on average, by including investors who are unable to correctly quantify risk. Fourth, by decomposing the CAPM alphas in a *Fake Alpha* component that accounts for risk premia earned by loading on beyond-market-risk factors and the residual *True Alpha* component, we provide a novel way to show that those two components exhibit important differences regarding their relationship with overall market conditions. Hence, it might make sense to treat them separately when examining the mechanisms of the active managed mutual fund industry.

Our results are in line with an attenuated version of a thesis stated by John Cochrane (2010), who claims that "There is no alpha. There is just beta you understand and beta you do not understand...", as it seems, like the only beta an average mutual fund investor understands is the market beta. Within our model, even if the managers didn't have any skill at all, investors would still provide flows to their funds in the long-run equilibrium, since they confuse skill with risk premia on omitted factors. We explicitly add the term *attenuated*, since our empirical results do not clearly support the first part of the thesis that there is no alpha at all. However, obviously, the overall level of skill you need to explain the size of the active management industry is smaller assuming investors who are not able to correctly quantify risk, than the level of skill you need to explain the industry size in a purely neoclassical framework.

The paper is organized as follows: We introduce our theoretical model in Section 2. In Section 3 we explain the data we use to test the model as well as our estimators. In this section we also point out econometric issues that might bias our empirical results later on as well as how we address those. In Section 4 we analyze the results, starting of with examining the returns to scale relationship before looking at the level of alphas in equilibrium states and finally elaborating on whether or not there is support for our no-leaning assumption in the data. We conclude in Section 5.

2 The Model

To clarify our arguments we develop a simple model that captures the relationship between fund flows and alphas. In general, the model is a derivation of the work by Berk and Green (2004) with an additional investment opportunity on the manager's side that is unknown to investors. To simplify notation, let us consider a universe with a single fund only but numerous investors. In general, we think of the entire investment/allocation process in rounds. One round comprises the following three steps in the outlined order:

1. Investors learn about parameters by observing costs and returns.

2. Investors make their investment decision, i.e. they provide flows to the fund.

3. The manager allocates the received flows into different investment strategies.

The first step is skipped in the initial loop, instead investors start by setting their beliefs about parameters. Returns are realized between the rounds. Note that to explain the mechanisms of the model in the following, we do not go through the above steps in the outlined order, but rather follow a more illustrative path by first explaining the general setting, to then referring to the manager's perspective (step 3), before finally focusing on the investors' decision problem (steps 1 and 2).

2.1 General Setting

We assume that the capital flows in and out of a fund are linked to the fund's past net performance measured against the $CAPM$. As compensation for the management services the manager charges a fix proportional fee on the total assets under management q_t. Further, there exists a true pricing kernel TP which is driven by the market-risk factor and at least one additional state variable. The manager can choose to invest the provided funds in either a true active strategy a, a factor strategy f, or the market portfolio (indexing) i. By the term *factor strategy* we refer to strategies based on beyond market-risk factors that are included in the true pricing kernel, e.g. on *HML* or on *SMB* if we assume the true pricing kernel to be $FF3$. In absence of of any cost or fee, we denote the excess of market return of the active strategy and the factor strategy respectively by

$$R_{a,t} = \kappa_a + e_{a,t} \tag{1}$$

$$R_{f,t} = \kappa_f + e_{f,t} \tag{2}$$

where κ_a, κ_f are positive constants with $\kappa_a > \kappa_f$, and $e_{a,t}$, $e_{f,t}$ are white noise variables with mean zero and precision ω_a, ω_f respectively.[3] The total excess of market-risk payoff to the

[3]Precision is defined as the inverse of the variance of a random variable.

investors equals

$$TEP_{CAPM,t} = q_{a,t}R_{a,t} + q_{f,t}R_{f,t} - C_a(q_{a,t}) - C_f(q_{f,t}) - bq_t, \tag{3}$$

with $q_{a,t}$ and $q_{f,t}$ being the amount invested in the active strategy and the factor strategy respectively at the beginning of period t, b being the fix proportional management cost, and

$$q_t = q_{a,t} + q_{f,t} + q_{i,t}, \tag{4}$$

defining the total assets under management. The difference between the total assets under management and the sum of the investments in the active and the factor strategy is defined as $q_{i,t}$, denoting the amount which is indexed. Indexing comes along with zero expected CAPM excess return and zero cost, therefore it is not part of Equation (3). The costs $C_a(q_{a,t})$ and $C_f(q_{f,t})$ of running the active and the factor strategy are defined as

$$C_a(q_{a,t}) = a_a[q_{a,t}]^{p_a}, \tag{5}$$

$$C_f(q_{f,t}) = a_f[q_{f,t}]^{p_f} \tag{6}$$

with $a_a, a_f > 0$, $p_f > 1$, and $p_a > p_f$. Thereby, investing in either one of the two strategies is associated with positive cost which is convex in the amount invested in the respective strategy. Also, note that due to $p_a > p_f$ the cost of investing in the active strategy increases faster compared to the cost of the factor strategy (for large amounts invested).

Dividing Equation (3) by the total amount invested q_t leads to a relative excess of market return for investors of

$$\alpha_{CAPM,t} = \frac{q_{a,t}}{q_t}R_{a,t} + \frac{q_{f,t}}{q_t}R_{f,t} - \frac{[C_a(q_{a,t}) + C_f(q_{f,t})]}{q_t} - b. \tag{7}$$

2.2 Manager's Choice of the Investment Strategy

Facing the decision problems of the agents within the model, we start off looking at the manager's perspective (step 3), who has to decide upon his allocation of the received flows across the three strategies, thereby maximizing his expected revenue bq_t. He does so by maximizing the expected α_{CAPM}. This is a direct consequence of the assumption that investors, whose actions are described in the next paragraph, are chasing α_{CAPM}. For simplicity, we assume that the manager does know κ_a and κ_f from Equations (1) and (2) as well as a_a, a_f, p_a, and p_f from Equations (5) and (6).[4] Assuming that the total investment amount q_t is large enough to make the maximization problem unbounded in this dimension (i.e. $q_{a,t} + q_{f,t} \leq q_t$), the expected CAPM excess return to investors from the manager's perspective in the optimum is characterized by

$$E(\alpha_{CAPM,t}|K_M) = \frac{\hat{q}_a \kappa_a}{q_t} + \frac{\hat{q}_f \kappa_f}{q_t} - \frac{[C_a(\hat{q}_a) + C_f(\hat{q}_f)]}{q_t} - b, \tag{8}$$

with

$$\hat{q}_a = \left[\frac{\kappa_a}{p_a a_a}\right]^{\frac{1}{p_a - 1}}, \tag{9}$$

$$\hat{q}_f = \left[\frac{\kappa_f}{p_f a_f}\right]^{\frac{1}{p_f - 1}} \tag{10}$$

denoting the optimal amount invested in either strategy and K_M being the manager's information set.[5] Optimal amounts are characterized such that the manager invests in either strategy up to the point where the expected return of the last invested dollar equals the marginal cost of the respective strategy. Any additional funds are indexed at zero cost and zero expected excess return.

[4] None of our core results change if we let them Bayesian-learn these parameters using past results.

[5] We neglect the case $q_t \leq \hat{q}_a + \hat{q}_f$ for simplicity here. If we did not, a simple rule that the manager primarily invests in the strategy with higher marginal net return would be the most rational assumption.

2.3 The Investors' Problem

The investors, on the other side, have to decide how much to invest in the fund. They are not aware of the manager's factor investing option, hence they have a coarser information set and assume the return generating process to be

$$R_{af,t} = \kappa_{af} + e_{af,t}, \tag{11}$$

with $e_{af,t}$ being normally distributed with zero mean and precision ω_{af}. Investors do not know κ_{af} and a priori assume it to be normally distributed with mean ϕ_0 and precision γ_{af}. In line with the investors' knowledge about available investment strategies, they assume the cost to scale by the following function

$$C_{af}(q_{af,t}) = a_{af}[q_{af,t}]^{p_a}, \tag{12}$$

with $a_{af} > 0$ and $p_a > 1$ (as well as $p_a > p_f$, see Equations (5) and (6)). Also, investors know p_a as they have some sense about the cost scaling for (true) active investments, however they are not aware of the value of a_{af} which they instead assume to be normally distributed with unknown mean and known precision τ_{af}. Similarly as for κ_{af}, they expect the unknown mean of a_{af} a priori to be normally distributed with expected value θ_0 and precision ψ_{af}. Note that the cost function assumed by the investors, $C_{af}(\cdot)$, is of the same power-law family (linked via p_a) as the manager's active cost function, $C_a(\cdot)$, which is justified by the assumption that investors solely know about active investing and are clueless about factor investing as well as about the associated cost. Intuitively, this means that investors associate the manager's past returns with a steeper cost function then he actually is facing. This misperception leads them to conclude that the manager is more skillful than he is, which eventually results in over-investment into the fund as we will show later on.

According to their coarser information set, investors are only able to observe past aggregated non-index investments, $q_{af,t-1}$, past aggregated returns, $R_{af,t-1}$, as well as past aggregated cost, $C_{af}(q_{af,t-1})$, but they have no knowledge about the individual components, i.e by looking backwards they observe the left-hand sides of the following equations, but not their compositions on the right-hand sides:

$$q_{af,t-1} = q_{a,t-1} + q_{f,t-1}, \tag{13}$$

$$R_{af,t-1} = \frac{q_{a,t-1}R_{a,t-1} + q_{f,t-1}R_{f,t-1}}{q_{a,t-1} + q_{f,t-1}}, \tag{14}$$

$$C_{af}(q_{af,t-1}) = C_f(q_{f,t-1}) + C_a(q_{a,t-1}). \tag{15}$$

Based on this setting, the net CAPM excess return from the investors' perspective looks like

$$\alpha_{CAPM,t} = \frac{q_{af,t}R_{af,t}}{q_t} - \frac{C_{af}(q_{af,t})}{q_t} - b. \tag{16}$$

We assume positive α_{CAPM} generating opportunities to be in short supply, i.e. whenever such opportunities exist, there is always a matching demand by investors and due to the negative relationship of investment amount and alpha, the result is that those positive opportunities should be competed away immediately. Hence, in equilibrium the expectation of $\alpha_{CAPM,t}$ from the investors' perspective has to be zero

$$E(\alpha_{CAPM,t}|K_{I,t}) \stackrel{!}{=} 0, \tag{17}$$

where $K_{I,t}$ denotes the investors' information set at the beginning of period t which includes all information from the previous period $t-1$.

Using condition (17) leads to the *optimal* investment amount from the investors' perspective

$$\hat{q}_t = \left[\frac{[\phi_t]^{\frac{p_a}{p_a-1}}}{[p_a\theta_t]^{\frac{1}{p_a-1}}} - \frac{\theta_t \left[[\phi_t]^{\frac{p_a}{p_a-1}} \right]}{[p_a\theta_t]^{\frac{p_a}{p_a-1}}} \right] \frac{1}{b} \tag{18}$$

with ϕ_t and θ_t being Bayesian-updated according to the recursions

$$\phi_t = \left[\frac{\gamma_{af} + [t-1]\omega_{af}}{\gamma_{af} + t\omega_{af}} \right] \phi_{t-1} + \left[\frac{\omega_{af}}{\gamma_{af} + t\omega_{af}} \right] R_{af,t-1}, \tag{19}$$

and

$$\theta_t = \left[\frac{\psi_{af} + [t-1]\tau_{af}}{\psi_{af} + t\tau_{af}} \right] \theta_{t-1} + \left[\frac{\tau_{af}}{\psi_{af} + t\tau_{af}} \right] \left[\frac{C_{af}(q_{af,t-1})}{[q_{af,t-1}]^{p_a}} \right]. \tag{20}$$

Proofs see Appendix.

Combining Equation (19) with (9), (10), and (14) it is easy to show that in the long-run equilibrium the posterior mean of κ_{af} converges to

$$\phi^\star = \frac{\hat{q}_a\kappa_a + \hat{q}_f\kappa_f}{\hat{q}_a + \hat{q}_f}. \tag{21}$$

Looking at Equation (20) it is straight forward that the long-run posterior mean of a_{af} converges to

$$\theta^\star = \frac{a_a[\hat{q}_a]^{p_a} + a_f[\hat{q}_f]^{p_f}}{[\hat{q}_a + \hat{q}_f]^{p_a}}. \tag{22}$$

Using Equation (19) and (20) together with (18) results in the long-run equilibrium *optimal* investment amount

$$q^\star = \left[\frac{[\phi^\star]^{\frac{p_a}{p_a-1}}}{[p_a\theta^\star]^{\frac{1}{p_a-1}}} - \frac{\theta^\star \left[[\phi^\star]^{\frac{p_a}{p_a-1}} \right]}{[p_a\theta^\star]^{\frac{p_a}{p_a-1}}} \right] \frac{1}{b}. \tag{23}$$

It is important to note that the equilibrium values themselves are all expectations of – possibly very noisy – random variables, which makes it unlikely that investors effectively learn about the true costs of both the active and the factor strategy.

Solving the investors' problem using the true information set K_M would lead to a long-run optimal investment amount of

$$q^+ = \frac{\hat{q}_a[\kappa_a - a_a[\hat{q}_a]^{p_a-1}] + \hat{q}_f[\kappa_f - a_f[\hat{q}_f]^{p_f-1}]}{b}. \tag{24}$$

Proof see Appendix.

If investors chose to invest q^+ they would receive zero CAPM excess return in the long-run equilibrium. But instead they invest $q^\star$, therefore from an ex-post perspective the long-run equilibrium expected excess return is equal to

$$\alpha^\star_{CAPM} = \left[\frac{q^+ - q^\star}{q^\star} \right] b. \tag{25}$$

As $q^\star > q^+$ due to the imposed relations between κ_a and κ_f, a_a and a_f, as well as p_a and p_f, the ex-post expected long-run equilibrium excess return to investors is negative. If $q^\star >> q^+$, $\alpha^\star_{CAPM}$ approaches $-b$ which is the natural lower bound for the ex-post expected long-run excess return in the model. This negative return to investors is the main difference of our model compared to Berk and Green (2004) according to which econometricians should observe zero excess returns in the long-run equilibrium.

3 Data and Methodology

3.1 Data

We use data provided by the Center for Research in Security Prices (CRSP) and Morningstar. Our sample contains 3,292 actively managed mutual funds from the United States that primarily invest in domestic equity (>95% AUM invested in domestic stocks), covering the period from March 1993 to December 2014 on a monthly grid. We merge the two databases using an approach similar to Pástor et al. (2015), thereby trying to eliminate possible reporting errors in our sample. Specifically, we perform an inner join of the two databases looking at matching values for net returns and AUM. We only include funds that contain almost exactly the same return and AUM figures across the two databases on the fund level. This enables us to sort out database errors leading to a high data accuracy in the final sample.

To gain more intuition for our analyses we label the components of Equation (7) from the model in Section 2 as follows:

$$
\begin{aligned}
Investors'\ Alpha &= \alpha_{CAPM,t} \\
True\ Alpha &= \frac{q_{a,t}}{q_t} R_{a,t} \\
Fake\ Alpha &= \frac{q_{f,t}}{q_t} R_{f,t} \\
Operating\ Costs &= \frac{C_a(q_{a,t}) + C_f(q_{f,t})}{q_t} \\
Management\ Fee &= b
\end{aligned}
$$

Note from Equation (7) that *Investors' Alpha* is a net figure, whereas *True Alpha* and *Fake Alpha* are both gross figures, which means associated costs are not deducted.

The *Investors' Alpha* is the fund's risk adjusted net excess return with respect to the *CAPM*. To account for the possible time variation of the exposure to the systematic risk source, we estimate α_{CAPM} by applying a rolling window approach. For a specific month t, we apply the following two-stage estimation: First, we use the fund's *Net Return* data from the previous 24 month including $t-1$ to estimate the risk exposure towards *Kenneth French's* value weighted *Market Risk* Factor. Second, we take the estimated exposure and the data from month t to estimate the alpha. On average a fund within our sample has a positive *Net Return* of 73 bp per month while having a negative *Investors' Alpha* of -5 bp per month (see Table 1).

We estimate *True Alpha* in the same manner as *Investors' Alpha* but instead of using the one factor *CAPM* as risk adjustment model, we use either the three factor model by Fama and French (1993) (*FF3*) or the four factor model by Carhart (1997) (*CH4*).[6] To obtain the gross values, we take the net estimates and add the monthly *Expense Ratio*. Additionally, we calculate a Pástor et al. (2015) style *True Alpha* by simply taking the difference between a fund's *Gross Return* and the corresponding return of the benchmark assigned by Morningstar for each month. We label these Morningstar alphas with *MS*.

The *Fake Alphas* are calculated by simply taking the difference between *Investors' Alpha* and the net *True Alpha*[7]. Depending upon the risk adjustment model the average *True Alpha* lies between 1 bp and 2 bp per month whereas the average *Fake Alpha* ranges from 3 bp to 5 bp per month (see Table 1). A more detailed explanation of our alpha estimation methodology can be found in Appendix B.

A fund's size measured in end-of 2014 inflated dollars is defined as

$$Infl.\ AUM_t = AUM_t \frac{Total\ MCap\ Equity\ CRSP_{12-31-2014}}{Total\ MCap\ Equity\ CRSP_t}. \tag{26}$$

By using this definition, where we multiply the fund's AUM at t with the ratio of the total market capitalization of CRSP stocks at the end of 2014 to the total market capitalization of CRSP stocks at t, we account for variations in overall equity-market size that might be

[6]We obtain the factor data from *Kenneth French's* website as well.
[7]The net *True Alpha* is equal to the *True Alpha* minus the monthly *Exp. Ratio*.

mirrored in a fund's AUM as well. This simply means that we assume the supply of equity to be driven by the overall equity market size, and we control for that using inflated dollars instead of usual dollars. The average size of a fund in 2014 inflated dollar is $1,952 *million*. Over the entire sample the fund size varies with a standard deviation of $4,857 *million* (see Table 1) which is mainly due to severe size differences in the cross section: The average cross-sectional standard deviation of $Infl.\ AUM_t$ is $4,089 *million* compared to an average time series standard deviation of $475 *million*.

To account for these cross-sectional differences in fund size we introduce the *Standardized Size* of a fund as

$$Std.\ Size_t = \frac{Infl.\ AUM_t}{Mean(Infl.\ AUM_s)} \quad \{s \in all\ observations\ for\ the\ fund\}. \tag{27}$$

Using the $Std.\ Size_t$ instead of the absolute size prevents our panel analyses to be dominated by large funds. The mean *Std. Size* is at roughly 100% [8], which is not surprising having in mind the definition above, and it varies with a standard deviation of about 61% on a month by month basis (see Table 1).

We define the *Industry Size* in the spirit of Pástor et al. (2015) as

$$Industry\ Size_t = \frac{\sum_{all funds} AUM_t}{Total\ MCap\ Equity\ CRSP_t}, \tag{28}$$

and similarly to their work, we use it as a measure for professional competitiveness in the equity market as well. The basic idea of the estimator is that the higher the portion of equity managed by professionals (in our case represented by the managers of active managed equity mutual funds) the higher the competitiveness regarding the generation potential of excess returns. As competitiveness rises it gets harder for a manager to generate an excess return running an active strategy which is shown in a theoretical context by Pástor and Stambaugh

[8] Due to the applied winsorizing of all variables at the 1% and the 99% quantile this is only *roughly* the case

(2012).

Further, we make use of a fund's reported *Expense Ratio*, its *Management Fee* as well as its *Turnover Ratio*. For the *Exp. Ratio* and the *Mgmt. Fee* we take the reported yearly figures and divide them by 12 to obtain the monthly equivalents. We then calculate the *Operating Costs* as the difference between *Exp. Ratio* and *Mgmt. Fee* on a monthly basis. The average *Operating Costs* are 6 bp per month, the mean *Mgmt. Fee* is 5 bp on a monthly basis (see Table 1).

The number of funds in our analysis varies strongly over time. Figure 1 shows the timeline of funds that report returns (black) as well as of those we included in the following regressions where either FF3 or CH4 as benchmark is used (blue). The gap between the two lines is due to the fact that not all return reporting funds report e.g. *Mgmt. Fee* or *Exp. Ratio* as well, therefore we cannot include these observations in analyses where we require all of theses variables. Note that especially between 1995 and 2005 the number of funds in our sample strongly increases, whereas between 2005 and the end of 2014 is roughly stays at a stable level which is between 1500 and 2000 funds.

3.2 Econometric Issues

As outlined in detail by Pástor et al. (2015), estimating simple OLS coefficients for a regression model of type

$$\alpha_{CAPM,it} = a + \beta Std.\ Size_{it-1} + e_{it} \tag{29}$$

is prone to inherit an omitted-variable bias. Suppose the alpha of a fund as well as the fund's size are both correlated with the fund manager's skill. Then the estimated OLS β from equation (29) is biased, as the resulting error terms of the model are correlated with skill. The omitted-variable bias in this case arises from the neglected variable skill. Unfortunately, managerial skill is not easy to observe but we can at least control for a time-independent cross-sectional effect by running a similar regression model as above adding fund-fixed effects

$$\alpha_{CAPM,it} = a_i + \beta Std.\ Size_{it-1} + e_{it}. \tag{30}$$

Accordingly, we assume that each manager has an individual skill which is fix over the observation period. It does not completely solve the omitted-variable concern, but at least enables to correct for the time-fixed effects in the cross section.

As Pástor et al. (2015) correctly point out, introducing the fund-fixed effects however leads to another bias, the so called finite-sample bias if the parameters of the model in Equation (30) are estimated using standard OLS. They give a detailed explanation of the problem in their paper including a simulation study concerning the matter. Our goal is simply to summarize the core of the finite-sample issue to provide reason why we address this in our analyses. According to Stambaugh (1999) the bias arises in finite-samples (and by introducing the fixed effects we estimate our parameters sort of on a fund by fund level thereby extremely shrinking the sample size per parameter estimation) if the innovations of $Std.\ Size_{it}$ and the e_{it} are correlated. In our case we should expect such a correlation for the following two reasons: (1) A lucky positive return in period t (high e_{it}) also automatically increases $Std.\ Size_{it}$ (measured at the end of period t), which is due to the defined linkage of the return and the subsequent size of a fund. (2) By theory, any positive return during the period also attracts new flows leading to a higher $Std.\ Size_{it}$ at the end of it. This is as well the case if the positive return was earned by luck. Consequently, to avoid this bias we again follow Pástor et al. (2015) using a recursive demeaning approach to estimate the coefficients in the fund-fixed effect models to follow. Details of the recursive demeaning procedure can be found in Pástor et al. (2015). We exactly apply their procedure of recursively forward demeaning all our variables (dependent and independent) and then instrumenting for forward demeaned *Std. Size* by using backward demeaned *Std. Size* in the analyses to follow.

Further, we address the problem of possible correlations within the dependent variable in our further analyses: The average correlation of *Investors' Alpha* for funds within the same *Morningstar Category* is 0.33 whereas for funds of different *Morningstar Category* membership it drops to 0.06. This pattern holds for *True Alphas*, *Fake Alphas* as well as the *Operating Costs*. Hence, we calculate the standard errors of the estimators in our empirical analyses allowing for correlations within the clusters $MSCategory \times date$ and $fund$, thereby also accounting for serial correlation on the fund level possibly induced by the alpha estimation or the recursive demeaning.

4 Empirical Tests

One of the model's important pillars is the assumption that investors chase CAPM alpha. Before we test whether there is empirical evidence for our model, we need to address this assumption first. In a recent paper, Berk and van Binsbergen (2016) show that amongst all major asset pricing models it is most likely that investors form their decisions based on CAPM alpha.[9] Yet Berk and van Binsbergen (2016) use a slightly different approach estimating their alphas. In particular, they are not applying a rolling estimation window as we do, which is why we reaffirm their results using our way of estimating alphas. The results of these tests are in Appendix C and are qualitatively in line with their work, yielding that indeed the CAPM alpha is most likely guiding investors' decisions. Given theses results we focus on testing the model's main predictions in the following.

4.1 Implications Derived from our Model

If investors do not truly understand true market mechanisms while managers are well informed, the model gives the following two main implications:

Implication 1 *Decreasing Returns and Costs to Scale*

Based upon Equation (8), *Investors' Alpha*, *True Alpha*, *Fake Alpha*, as well as the *Operat-*

[9] The results of Barber et al. (2016) point towards the same direction although they use a different approach.

ing Costs should all be inversely related to the fund's size. This should hold for equilibrium as well as for non-equilibrium states.

Implication 2 *Negative Investors' Alpha in Equilibrium*
Ex-post expected *Investors' Alpha* should be negative but not exceeding the level of the *Mgmt. Fee* in the long-run equilibrium states. This is due to the outcome that under the imposed restrictions investors overinvest into the funds in the long-run equilibrium ($q^\star > q^+$). At the same time, the *True Alpha* and the *Fake Alpha* should be non-negative.

While the *Decreasing Returns and Costs to Scale* from *Implication 1* of our model are basically in line with Berk and Green (2004), the negative *Investors' Alpha* in equilibrium from *Implication 2* is clearly against their theory. However, the negative overall mean of the *Investors' Alpha* in our sample (including possible non-equilibrium states) as reported in Table 1 could be explained by their model as well. Berk and Green do not impose restrictions for the way towards the long-run equilibrium. For example, if investors start off overestimating a specific fund manager's skill in the Berk and Green model they will receive negative excess returns at the beginning of the fund's lifetime to then converge towards receiving zero excess returns in equilibrium. Hence, the average of all excess return observations for that fund still would be negative, with the level depending upon the time it takes to reach equilibrium. The fact that possibly not all observations are at equilibrium states is largely ignored by the empirical literature discussing Berk and Green (2004). In contrast, we exploit our model's equilibrium relation between age and size to carefully identify equilibrium states.

Literature discussing decreasing returns to scale, as stated in *Implication 1*, show diverging results so far. Pástor et al. (2015) point out that most of the existing work on estimating the size-performance relation is likely to be biased, and consequently they derive an unbiased estimator by applying the recursive demeaning procedure mentioned in Section 3.2. Using this estimator they find that a fund's size is negatively correlated to its gross alpha. However, the observed negative relation is not very robust, especially if *Ind. Size* as competitiveness measure is considered as well. We make use of their recursive demeaning procedure and provide first evidence regarding the fund size-performance relation with respect to individual alpha components, i.e. *Investors Alpha* (net), *True Alpha* (gross) and, *Fake Alpha* (gross). We also reexamine the fund size/gross alpha relation using a standardized size measure to facilitate cross-sectional comparisons.

4.2 Decreasing Returns and Costs to Scale

According to *Implication 1* the model yields negative relations of the different alphas and the costs with the size of a fund. We will empirically test theses relations starting off with the *Investors' Alpha*, then looking at different estimators for the *True Alpha*, and examining the *Operating Costs* last. In each of the tests we perform a regression using one of the alphas or the costs as dependent variable and *Std. Size* along with several control variables as independent ones. First, we control for *Ind. Size* as a measure for market competitiveness. Second, *Age* is included as we expect a fund's *Age* to be correlated with its size if the investors' prior parameter expectations deviate from the posterior equilibrium ones. Third, we include *Mgmt. Fee* as a control variable to asses possible performance boosting in exchange for management income. Fourth, we consider the *Turn. Ratio* as well, to account for portfolio rebalancing. Lastly, we control for the *SMB*, the *HML*, and the *MOM* factor returns to account for time variation in risk premia, i.e. variations of the supply side of the *Fake Alpha*.

For each model the slope coefficients are estimated using two different samples: One including all available observations and one in which observations with $Age < 4\ years$ are excluded, thereby accounting for a possible incubation bias.[10] For the *Investors' Alpha*, the *True Alpha*, and the *Operating Costs* regressions we use the recursive demeaning procedure on the fund level. As the average level of *Fake Alpha* for a fund primarily varies between funds of different Morningstar Categories and only exhibits a much lower

[10]This is in line with Evans (2010) who finds that removing the first three years of data eliminates the incubation bias.

variation for funds belonging to the same Morningstar Category (i.e. 0.25% compared to 0.16% in the *CH4* specification), we use fixed effects on the Morningstar Category level in the *Fake Alpha* regressions instead of the fixed effects on the fund level we apply in the other models. The *Fake Alpha* regressions are therefore less prone to inherit a finite-sample bias, as we have more observations within each *fixed-effect* category (see Pástor et al. (2015) for details). This is the reason why we use plain OLS estimators (including the fixed effects on the Morningstar Category level) in the *Fake Alpha* regressions.

Investors' Alpha

Results for the regression models with *Investors' Alpha* as dependent variable are presented in the first two columns of Table 2. Model (1) includes all observations whereas observations with $Age < 4$ *years* are excluded in model (2). In the specification with all observations, the estimated slope coefficient for *Std. Size* is -0.26, which yields that as a fund grows from zero (0%) to its mean size (100%) the *Investors' Alpha* on average declines by 26 bp per month. With a t-statistic of -6.66 the negative relation is statistically significant. The decreasing returns to scale relation is not only statistically but also economically significant. For example, a one-standard-deviation increase in the *Std. Size* leads to a sizable decline in *Investors' Alpha* of about 16 bp per month. These findings clearly support the first part of *Implication 1* regarding the negative relationship between *Investors' Alpha* and the size of a fund.

Looking at the controls, the *Industry Size* has no significant relation with the *Investors' Alpha*. While we observe a negative relation between a fund's *Age* and *Investors' Alpha* with a slope coefficient of -3.01 and a t-statistic of -1.83. In terms of our model, the the negative *Investors' Alpha*/*Age* relationship goes well with investors that start off investing $q_t < q^\star$ on average at the time the funds get issued. Note that the slope coefficient declines to -2.67 with a t-statistic of -1.61 in model (2). Thus, the learning effect seems to be most severe amongst the very young funds.

The *Mgmt. Fee* and *Investors' Alpha* are negatively connected in model (1) but the slope coefficient drops to about one third of its size in model (2), which gives reason to believe that there might be an incubation bias: Managers set a low *Mgmt. Fee* at the very beginning of a fund's lifetime, thereby attract new investors, then quickly raise the *Mgmt. Fee* (within the first 4 years), and leave it on a constant level for the remainder.

For the *Turn. Ratio* we observe a significantly positive relation with *Investors' Alpha*, indicating that managers only trade when they expect positive net outcome. If the entire portfolio is reallocated over the period of a year (*Turn. Ratio*=100%) investors on average receive about 10 bp more alpha on a monthly basis compared to the case when there is no reallocation at all (*Turn. Ratio* =0%). In all cases the risk premia are positively correlated with *Investors' Alpha*, which implies that the average fund loads positively on all three factors.

True Alphas

Next, we examine the regression models with *True Alpha* as dependent variable for which the results are presented in models (3) to (8) of Table 2. We run the regressions for the different versions of *True Alpha* introduced in Section 3.1, using either $FF3$, $CH4$, or MS as the benchmark for the full sample and the sample where we exclude the young funds respectively. Throughout all specifications we observe a significant negative relationship between *True Alpha* and *Std. Size* with slope coefficients ranging between -0.20 and -0.10 and t-statistics between -7.96 and -3.63. Looking at model (6) for example, where the benchmark is $CH4$ and we exclude young observations, a rise in *Std. Size* from 0% to 100% yields an average decline in monthly *True Alpha* of 17 bp. These results strongly support the second part of *Implication 1* regarding the negative relation between the *True Alpha* and the size.

Examining the controls, we observe the relationship between *Ind. Size* and *True Alpha* to be mainly negative but usually insignificant. Model (5) depicts the only exception where we obtain a slightly significant negative relationship between *Ind. Size* and *True Alpha* with a slope coefficient of -7.53. This means that as industry size rises by 1% point the $True\ Alpha_{CH4}$ drops on average by about 8 bp per month in model (5). The relation between *Age* and the *True Alphas* is negative as well, yet only significant when we use *MS* (model (7) and (8)) as benchmark to calculate the alphas. For the *Mgmt. Fee* we obtain similar results as in the *Investors' Alpha* case, always having a negative relation between *Mgmt. Fee* and *True Alpha* that is significant for the models we estimate using the full sample, but insignificant whenever we exclude the young observations. The results reinforce the above evidence of an incubation bias. For the *Turn. Ratio* the regression coefficients are consistently positive and significant, hinting again at managers that only seem to trade when they expect a positive risk adjusted return. In contrast to the positive relation to *Investors' Alpha* we find that risk premia have a negative effect on *True Alpha* whenever significant (except for the Momentum premium in the *MS* specifications). We will come back to this point shortly after discussing *Fake Alpha* and *Operating Cost*.

Fake Alphas

Table 3 reports *Fake Alpha* regression results. For each specification the *Std. Size* is significantly negatively related to *Fake Alpha* with slope coefficients ranging between -0.10 and -0.03 and t-statistics between -6.42 and -2.08. Focusing again at the $CH4$ benchmark case with excluded young observations in model (4), the slope coefficient of -0.04 yields that as *Std. Size* of a fund grows from zero (0%) to its average size (100%) the monthly *Fake Alpha* decreases by 6 bp on average. This is in line with the third part of *Implication 1* regarding the negative relationship between *Fake Alpha* and the size of a fund.

Examining the controls, the relationship between *Ind. Size* and the *Fake Alphas* is positive but only significant in the $CH4$ case with slope coefficients in both specifications at about 8 and t-statistics at 2.25 and 2.40 respectively. Economically this means that if the *Ind. Size* rises by 1% point the average $Fake\ Alpha_{CH4}$ increases by 8 bp per month. Note that

Ind. Size has a standard deviation of 1.63% as you can observe in Table 1, correspondingly a one standard deviation increase in *Ind. Size* leads to an average increase in $Fake\ Alpha_{CH4}$ of about 13 bp. Also, note that we observe a positive relation between *Ind. Size* and *Fake Alphas* in general for all *Fake Alpha* specifications, whereas the relation between *True Alphas* and *Ind. Size* is rather negative. Again, we will address this issue in more detail later on.

Age is negatively correlated with *Fake Alpha*, similarly as in the *Investors' Alpha* and *True Alphas* regressions but always insignificant. We also do not observe a significant relation between *Mgmt. Fee* and *Fake Alpha* which is expected due to our alpha estimation methodology. Any alpha which is not linked to the pricing kernel is contributed to *True Alpha.* The *Turn. Ratio* and the *Fake Alphas* are positively connected, although not significant in the MS specification. This could again be interpreted as evidence that the managers only shift their portfolio if they think the shift has a positive expected gross return.

The risk premia are in general significantly positively related to the *Fake Alphas*, which is plausible, since if a specific portfolio's factor loadings do not change, higher risk premia directly lead to higher *Fake Alpha.* Nevertheless, against expectations we observe a significantly negative relation of the *MOM prem.* with the $Fake\ Alpha_{MS}$, which hints at difficulties by the Morningstar benchmark in capturing variations in risk-factor exposure.

Operating Costs

Now let us look at the regression results for *Operating Costs* in Table 4. In line with the alpha regressions, we also observe a significant negative relation between *Std. Size* and *Operating Costs* which is coherent with the last part of *Implication 1* regarding the decreasing relationship between the size of a fund and its *Operating Costs.* The magnitude of the slope coefficient in both cases is -0.01, yielding that as *Std. Size* increases from 0% to 100%, cost decreases on average by 1 bp per month.

Examining the results for the controls, we observe no significant relation between *Ind. Size* and the cost. *Age* is again negatively connected to the dependent variable, this is also the case for the *Mgmt. Fee.* The large t-statistics in the latter case are due to the definition of *Operating Costs* as they are calculated by subtracting the *Mgmt. Fee* from the *Exp. Ratio.* As expected, *Turn. Ratio* is significantly positively related with *Operating Costs*, meaning more trading comes along with higher costs.[11] We obtain tremendously higher R^2s within these regressions ($\geq$0.66) compared to the ones we get in all the alpha regressions ($\leq$0.09). This is due to the fact that we estimate the alphas using market data thereby introducing estimation noise, whereas *Operating Costs* are calculated based on reported figures only.

Discussing Risk Premia Relations

Besides the fact that *Implication 1* is generally confirmed by our empirical results, cross-comparing the slope coefficients of *True Alpha* (Table 2) and *Fake Alpha* (Table 3) is instructive: The relationship between risk premia and *True Alphas* is either insignificant or negative, whereas for the *Fake Alphas* it is the other way around. When taken together, these findings are consistent with the following storyline: Fund managers consider active and factor strategies as substitutes. Once factor premia increase, the benefits for running factor strategies increase as well (positive risk premia coefficients in Table 3). Therefore, managers focus less on the active part and *True Alpha* declines (negative risk premia coefficients in Table 2). Also if competitiveness rises, it gets harder for managers to generate excess returns running true active strategies (negative *Ind. Size* coefficients in Table 2). Thus, they lose incentives for running active strategies in more competitive times and instead they shift towards factor investing (positive *Ind. Size* coefficients in Table 3).

4.3 Addressing Equilibrium Alphas

As emphasized in *Implication 2*, the model implies negative *Investors' Alpha* in equilibrium. To test this hypothesis with our data, we first need a way to identify equilibrium

[11]Note that this does not cover cross-sectional differences in cost levels, since the *Turn. Ratio* is a measure relative to the fund's individual portfolio size.

states empirically. One of the theoretical properties of our model is helpful in this context, namely that the newly observed returns when being in the long-run equilibrium should not influence the investors' posterior believes. Hence, for long-run equilibrium states, *Age* and size of a fund should be uncorrelated. In our empirical specification we allow for multiple equilibrium states during a fund's lifetime and we label these as *EQ phases*.

To identify the *EQ phases* we first include all observations with a non-significant 24 month rolling window correlation coefficient (at the 10% level) between *Age* and *Infl. AUM* in the candidate *EQ sample*. From the candidate *EQ sample* we then exclude all observations during NBER recession periods to obtain the final *EQ sample*. In this *EQ sample* we now merge all observations of one fund that are within a window of 24 months to one *EQ phase*. We end up having N=26,789 observations merged to 4,379 *EQ phases* of 2,406 different funds in the *EQ sample*. Based on this identification, we recalculate the standardized size ($Standardized\ Size_{EQ}$) but only use observations within the same *EQ phase* for standardization. Specifically, this means that the mean of *Infl. AUM*, which is used to standardize the size of a fund, is calculated over only those observations that are included in one *EQ phase*.

Figure 2 shows the empirical distribution of the $Std.\ Size_{EQ}$ within the *EQ sample*. Restricting our empirical analysis to the *EQ sample* allows us to consider this distribution as the empirical counterpart of the optimal investment amount's distribution with $q^{\star}$ being the corresponding expected value. While Figure 2 reveals a peak around $Std.\ Size_{EQ} =$ 100% both lower and higher values also occur. In terms of our model, this variation is induced by the posterior distributions of the model parameters. Note that *Implication 2* makes statements about ex-post expected alphas and cost in equilibrium, i.e. those alphas we should observe as econometricians when focusing on equilibrium states and conditioning to investors investing the expected equilibrium amount $q^{\star}$. Therefore, we are most interested in alphas and cost corresponding to a $Std.\ Size_{EQ}$ of 100%. According to *Implication 2* the corresponding *Investors' Alpha* should be negative but bounded by the level of the *Mgmt. Fee* while we expect non-negative *True Alpha* and *Fake Alpha* at the same time. Based on our model, we can also make predictions for invested amounts below and above $q^{\star}$: For some level of the *Std.* $Size_{EQ}$ below 100% and below we expect to see positive *Investors' Alpha*, while observations at some level of the *Std.* $Size_{EQ}$ with *Std.* $Size_{EQ}$ above 100% and above should come along with a negative *Investors' Alpha* that clearly exceeds the *Mgmt. Fee* in absolute terms, as well as possibly negative *True Alpha* and *Fake Alpha*.

To empirically test the above hypotheses and examine the alphas and cost for the *EQ sample*, we subsequently group the observations in quintiles based on *Std.* $Size_{EQ}$ and then calculate the mean values for the variables under investigation for each quintile. Results are presented in Table 5. Panel A depicts the mean *Investors' Alphas* for the five quintiles with t-statistics in parentheses[12]. Our focus lies on quintile 3 as its mean *Std.* $Size_{EQ}$ matches exactly the *Std.* $Size_{EQ}$ =100% level. The average *Investors' Alpha* within this quintile is -7.21 bp per month with a significant t-statistic of -2.92. Note that the absolute level exceeds the average *Mgmt. Fee* of 4.74 bp per month (see Panel E). Still, we conclude that the first part of *Implication 2* regarding the negative *Investors' Alpha* is clearly supported by the data, even though this is not the case for the lower bound imposed by the model. For the remaining quintiles, the *Investors' Alphas* are well in line with the model as we observe positive values for quintile 1 and 2 (not significant) and negative values for quintiles 4 and 5.

Panel B reports the average *True Alphas* for the three different benchmark cases. Again, we first focus on quintile 3. For none of the three benchmark cases the average *True Alpha* is significantly different from zero which is in line with the second part of *Implication 2*. The values for the remaining quintiles are in line with the model, with the average *True Alphas* being positive for quintiles 1 and 2 and negative for quintiles 4 and 5.

Panel C shows the *Fake Alphas* for the three different benchmark cases. Within quintile 3, for the $FF3$ and the $CH4$ specification we obtain mean estimates of 3.09 bp per month and

[12]Note that due to the reduced sample size we cluster the standard errors on the fund basis in the *EQ sample*, as otherwise we would have many clusters containing a single observation only.

2.80 bp per month respectively, both with t-statistics close to being significant. As the level of these *Fake Alphas* exceeds the level of the corresponding *True Alphas* from Panel B (FF3 and CH4 both at 0.16 bp per month) this could be suggestive that the *Fake Alpha* is the more important component of the two from an economic perspective. The mean $Fake\ Alphas_{MS}$ is insignificant in quintile 3. Collectively, these results support the second part of *Implication 2* referring the non-negative *Fake Alpha* in equilibrium. Looking at the quintiles to the left and to the right we observe positive average *Fake Alphas* for quintiles 1 and 2 and negative ones for quintiles 4 and 5, which is also in line with the model.

To complete the picture, we supplement information about the average *Operating Costs* as well as the average *Mgmt. Fees* for all quintiles in Panels D and E.

Overall, based on the *EQ sample* we find strong evidence for negative expected *Investor's Alpha* in equilibrium. Also, the behavior of the *Investor's Alphas*, the *True Alphas*, and the *Fake Alphas* across size quintiles fit into the picture and clearly supports our model.

4.4 Do Investors Learn?

So far we assume that investors do not learn about the structural mistake they make, i.e. even if they receive negative excess returns in equilibrium they do not reconfigure the cost function they associate with these returns. This might be a strong assumption in particular in a model with constant unknown return distributions and cost. However, in real-world decision-making situations general market conditions are very dynamic, which means that a specific strategy to gain *True Alpha* run by a manager today might not work tomorrow. Also, risk premia vary a lot over time and on top of that there might be time variation in the cost functions as well. Thus, a more realistic economic setting would assume that investors always get fresh injections of uncertainty and are never able to pin down true distributions. The fact that only 31,273 of possibly 340,049 observations are within the *EQ sample* can be seen as empirical support.

To get an impression whether and to what extent learning is possible in a more dynamic economic setting we consider an empirical specification that allows investors to learn about their mistake once they remain in equilibrium state for a while. Think of it as a two staged process where the second stage can only be reached after completion of the first one. If such learning takes place, the prediction would be that investors leave the first stage equilibrium (as a consequence of the learning) by reducing their investments into the fund under consideration to then reach another second stage equilibrium. Given that structural learning is completed, the *Investors' Alpha* in the second stage equilibrium should be zero as second stage equilibrium then corresponds to an optimal investment level of q^+.

We first analyze whether or not such structural learning is observable in our overall data. To this end, we split the *EQ sample* in two cohorts by looking at the counter of *EQ phases* on the fund level. If a fund reaches equilibrium for the first time, the observations of the corresponding *EQ phase* are assigned to the first cohort. The observations of any following *EQ phase(s)* for that fund are assigned to the second cohort. Note that thereby observations of funds that reach only one equilibrium phase during their entire observation period are assigned to the first cohort as well. Of the 4,379 *EQ phases* identified in total in our sample, by applying this splitting rule 2,406 are assigned to the *1st stage EQ* cohort and 1,973 are assigned to the *2nd stage EQ* cohort. This corresponds to N=14,113 observations in the *1st stage EQ* cohort and N=12,676 observations in the *2nd stage EQ* cohort. To compare the *Investors' Alphas* for the two cohorts we again split observations into quintiles based on their $Std.\ Size_{EQ}$. The results are presented in Panel A of Table 6. As in the previous section our focus is on the 3rd quintile. Within this quintile *Investors' Alpha* is on average -7.00 bp amongst the *1st stage EQ* observations compared to an average *Investors' Alpha* of -6.56 bp for the observations belonging to the *2nd stage EQ* cohort. This insignificant difference from both the economical and the statistical point of view supports our assumption that on average investors seem not to learn about the structural mistake they are making, even if they have reached an equilibrium state.

However, to get a sense about the economic magnitude of the mistake investors make, it would help to observe at least some funds whose investors seem to learn when they have reached an equilibrium state. Thus, instead of examining from an overall perspective whether or not investors learn from *1st* to *2nd stage EQ*, we now check if there are at least some investors within the sample who seem to structurally learn. To be identified as a learning candidate pair of *EQ phases* we require the following two conditions on the fund level: (1) The *1st stage EQ* phase has to be left due to a negative (rolling) correlation of *Age* and size, i.e. investors of the specific fund redeemed investments at the time the *EQ phase* was left. (2) To diminish possible effects that arise from dynamic markets, the *2nd stage EQ* phase has to be entered shortly after *1st stage EQ* was left. We set the threshold to 4 years[13] after the *1st stage EQ phase* was left. Within the *EQ sample* 530 pairs of *EQ phases* fulfill these two criteria. Splitting the 1,060 phases into a cohort containing observations belonging to the *1st stage EQ* phases and a cohort that comprises the observations for the *2nd stage EQ* phases results in a size of N=2,756 observations for the *1st stage EQ* cohort and a size of N=3,553 observations for the *2nd stage EQ* cohort.

We again compare the *Investors' Alphas* around the $Std.\ Size_{EQ}$ level of 100% by looking at Panel B of Table 6 focusing on quintile 3. The *Investors' Alpha* of the *1st stage EQ* cohort is at -14.84 bp per month which is below the *Investors' Alpha* of the entire *EQ sample* in quintile 3 of -7.21 bp (see Table 5). For the *2nd stage EQ* cohort the average *Investors' Alpha* in quintile 3 is equal to -1.25 bp. With a t-statistic of -0.11 the alpha is not significantly different from zero. This evidence is consistent with the prediction of zero *Investors' Alpha* in equilibrium once the investors have learned the true cost function and hence might be an indication that the group of investors we examine here has indeed learned about its structural mistake. To get a sense about the economic magnitude of the investors' mistake we compare the average *Infl. AUMs* for each fund between the two equilibrium phases: Across all 530 *EQ phase* pairs on average investors reduce their investments by 23% (measured in 2014

[13]The results are qualitatively the same if we instead change this criteria to 3 years. Note that we cannot require a proximity of subsequent *EQ phases* of <= 2 years as due to the identification of the *EQ phases*, observations in the *EQ sample* that are closer than 2 years apart are combined in one phase.

inflated dollars) coming from *1st stage EQ* to *2nd stage EQ*.

In conclusion, two important observations emerge from this analysis. First, there is some evidence for structural learning, but only within a very small group and this small group seems to have a negligible effect on our overall sample. Second, from an in-depth analysis of this small group we can roughly estimate that the active mutual fund industry focusing the US equity market is 23% too large due to investors erroneous assessment of the managers skill.

5 Conclusion

By assuming investors who are not aware of the true market mechanisms on the one side, and managers that are able to exploit this on the other side, we have shown that negative excess returns to investors in equilibrium might be simply a consequence of rational behaving agents given the circumstances. Supporting our model, we observe significant negative relationships between each *Investors' Alpha*, *True Alpha*, *Fake Alpha*, and *Operating Costs*, with the size of a fund. Identifying and examining an equilibrium sample yields that the implied negative returns to investors in the long-run equilibrium are mirrored by empirical data: In equilibrium, investors earn negative net CAPM alphas by about the magnitude of the *Mgmt. Fee* of a fund. According to our model such a severe magnitude implies that investors tremendously overinvest into active managed mutual equity funds, since negative net CAPM alphas to investors by the level of the *Mgmt. Fee* should only be observed if the actual amount invested is by far larger than the amount investors would invest having access to the full information set ($q^{\star} >> q^{+}$). We show that on average investors do not learn about the structural mistake they are making, but we are able to identify a small group of investors that seems to learn. Observing this group's behavior we estimate that the US equity active mutual fund industry is too large by about 23% of its actual size. The model implies that the excess investment amount should not improve market efficiency too much, as it is simply indexed by the managers. Thus, the portion of money that is put into true active strategies should be unaffected by these excess investments.

To put this in a more tangible perspective consider the scenario where the actual size of the industry exceeds by far its optimal size q^+. This would roughly mean that on average 5 bp per month of the entire industry value (this is about equal to the average *Mgmt. Fee* in the *EQ sample*) are erroneously paid by investors to the managers as a consequence of their false assessment of managerial skill. But since the model surely doesn't capture all restrictions, let us look at a more conservative estimate of the overall economic magnitude by referring to the results we obtain for the small group that seems to learn about its mistake. Projecting their behavior onto the overall sample about 23% of the 5 bp per month are due to overinvestment. Taking the size of the entire US active mutual fund market[14] in 2014 which is equal to \$15.9 trillion (see Investment Company Institute (2015)), this means that within 2014 \$21.9 billion would have been *falsely* paid by investors of active managed US mutual funds to their managers. This is equal to 0.13% of the US GDP in 2014, or on average \$244 per mutual fund holder.

[14]This includes active managed mutual funds that invest in non-equity asset classes and out-of-US regions as well.

References

Barber, Brad M., Xing Huang, and Terrance Odean, 2016, Which Factors Matter to Investors? Evidence from Mutual Fund Flows, *Review of Financial Studies* 29, 2601–2642.

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, 2016, Assessing asset pricing models using revealed preference, *Journal of Financial Economics* 119, 1–23.

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

Chu, Yongqiang, David A. Hirshleifer, and Liang Ma, 2016, The Causal Effect of Limits to Arbitrage on Asset Pricing Anomalies .

Cochrane, John H., 2010, Presidential Address : Discount Rates, *Journal of Finance* 66, 1047–1108.

DeGroot, Morris H., 1970, *Optimal Statistical Decisions* (John Wiley & Sons).

Evans, Richard B., 2010, Mutual fund incubation, *Journal of Finance* 65, 1581–1611.

Fama, Eugene F., and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.

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.

French, Kenneth R., 2016 (http://mba.tuck.dartmouth.edu/pages/faculty/ken.french).

Investment Company Institute, 2015, *Investment Company Fact Book 2015* (http://www.ici.org).

Lintner, John, 1965, The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets, *The Review of Economics and Statistics* 47, 13–37.

Milgrom, Paul, and Nancy Stokey, 1982, Information, trade and common knowledge, *Journal of Economic Theory* 26, 17–27.

Mossin, Jan, 1966, Equilibrium in a Capital Asset Market, *Econometrica* 34, 768.

Pástor, Luboš, and Robert F Stambaugh, 2012, On the size of the active management industry, *Journal of Finance* 67, 431–478.

Pástor, Luboš, Robert F. Stambaugh, and Lucian A. Taylor, 2015, Scale and skill in active management, *Journal of Financial Economics* 116, 23–45.

Sharpe, William F., 1964, Capital Asset Prices: A Theory of Equilibrium under Conditions of Risk, *The journal of finance* 19, 425–442.

Stambaugh, Robert F., 1999, Predictive regressions, *Journal of Financial Economics* 54, 375–421.

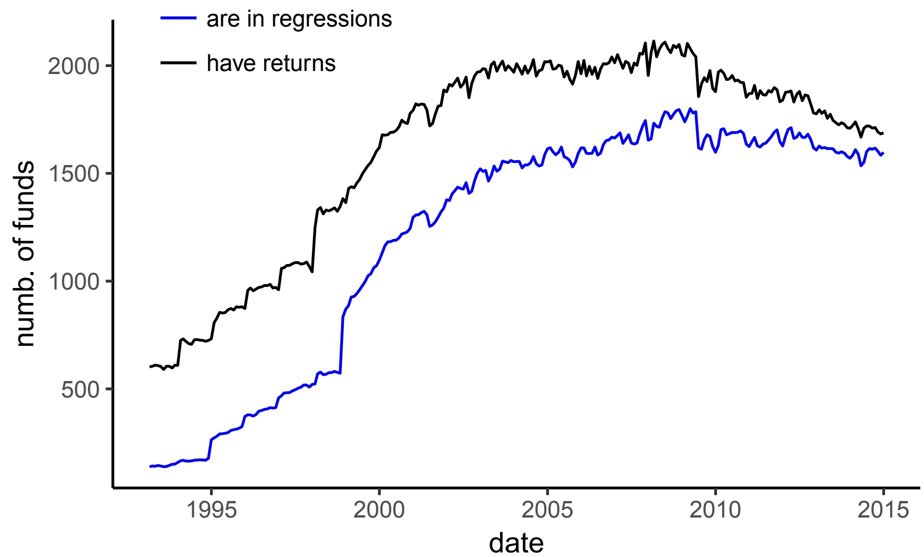

Figure 1: The black line shows the number of funds in the full sample for which we at least have a return at the given point in time. The blue line depicts the number of funds in the full sample which are included when estimating the full sample $CAPM$, $FF3$, and $CH4$ regression models in Table 2 and 3.

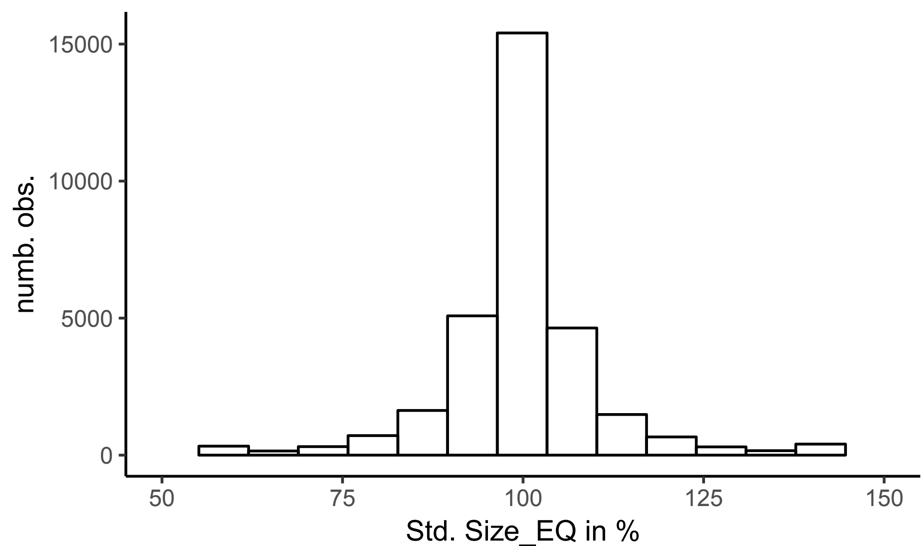

Figure 2: Absolute histogram over $Std.\ Size_{EQ}$ for the equilibrium sample. The equilibrium sample is constructed by including all observations that have a non-significant (at the 10% level) past 24 month correlation coefficient between *Age* and *Infl.* AUM, $Age > 4\ years$, and correspond to dates outside of the NBER recession periods.

Table 1: Summary Statistics

The table shows summary statistics for our sample covering US equity active managed mutual funds for the period from 1993 to 2014. *Net Return* is the return received by investors. *Investors' Alpha* is the net alpha measured using the CAPM. *True Alphas* are estimated using the *FF3* or the *CH4* model respectively, or in the *MS* case it is calculated by taking the difference between the return of the fund and the return of the corresponding *Morningstar Category Benchmark* index. For all *True Alphas* the monthly *Exp. Ratio* is added to obtain the figures. *Infl. AUM* the fund's AUM multiplied with the ratio of the entire stock market capitalization at the end of 2014 to the value of the entire stock market capitalization at the corresponding time. *Std. Size* is calculated on the fund level always taking the fund's AUM measured in 2014 inflated dollars and dividing it by the mean AUM of the fund measured in 2014 inflated dollars over its entire observation period, $Std.\ Size_{EQ}$ is calculated in the same manner as *Std. Size* but solely for assumed equilibrium states and using observations within the same *EQ phase* for standardization only. *Ind. Size* is the sum over the AUM for all active equity mutual fund AUMs for one month divided by the sum of the entire market capitalization of all CRSP Stocks within the same month.*Operating Costs* are calculated subtracting the *Mgmt. Fee* from the *Exp. Ratio*. *Turnover Ratio* is the fraction of the portfolio being reallocated over the entire year. *Flow* is the difference of AUM from one month to another correcting for the return over that period. All alphas, costs, and fees, as well as the flows are given in monthly units, only the *Turn. Ratio* is yearly. We winsorized all variables at the 1% and the 99% quantile.

Statistic	N	Mean	St. Dev.	Pctl(1)	Pctl(25)	Pctl(50)	Pctl(75)	Pctl(99)
Net Return (%)	437,251	0.733	5.023	−15.188	−2.016	1.219	3.837	13.223
Investors′ Alpha (%)	375,183	−0.050	2.199	−6.846	−1.122	−0.081	0.977	7.163
$True\ Alpha_{FF3}$ (%)	375,183	0.008	1.801	−5.570	−0.893	−0.001	0.886	5.956
$True\ Alpha_{FF4}$ (%)	375,183	0.007	1.777	−5.615	−0.877	0.004	0.879	5.790
$True\ Alpha_{MS}$ (%)	358,919	0.020	1.873	−5.975	−0.882	−0.006	0.883	6.410
$Fake\ Alpha_{FF3}$ (%)	375,183	0.050	1.552	−5.138	−0.537	0.007	0.608	5.459
$Fake\ Alpha_{FF4}$ (%)	375,183	0.050	1.688	−5.655	−0.619	0.009	0.693	5.831
$Fake\ Alpha_{MS}$ (%)	324,993	0.033	1.877	−5.935	−0.801	0.017	0.834	5.978
Infl. AUM (mill.)	427,774	1,952	4,857	2	92	375	1,444	34,212
Std. Size (%)	427,774	99.176	61.373	3.278	54.041	93.046	131.367	311.858
$Std.\ Size_{EQ}$ (%)	26,789	99.895	10.897	60.927	96.182	99.942	103.209	144.652
Ind. Size (%)	427,774	12.544	1.629	5.029	12.111	13.039	13.610	14.194
Age (years)	427,774	12.927	12.879	0.928	4.654	8.997	15.830	68.414
Mgmt. Fee (%)	351,376	0.049	0.064	−0.432	0.045	0.061	0.075	0.127
Oper. Costs (%)	351,376	0.056	0.075	0.000	0.018	0.037	0.067	0.582
Turn. Ratio (%)	411,763	83.159	72.143	3.000	34.000	64.000	109.000	411.380
Flow (mill.)	433,088	2	193	-12,079	-9	0	5	8,153

Table 2: Investors' Alpha and True Alpha Relations

The dependent variable is either the net alpha measured using the *CAPM* or the gross alpha using the *FF3*, or the *CH4* model respectively. In the *MS* case the gross alpha is calculated by taking the difference between gross returns of the fund and *Morningstar Benchmark* returns. The *Std. Size* is calculated on a fund by fund level, always taking the fund's AUM measured in 2014 inflated dollars and dividing it by the mean AUM of the fund measured in 2014 inflated dollars over its entire observation period. The *Ind. Size* is calculated by summing over the AUM for all active equity mutual funds for one month and then dividing that sum by the entire market capitalization of all CRSP within the same month. The unit of the dependent variable is always bp per month, *Std. Size*, *Ind. Size*, *Mgmt. Fee*, *Turn. Ratio*, and the risk premia are in % per month, *Age* is in years. We estimate the coefficients via a recursive demeaning approach, i.e. all variables are recursively forward demeaned on the fund level and we instrument for forward demeaned *Std. Size* by using backward demeaned *Std. Size*. We estimate each regression model using two samples, one with all observations, and another one where we exclude observations with *Age* < 4 years. Heteroskedasticity robust t-statistics based on standard error clustered by $MS\ Category \times month$ as well as by $fund$ are in parentheses.

	Dependent variable:							
	$Investors'\ Alpha$		$True\ Alpha_{FF3}$		$True\ Alpha_{CH4}$		$True\ Alpha_{MS}$	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Std. Size	−0.26	−0.25	−0.10	−0.12	−0.16	−0.17	−0.20	−0.20
	(−6.66)	(−7.50)	(−3.63)	(−4.61)	(−6.59)	(−7.35)	(−6.67)	(−7.96)
Ind. Size	−2.62	0.23	−1.38	1.10	−7.53	−5.03	−2.79	−0.62
	(−0.39)	(0.03)	(−0.31)	(0.25)	(−1.88)	(−1.25)	(−0.80)	(−0.18)
Age	−3.01	−2.67	−1.42	−0.94	−1.33	−0.92	−1.77	−1.42
	(−1.83)	(−1.61)	(−1.40)	(−0.94)	(−1.38)	(−0.96)	(−1.96)	(−1.63)
Mgmt. Fee	−67.23	−19.71	−98.26	−48.62	−80.76	−31.15	−52.52	21.64
	(−2.43)	(−0.51)	(−4.26)	(−1.36)	(−3.76)	(−0.89)	(−2.39)	(0.53)
Turn. Ratio	0.10	0.09	0.05	0.04	0.04	0.04	0.09	0.08
	(4.06)	(3.31)	(2.47)	(1.83)	(2.31)	(1.85)	(4.87)	(4.30)
HML prem.	3.77	3.40	−0.04	0.06	0.05	0.07	1.18	0.93
	(2.33)	(2.03)	(−0.05)	(0.06)	(0.05)	(0.07)	(1.22)	(1.00)
SMB prem.	11.38	10.92	−2.38	−2.01	−1.47	−1.05	−0.74	−0.88
	(5.68)	(5.47)	(−2.88)	(−2.41)	(−1.81)	(−1.28)	(−0.85)	(−1.02)
MOM prem.	0.35	0.40	−0.77	−0.89	−1.86	−2.24	3.02	2.70
	(0.45)	(0.51)	(−1.35)	(−1.60)	(−2.99)	(−3.51)	(5.61)	(5.27)
Excl. Age < 4	no	yes	no	yes	no	yes	no	yes
Observations	292,836	242,716	292,836	242,716	292,836	242,716	299,393	245,004
Adjusted R^2	0.04	0.04	0.01	0.01	0.01	0.01	0.02	0.02

Table 3: Fake Alpha Relations

The dependent variable *Fake Alpha* is calculated by first taking the difference between the *CAPM* alpha (net) of a fund and either one of the *FF3*, the *CH4* or the *MS* alpha (all gross) and then adding the monthly *Exp. Ratio* to obtain the gross figure. The *Std. Size* is calculated on a fund by fund level, always taking the fund's AUM measured in 2014 inflated dollars and dividing it by the average AUM of the fund measured in 2014 inflated dollars over its entire observation period. The *Ind. Size* is calculated by summing over the AUM for all active equity mutual funds for one month and then dividing that sum by the entire market capitalization of all CRSP within the same month. The unit of the dependent variable is always bp per month, *Std. Size*, *Ind. Size*, *Mgmt. Fee*, *Turn. Ratio*, and the risk premia are in % per month, *Age* is in years. Fixed effects are added on $MS\ Category$ level. We estimate each regression model using two samples, one with all observations, and another one where we exclude observations with $Age < 4$ years. Heteroskedasticity robust t-statistics based on standard error clustered by $MS\ Category \times month$ as well as by $fund$ are in parentheses.

	Dependent variable:					
	$Fake\ Alpha_{FF3}$		$Fake\ Alpha_{CH4}$		$Fake\ Alpha_{MS}$	
	(1)	(2)	(3)	(4)	(5)	(6)
Std. Size	−0.10	−0.08	−0.08	−0.06	−0.04	−0.03
	(−6.42)	(−5.51)	(−5.42)	(−4.35)	(−2.65)	(−2.08)
Ind. Size	3.89	4.42	7.86	8.28	4.21	4.77
	(1.21)	(1.38)	(2.25)	(2.40)	(0.95)	(1.09)
Age	−0.03	−0.02	−0.03	−0.03	−0.03	−0.01
	(−0.78)	(−0.81)	(−0.77)	(−1.02)	(−0.56)	(−0.48)
Mgmt. Fee	6.68	−0.83	3.65	−0.03	−5.06	−4.63
	(1.14)	(−0.10)	(0.63)	(−0.003)	(−0.79)	(−0.47)
Turn. Ratio	0.03	0.02	0.04	0.03	0.01	0.01
	(2.25)	(2.06)	(2.82)	(2.50)	(1.38)	(1.11)
HML prem.	4.02	3.53	4.16	3.73	2.32	2.04
	(2.61)	(2.21)	(3.06)	(2.65)	(1.17)	(1.03)
SMB prem.	13.47	12.70	12.52	11.66	13.09	12.71
	(7.58)	(7.19)	(6.79)	(6.39)	(5.70)	(5.62)
MOM prem.	0.81	1.04	2.16	2.63	−2.48	−2.18
	(1.13)	(1.48)	(2.54)	(3.00)	(−2.63)	(−2.36)
Excl. Age < 4	no	yes	no	yes	no	yes
Observations	292,836	242,716	292,836	242,716	289,256	239,376
Adjusted R^2	0.09	0.08	0.07	0.06	0.06	0.05

Table 4: Operating Costs Relations

The dependent variable *Operating Costs* is calculated by subtracting the *Mgmt. Fee* from the *Exp. Ratio*. The *Std. Size* is calculated on a fund by fund level, always taking the fund's AUM measured in 2014 inflated dollars and dividing it by the average AUM of the fund measured in 2014 inflated dollars over its entire observation period. The *Ind. Size* is calculated by summing over the AUM for all active equity mutual funds for one month and then dividing that sum by the entire market capitalization of all CRSP within the same month. The unit of the dependent variable is bp per month, *Std. Size*, *Ind. Size* and *Turn. Ratio* are in % per month, *Age* is in years. We estimate the coefficients via a recursive demeaning approach, i.e. all variables are recursively forward demeaned on the fund level and we instrument for forward demeaned *Std. Size* by using backward demeaned *Std. Size*. We estimate each regression model using two samples, one with all observations, and another one where we exclude observations with $Age < 4$ years. Heteroskedasticity robust t-statistics based on standard error clustered by $MS\ Category \times month$ as well as by $fund$ are in parentheses.

	Dependent variable:	
	Operating Costs	
	(1)	(2)
Std. Size	−0.01	−0.01
	(−13.80)	(−12.12)
Ind. Size	−0.01	−0.002
	(−0.25)	(−0.10)
Age	−0.07	−0.05
	(−7.87)	(−5.07)
Mgmt. Fee	−86.40	−70.93
	(−80.57)	(−17.88)
Turn. Ratio	0.001	0.002
	(4.09)	(5.44)
Excl. Age < 4	no	yes
Observations	304,072	248,937
Adjusted R^2	0.83	0.66

Table 5: Equilibrium Alphas and Costs

The table holds the alphas and costs for the equilibrium phases. To identify equilibrium states the 24 month rolling window correlation between *Infl.* AUM and *Age* is calculated on a fund by fund basis. All observations with a non-significant correlation at the 10% level and $Age > 4\ years$ are included in the candidate *EQ sample*. Then we drop all observations during NBER recession periods to obtain the final *EQ sample*. Observations of the same fund in the *EQ sample* that are within a window of 24 month are assigned to one *EQ phase*. To calculate the $Std.\ Size_{EQ}$ we divide the *Infl.* AUM through the mean of *Infl.* AUM for all observations within the same *EQ phase*. Then the observations are grouped in quintiles based on their $Std.\ Size_{EQ}$. The values in the table present the means over all observations for the corresponding variable-quintile combination within the equilibrium sample in bp per month. We test the null whether the mean of observations belonging to one quintile is zero. Heteroskedasticity robust t-statistics based on standard errors clustered by $fund$ are in parentheses.

Quintile	1	2	3	4	5
Quintile mean ($Std.\ Size_{EQ}$ in %)	86	97	100	102	114
Panel A			*Investors' Alpha*		
$Investors'\ Alpha$	15.31	7.94	-7.21	-32.92	-45.39
	(4.36)	(2.82)	(-2.92)	(-12.92)	(-14.11)
Panel B			*True Alpha*		
$True\ Alpha_{FF3}$	20.29	11.46	0.16	-16.29	-20.57
	(7.38)	(5.03)	(0.08)	(-7.83)	(-7.88)
$True\ Alpha_{CH4}$	19.90	11.44	0.16	-17.92	-21.94
	(7.34)	(5.16)	(0.09)	(-8.81)	(-8.59)
$True\ Alpha_{MS}$	20.87	11.67	2.02	-13.61	-19.78
	(7.41)	(5.50)	(1.01)	(-6.29)	(-7.01)
Panel C			*Fake Alpha*		
$Fake\ Alpha_{FF3}$	5.55	6.04	3.09	-5.91	-10.85
	(2.22)	(3.02)	(1.66)	(-3.00)	(-4.27)
$Fake\ Alpha_{CH4}$	5.95	6.12	2.80	-4.33	-10.82
	(2.25)	(2.94)	(1.42)	(-2.11)	(-3.98)
$Fake\ Alpha_{MS}$	4.83	5.92	1.83	-8.13	-13.14
	(1.72)	(2.44)	(0.80)	(-3.37)	(-4.27)
Panel D			*Operating Costs*		
$Operating\ Costs$	5.07	4.60	4.77	4.28	4.82
	(29.06)	(29.98)	(12.83)	(31.70)	(31.74)
Panel E			*Mgmt. Fee*		
$Mgmt.\ Fee$	5.80	5.84	4.74	5.98	5.82
	(33.67)	(45.23)	(9.20)	(54.96)	(45.68)

Table 6: Equilibrium Alphas at Different Stages

The table holds the alphas and costs for the equilibrium phases divided up in a *1st* and a *2nd stage EQ* cohort. To identify equilibrium states the 24 month rolling window correlation between *Infl. AUM* and *Age* is calculated on a fund by fund basis. All observations with a non-significant correlation at the 10% level and $Age > 4$ *years* are included in the candidate *EQ sample*. Then we drop all observations during NBER recession periods to obtain the final *EQ sample*. Observations of the same fund in the *EQ sample* that are within a window of 24 month are assigned to one *EQ phase*. To calculate the $Std.\ Size_{EQ}$ we divide the *Infl. AUM* through the mean of *Infl. AUM* for all observations within the same *EQ phase*. In Panel A we split up the entire *EQ sample* assigning the first *EQ phase* of each fund to the *1st stage EQ* cohort and any following *EQ phase* of the same fund to the *2nd stage EQ* cohort. In Panel B we identify *1st* and *2nd stage EQ* pairs by requiring that (1) the *1st stage EQ* phase has to be left due to a negative significant correlation at the 10% level of age and size of the fund, and (2) the two *EQ phases* have to be closer than 4 years apart. Within their cohort the observations are grouped in quintiles based on their $Std.\ Size_{EQ}$. The values in the table present the means over all observations for the corresponding variable-quintile combination for the respective cohort in bp per month. We test the null whether the mean of observations belonging to one cohort quintile is zero. Heteroskedasticity robust t-statistics based on standard errors clustered by *fund* are in parentheses.

	Quintile	1	2	3	4	5
Panel A				*Overall EQ sample*		
EQ stage 1	Quintile mean ($Std.\ Size_{EQ}$ in %)	85	97	100	103	115
	Investors' Alpha	20.86	4.09	-7.00	-38.72	-49.94
		(3.78)	(1.00)	(-2.01)	(-10.43)	(-10.34)
EQ stage 2	Quintile mean ($Std.\ Size_{EQ}$ in %)	88	97	100	102	112
	Investors' Alpha	9.59	10.35	-6.56	-27.20	-39.28
		(2.17)	(2.93)	(-1.90)	(-7.97)	(-9.69)
Panel B				*Learning candidates sample*		
EQ stage 1	Quintile mean ($Std.\ Size_{EQ}$ in %)	89	98	100	102	111
	Investors' Alpha	-14.38	8.55	-14.84	-44.29	-83.80
		(-1.40)	(1.11)	(-1.96)	(-5.61)	(-6.68)
EQ stage 2	Quintile mean ($Std.\ Size_{EQ}$ in %)	87	97	100	102	113
	Investors' Alpha	-0.83	14.60	-1.25	-28.39	-43.66
		(0.02)	(1.91)	(-0.11)	(-3.73)	(-4.67)

A The Model - Proofs

Proposition 1 *Observing aggregated returns* $R_{af,t}$ *investors update their expectation* ϕ *about* κ_{af} *according to*

$$\phi_t = \left[\frac{\gamma_{af} + [t-1]\omega_{af}}{\gamma_{af} + t\omega_{af}}\right]\phi_{t-1} + \left[\frac{\omega_{af}}{\gamma_{af} + t\omega_{af}}\right] R_{af,t-1}.$$

Proof. Using Theorem 1 from page 1967 in DeGroot (1970) it is straight forward to show that the posterior mean of κ_{af} is updated by the recursion in *Proposition 1.* ■

Proposition 2 *Observing aggregated cost* $C_{af}(q_{af,t-1})$ *investors update their expectation* θ *about* a_{af} *according to*

$$\theta_t = \left[\frac{\psi_{af} + [t-1]\tau_{af}}{\psi_{af} + t\tau_{af}}\right]\theta_{t-1} + \left[\frac{\tau_{af}}{\psi_{af} + t\tau_{af}}\right]\left[\frac{C_{af}(q_{af,t-1})}{[q_{af,t-1}]^{p_a}}\right].$$

Proof. Solving Equation (15) for a_{af} at t-1 leads to

$$a_{af} = \frac{C_{af}(q_{af,t-1})}{[q_{af,t-1}]^{p_a}} \tag{31}$$

Using (31) and applying Theorem 1 from page 1967 in DeGroot (1970) one can easily show that the posterior mean of a_{af} is updated by the recursion in *Proposition 2.* ■

Proposition 3 *For given ϕ_t and θ_t the optimal investment amount for investors in period t is*

$$q_t = \left[\frac{[\phi_t]^{\frac{p_a}{p_a-1}}}{[p_a\theta_t]^{\frac{1}{p_a-1}}} - \frac{\theta_t \left[[\phi_t]^{\frac{p_a}{p_a-1}} \right]}{[p_a\theta_t]^{\frac{p_a}{p_a-1}}} \right] \frac{1}{b}.$$

Proof. In a first step, the investors predict the amount $q_{af,t}$ they expect the manager to actively invest conditional on their coarser information set, $K_{I,t}$, and the overall investment q_t. To do so they basically solve the manager's problem from the investors' perspective using the rationality assumption that manager invests up to the point where the marginal expected return equals the marginal cost: $\phi_t \overset{!}{=} E(C'_{af}(q_{af,t})|K_{I,t})$. This leads to

$$E(q_{af,t}|K_{I,t}) = \left[\frac{\phi_t}{p_a\theta_t} \right]^{\frac{1}{p_a-1}}. \tag{32}$$

Inserting (32) into the equilibrium condition $E(\alpha_{CAPM,t}|K_{I,t}) \overset{!}{=} 0$, taking expectations, and moving terms leads to

$$q_t = \left[\frac{[\phi_t]^{\frac{p_a}{p_a-1}}}{[p_a\theta_t]^{\frac{1}{p_a-1}}} - \frac{\theta_t \left[[\phi_t]^{\frac{p_a}{p_a-1}} \right]}{[p_a\theta_t]^{\frac{p_a}{p_a-1}}} \right] \frac{1}{b}.$$

■

Proposition 4 *The optimal investment amount for investors having the full information set K_M is equal to*

$$q^+ = \frac{\hat{q}_a[\kappa_a - a_a[\hat{q}_a]^{p_a-1}] + \hat{q}_f[\kappa_f - a_f[\hat{q}_f]^{p_f-1}]}{b}.$$

Proof. Using the equilibrium condition $E(\alpha_{CAPM,t}|K_M) \overset{!}{=} 0$ leads to

$$E\left(\frac{q_{a,t}}{q_t}R_{a,t} + \frac{q_{f,t}}{q_t}R_{f,t} - \frac{[C_a(q_{a,t}) + C_f(q_{f,t})]}{q_t} - b|K_M\right) = 0.$$

Now taking together Equation (9) with (10) and rearranging leads to

$$\begin{aligned} q_t &= \frac{\hat{q}_a\kappa_a + \hat{q}_f\kappa_f - [C_a(\hat{q}_a) + C_f(\hat{q}_f)]}{b} \\ \Leftrightarrow q_t &= \frac{\hat{q}_a[\kappa_a - a_a[\hat{q}_a]^{p_a-1}] + \hat{q}_f[\kappa_f - a_f[\hat{q}_f]^{p_f-1}]}{b}. \end{aligned}$$

■

B Alpha Estimation

To estimate the alpha for a fund at one specific point in time t, we use the following two stage approach: First, we estimate the risk exposure of the fund regarding the market factor via the linear regression model

$$r_s - r_{rf,s} = \rho_{CAPM,t-1} + \beta_{Mkt,t-1} Mkt_s + \epsilon_s \qquad \forall s \in [t-w, t-1], \tag{33}$$

using data over the observation window of length w months preceding t with r_{t-1} being the return received by investors holding the fund over period $t-1$ to t, and $r_{rf,t-1}$ representing the return of the risk-free asset over the same period of time. Second, we use the estimated $\beta_{Mkt,t-1}$ from step one to calculate the risk adjusted excess return of the fund i with respect to the $CAPM$ over the period from $t-1$ to t as

$$\alpha_{CAPM,t} = r_t - r_{rf,t} - \beta_{Mkt,t-1} Mkt_t. \tag{34}$$

By excluding data from t in step one we prevent correlation of the errors in α_{CAPM} and β_{Mkt}.

In general, for each t we use monthly data of the previous 24 months including $t-1$ ($w = 24m$) for the estimation of step one. We then take the observations from t to calculate the alpha in step two. Relying solely on this procedure would imply that we couldn't analyze returns of funds with an age below 25 months. To also include these young observations, we use daily (working day) data starting off with window length $w = 40wd$ and step by step widen the estimation window up to $w = 500wd$ for the estimation and then switch to the monthly grid as soon as we reach the 24 month ($= 500wd$). We accumulate the daily alphas to a monthly figure to make them comparable with the monthly estimates. As a result we obtain an enlarged sample with alpha estimates starting whenever we have at least 40 working days of observation data for a fund. For each $t < 25m$ the returns are estimated using daily data, for the remainder they are based on the monthly data set.

We estimate *True Alpha* in the same manner as *Investors' Alpha* but instead of using the *CAPM* as risk adjustment model in Equations (33) and (34) we use either the three factor model by Fama and French (1993) (*FF3*) or the four factor model by Carhart (1997) (*CH4*). Additionally, we calculate a Pástor et al. (2015) style *True Alpha* by simply taking the difference between a fund's *Gross Return* and the corresponding return of the benchmark assigned by Morningstar for each month. This means that estimation of one or multiple betas as in Equation (33) is not necessary, instead the beta towards the assigned benchmark is always assumed to be equal to one. Thus, the procedure reduces possible estimation noise on the one hand but might lack accuracy one the other. We label these Morningstar alphas with *MS*. To obtain the gross values we use the net estimates and add the monthly *Exp. Ratio*.

The *Fake Alphas* are calculated by simply taking the difference between *Investors' Alpha* and the net *True Alpha*. The net *True Alpha* is equal to the *True Alpha* minus the monthly *Exp. Ratio*.

C The Basis of Investors' Investment Decisions

In Section 2 we presume that investors chase CAPM alpha when making investment decisions. We apply the method of Berk and van Binsbergen (2016) in order to show that investors actually base their capital allocation decisions on the CAPM.

Berk and van Binsbergen (2016) assume that investors hunt for positive net present value (NPV) opportunities and eliminate them by submitting buy or sell orders. Positive NPV opportunities are misspricings in the context of a specific asset pricing model and thus they depend on model choice. The idea of Berk and van Binsbergen (2016) is to compare misspricings with investor reactions thereafter, in order to determine which model's misspricing triggers an investor reaction. By observing the reaction to a certain asset pricing model's misspricing, Berk and van Binsbergen (2016) conclude that the investor is using that particular model in which the misspricing occured. For example, from net buy orders in response to positive NPV opportunity with respect to the CAPM? We would infer that investors are using the CAPM to price their investments. To test different models, it is thus necessary to identify misspricing and to monitor investors' reactions thereafter. Especially the second part is difficult as most markets are so competitive that only the price change itself can be observed.[15] Berk and van Binsbergen (2016) exploit that the price of a mutual fund, i.e. its Net Asset Value (NAV), is fixed based on the value of its underlying assets. Thus, adjustments have to be realized via volume, i.e. fund flows. Under the assumption that a particular asset pricing model holds, Berk and van Binsbergen (2016) rely on the main insight from Berk and Green (2004) to show that positive abnormal mutual fund returns are equal to positive NPV opportunities. By looking at abnormal returns in the context of a specific asset pricing model and observing subsequent flows, it is hence possible to investigate which model investors use for their capital allocation decision. The approach can also be used to judge naive models in which investors ignore risk factor exposure and simply chase

[15] See for example Milgrom and Stokey (1982).

any outperformance relative to some benchmark.

We calculate the two model ingredients as follows. First, we use the two stage approach from Appendix B in order to calculate abnormal returns for different asset pricing models. Specifically, we deploy the rolling-window regressions for the CAPM as well as multi-factor models (see Equation (34) and (33) for details). For the naive models without any beta, the abnormal return is calculated linearly with a plain subtraction. In order to keep notations simple in the following, alpha / α_M refers to all tested measures, i.e. abnormal model returns, excess returns in the context of naive models as well as raw returns. Second, we calculate mutual fund flows by comparing subsequent assets under management and correcting for the funds' return, in order to counteract the possibility that a fund creates "inflows" just by having a positive return, as follows:

$$flow_t = q_{t-m} - q_t(1 + r_{t-m}), \tag{35}$$

where q is the funds total assets under management and r is the funds return over the observation horizon from $t - m$ to t. As before, we suppress a fund's index i in order to facilitate notations. Using both flows and alphas, we then run the single univariate panel regression:

$$\Phi(flow_t) = x + z\Phi(\alpha_{M,t}) + u_t, \tag{36}$$

in which $\Phi(\cdot)$ is a function that maps real values onto {-1,0,1} according to a variable's numeric value and the alpha in this setting corresponds to an accumulated alpha over the respective observation horizon m. While Berk and van Binsbergen (2016) test their model using various horizons (m) for the calculation of flows and alphas, we focus on a 3 months horizon simply due to the fact that results for other horizons do not provide any additional insights.

Table 7: Relation between Abnormal Returns and Fund Flows

This table reports estimates for the panel regression from Equation (36) for different asset pricing models M. For ease of interpretation, the table reports $(z+1)/2$ in percent, which is equivalent to $(Pr[\Phi(flow_t = 1 \mid \Phi(\alpha_{M,t}) = 1)] + Pr[\Phi(flow_t = -1 \mid \Phi(\alpha_{M,t}) = -1)])/2$ as shown in Berk and van Binsbergen (2016). The table comprises a different asset pricing model in each row, starting with the CAPM using the CRSP value-weighted index as the market portfolio in the first row. The three subsequent rows report results for the Fama-French three-factor model from Fama and French (1993) and its cousin, the Carhart four-factor model from Carhart (1997) as well as the risk correction proposed in Pástor et al. (2015) using a fund's Morningstar benchmark. Finally, the last three lines report the results for the fund's actual return, the fund's return in excess of the risk-free rate and the fund's return in excess of the return on the market as measured by the CRSP value-weighted index. The maximum number (the best performing model) is shown in bold face.

	$(z+1)/2$ in %
	Market Model
CAPM	**54.30**
	Multi-factor Models
Fama-French 3F (FF3)	53.33
Carhart 4F (CH4)	53.02
Morningstar Benchmark (MS)	52.73
	No Model
Return	53.04
Excess Return	52.88
Excess Market Return	53.64

Table 7 presents the result from regression (36), whereby the regression parameter z is scaled to $\frac{z+1}{2}$ and expressed in percent. As shown in Berk and van Binsbergen (2016), the scaled parameter is basically the fraction of decisions for which outperformance implies capital inflows and underperformance implies capital outflow. In other words, it is the average probability that the sign of the alpha predicts the sign of the flow.[16] So if flow and outperformance are completely unrelated, the scaled parameter amounts to 50%. In contrast, if outperformance is a perfect prediction for flow, the scaled beta is equal to 100%. You can see in Table 7 that the CAPM ranks first with a value for the scaled parameter of 54.30. Thus, investors seem to rely on the CAPM for investment decisions. The result is consistent with findings from Berk and van Binsbergen (2016) as well as Barber et al. (2016),

[16]See Berk and van Binsbergen (2016) for proof of $\frac{z+1}{2} = \frac{Pr[\Phi(flow_t=1|\Phi(\alpha_{M,t})=1)]+Pr[\Phi(flow_t=-1|\Phi(\alpha_{M,t})=-1)]}{2}$.

both depicting the CAPM as the main asset pricing model used by investors. Finally, a large part of flows remains unexplained and shows that investors appear to use other criteria as well. To assess whether the ranking of the different models depicts statistically significant differences, we deploy a further statistical test from Berk and van Binsbergen (2016) in Table 8 and conclude that the overall victory of the CAPM is statistically profound.

Table 8: Test of Statistical Significance

This table displays the test of the statistical significance of the ranking of the different asset pricing models. Specifically, the coefficient and double-clustered (by fund & date) t-statistics of the univariate regression of signed flow on signed outperformance (see Equation (36) is shown in the first two columns. The rest of the table shows the statistical significance of $\zeta_1 > 0$ from the pairwise regression tests: $\Phi(flow_t) = \zeta_0 + \zeta_1(\frac{\Phi(\alpha_{c,t})}{var(\Phi(\alpha_{c,t}))} - \frac{\Phi(\alpha_{d,t})}{var(\Phi(\alpha_{d,t}))}) + \xi_t$, in which $\zeta_1 > 0$ if and only if model c is a better approximation of the true asset pricing model than d. Hence, the double-clustered t-statistic of the test that the model in the row is a better approximation of the true asset pricing model than the model in the column is shown in each entry. In particular, that means that each entry tests $z_{row} > z_{column}$ on its statistical significance. The table (both rows and columns) is ordered so that the best performing model, according to the highest value for z, is on top. The underlying derivation and proof can be found in Berk and van Binsbergen (2016).

Model	z	t-statistic	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) CAPM	0.086	13.60	0.00	2.89	3.26	2.48	4.20	2.87	5.51
(2) Ex. Market	0.073	11.55	-2.89	0.00	0.98	1.40	1.93	1.82	3.86
(3) FF3	0.067	13.07	-3.26	-0.98	0.00	0.54	2.32	0.86	2.67
(4) Return	0.061	7.53	-2.48	-1.40	-0.54	0.00	0.04	2.54	0.67
(5) CH4	0.060	12.93	-4.20	-1.93	-2.32	-0.04	0.00	0.28	1.20
(6) Ex. Return	0.058	6.98	-2.87	-1.82	-0.86	-2.54	-0.28	0.00	0.31
(7) MS	0.055	11.60	-5.51	-3.86	-2.67	-0.67	-1.20	-0.31	0.00

In replicating the approach from Berk and van Binsbergen (2016), we differ slightly in the calculation of both flow and alpha. First, while the authors use a vanguard benchmark based on a fund's style as the return correction in the calculation of a fund's flow, we stick with the fund's own return as we think this is a more accurate way (see Equation 35). In particular, we opine that using an external benchmark leads to embellished regression results as an outperformance of this benchmark creates an *artificial* inflow and is likely to be correlated with the fund's alpha. Thus, regression results would be spuriously boosted towards one. This difference is seen in our somewhat weaker yet still significant results than the ones shown in Berk and van Binsbergen (2016). Second, we calculate alphas based on rolling windows and hence differing beta exposures, whereas Berk and van Binsbergen (2016) calibrate the betas over the lifetime of a fund and thus assume that the style of the fund does not change.

All in all, based on the approach of Berk and van Binsbergen (2016), mutual fund investors seem to favour the CAPM when it comes to investment decisions. Given the shortcomings of the CAPM depicted in papers such as Fama and French (1993) and Carhart (1997), this result seems somewhat surprising. However, the behavior benefits mutual fund managers, as it allows them to grow their fund by creating a CAPM alpha which can in turn be done by loading on additional systematic risk factors such as value, size, or momentum. Thus, our model or more specifically our assumption about the goal of a mutual fund manager to maximize the CAPM alpha in order to maximize his assets under management from Section 2 seems to be empirically justified.

SFB 649 Discussion Paper Series 2017

For a complete list of Discussion Papers published by the SFB 649, please visit http://sfb649.wiwi.hu-berlin.de.

001 "Fake Alpha" by Marcel Müller, Tobias Rosenberger, Marliese Uhrig-Homburg, January 2017.

SFB 649, Spandauer Straße 1, D-10178 Berlin
http://sfb649.wiwi.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".