Marketing Mutual Funds

Preliminary - do not cite without permission

Nikolai Roussanov*, Hongxun Ruan†, and Yanhao Wei‡

April 16, 2017

Abstract

Marketing expenses constitute a large fraction of the cost of active management in the mutual fund industry. We investigate the role of these costs on capital allocation and on returns earned by mutual fund investors by estimating a structural model of costly investor search and fund competition with endogenous marketing expenditures. We find that marketing is as important as performance and fees in determining fund size. Restricting the amount that can be spent on marketing substantially improves investor welfare, as more capital is invested with passive index funds and price competition drives down fees on actively managed funds. Average alpha increases as active fund size is reduced.

Keywords: mutual funds, distribution costs, broker commissions, performance evaluation, capital misallocation, investor welfare, financial regulation, structural estimation, search costs, information frictions, household finance
JEL codes: G11, G28, D14, M31

*The Wharton School, University of Pennsylvania and NBER
†The Wharton School, University of Pennsylvania
‡University of Southern California

1 Introduction

In 2016, active mutual funds in the U.S. managed a total of 11.6 trillion dollars, an amount comparable to the nation's GDP. This industry's revenue is on the order of $100 billion, with over one third of this amount representing expenditures on marketing, largely consisting of sales loads and broker commissions (known as 12b-1 fees). Although the relation between mutual fund size, performance, and fees has been actively debated in the academic literature (e.g. Berk and Green 2004, Chen et al 2004, Pástor and Stambaugh 2012, Berk and van Binsbergen 2015 , Pástor, Stambaugh and Taylor 2015), the contribution of marketing and distribution expenditures to steering investors into particular funds, and their resulting impact on the allocation of capital is not fully understood. While the literature documents a positive relationship between distribution costs and fund flows (e.g., Gallaher, Kaniel, and Starks 2006, Bergstresser, Chalmers, and Tufano 2009, Christoffersen, Evans, and Musto 2013), it is hard to assess their importance for capital allocation and investor welfare without a structural model. Is marketing a purely wasteful rat race, or does it enable capital to flow towards more skilled managers? If both effects are present, which one is quantitatively dominant?[1]

We study the role of marketing on capital allocation in the mutual fund industry, both within the universe of active funds and between active vs. passive (index) funds. We start with the benchmark model of Berk and Green (2004), which describes the efficient allocation of assets to mutual funds in a *frictionless* market. By estimating the model, we document substantial differences between the efficient allocation and the observed distribution of fund size. To explain these differences, we introduce information frictions by generalizing the search framework developed in Hortaçsu and Syverson (2004). In our model we allow the funds' marketing activities, as well as exogenous characteristics, to affect their inclusion in the investors' information set. In our setting, both the expense ratios (fees paid by investors) and the marketing/distribution costs (components of these fees related to broker compensation) are endogenous choices of each fund. By estimating

[1]This trade-off is apparent in the regulatory framework guiding mutual fund expenses: concerned with the amount of marketing expenditure and its potential impact on investor welfare, SEC currently restricts these distribution fees to be less than 1% of fund TNA.

the search model, we find that marketing expenses are as important as price (i.e., expense ration) or performance (i.e., manager skill estimated based on historical returns) for explaining the observed variation in fund size. Further, our counterfactual analysis indicates that tightening of the SEC restriction on marketing would cause capital allocation within the active fund sector to become less efficient, as lesser-known but highly skilled funds struggle to attract flows. At the same time, the overall allocation would become more efficient as more investors would choose (more prominent) passive funds, increasing the overall investor welfare.

We follow Hortaçsu and Syverson (2004) and model the impediments to investor's ability to allocate capital optimally across mutual funds as a search friction. This approach is intuitive at least when applied to retail investors: the task of choosing among thousands of funds can be daunting even for the most sophisticated investors. Since mutual funds spend considerable resources on marketing, it is reasonable to assume that by doing so they are able to influence the likelihood of being picked by investors. In our model investors conduct costly search to sample mutual funds to invest in every period. Investors care about the fund's performance and the expense ratio charged by the fund. Mutual fund performance is determined by managerial skill as well as the impact of decreasing returns to scale. Mutual funds choose expense ratio and the marketing expenses. The marketing expenditure can increase the fund's probability of being sampled but decreases its profit margin.

We structurally estimate our model using the data on well-diversified U.S. domestic equity mutual funds, which we assume to be representative of the industry as a whole. Our estimation results reveal sizable information frictions in the mutual fund market. The average investor implicitly incurs a cost of 39 basis points to sample an additional mutual fund. This friction's magnitude is about 2/3 of the mean annual gross alpha in our data sample. The large magnitude of the estimated search cost is a manifestation of the asset misallocation problem that we documented before. The intuition is simple: high search costs prevent investors from sampling more funds. Less intensive search leads to an inferior allocation. In comparison, Hortaçsu and Syverson (2004) find the mean search

cost for an average S&P 500 index fund investor is between 11 to 20 basis points. Our higher estimated search cost indicates that asset misallocation problem is more severe in mutual fund industry as a whole (including both active funds and passive funds) than it is within the S&P 500 index funds sector.

Our estimates imply that marketing via broker incentives is relatively useful as means of increasing fund size. On average, a one basis point increase in marketing expenses leads to 1% increase of fund's size. This effect is heterogeneous across funds. For high-skill funds, a one basis point increase in marketing expenses leads to a 1.15% increase of fund's size, while for low-skill funds a one basis point increase in marketing expenses only leads to 0.97% increase in fund size. This result is intuitive: since, conditional on being included in an investor's information set, a high-skill fund is more likely to be chosen by the investor, such funds benefit more from a higher probability of being sampled than low-skill funds. We find that marketing expenses alone can explain 10% of the variation in mutual fund size; this explanatory power is comparable to both fund manager skill and fund price.

We use our model to quantitatively study the importance of marketing expenses and search costs in shaping the equilibrium distribution of fund size and investor's welfare. We conduct three counterfactual experiments. First, we explore whether tightening the regulatory constraint on marketing could reduce allocational efficiency. The premise is that restricting marketing would steer investors from high-skill but "hard to find" funds to lower skill but "easy to find" funds. We find that, under some parameterizations, lowering the regulatory limit from 100 bp to 0 bp reduces the correlation between the model-implied and the "efficient" (in the Berk and Green sense) allocation, decreasing total value added by 2.53 billion dollars (using the measure of Berk and van Binsbergen 2015). This result shows that within the actively-managed sector, more marketing could potentially lead to a better allocation. And under this parameterization, total welfare decreases with a lowering of the regulatory limit. This result shows that the gain from allocation efficiency could potentially dominate the cost of marketing.

Next, we simulate the impact of preventing funds from doing any marketing using the

estimated model parameters. We find that if the cap on marketing is set to be zero the mean expense ratio drops from 160 bp in the current equilibrium to 83 bp. Interestingly, funds lower their prices by more than the original amount of marketing expenses. The observed average distribution cost is 62 basis points, but in the no-marketing equilibrium the average fund price drops by 77 basis points. This indicates that restricting funds from competing on non-price attributes (e.g. marketing) could significantly intensify price competition. We also find the total share of active funds drops from 74% to 68%. This drop is accompanied by an increase in average fund performance as measured by mean gross alpha. The increase in alpha is due to the effect of decreasing returns to scale on fund performance. In the no-marketing equilibrium, the "index fund" takes up the market share lost by the active funds. The total investor welfare increases by 57%. Three factors contribute to this increase: in the no-marketing equilibrium, (i) active funds are cheaper, (ii) active funds' alpha is on average higher, and (iii) more investors invest in the index fund, which is a better option than a large fraction of active funds that have low skill level. In order to further understand the large increase in investor's welfare, we examine the cross-section of investor search costs implied by our model. Naturally, high search cost investors search less and pay higher expense ratios than those with low search costs, while the funds they invest in have high marketing fees and lower alphas. Comparing the investors' welfare in the two equilibria, we show that the bulk of the welfare gain of eliminating marketing is driven by high search cost investors. The intuition is simple: the high search costs investors are the investors who invest with the worst funds (unless they are lucky to "find" the index fund). In the no-marketing equilibrium, even the worst funds are much cheaper than in the current equilibrium. This leads to a significant welfare gain for the high search cost investors.

Last, we examine the impact of search cost on equilibrium market outcomes. With the emergence of Internet, advancement in search technologies (e.g., Google), and more transparent comparison (e.g., Morningstar), we would expect the search frictions to drop in the future. In other words, investors should find it easier to sample mutual funds with the help of new technologies. In this counterfactual, we set the mean search cost to 35bp

and 20bp respectively. Given new search cost, funds reoptimize their prices and marketing expenses. We find that as search cost decreases from 39 bp to 35 bp, mean marketing expenses drops from 61 bp to 44 bp. But when search cost further drops to 20 bp, the equilibrium marketing expenses become zero. Notice that the regulation cap is still at 100 bp. The intuition is as follows: low search cost renders marketing less profitable. In the model with high mean search cost, a subset of funds specifically exploit the high search cost investors. Those funds invest aggressively in marketing so as to enter more of the high search cost investors' choice set. Since high search cost investors will not search much, they will invest with those funds. But when mean search cost drops to sufficiently low level, this strategy is no longer profitable anymore.

Our paper is related to several strands of literature focusing on the mutual fund industry, and industrial organization more generally. Our structural model builds on Hortaçsu and Syverson (2004). To be able to study actively managed funds, as opposed to index funds, which are the focus of their model, we extend it in the following ways. First, we allow for funds to have stock-picking ability that exhibits decreasing returns to scale, following Berk and Green (2004). Second, we allow funds to choose both price and marketing expenditure. Third, we incorporate investor's learning about funds' abilities over time, also in the spirit of Berk and Green (2004). The last extension is very important because the observed dispersion in fund size might be due to investors' expectations of funds' skill. To estimate investor's (rational) expectations of funds' skills, we derive the MLE estimator based on the generalized Berk and Green (2004) model. To the best of our knowledge, we are the first to utilize Berk and Green's model to empirically estimate the investors' (rational) beliefs about the skills of active funds. We also use Berk and Green model's prediction as the benchmark for capital allocation across funds in a frictionless economy.

There is a growing literature examining the role of financial advisors. Hastings, Hortaçsu and Syverson (2016) study the impact of sales force on observed market outcomes in the Mexico privatized retirement savings systems. In their model, a fund's sales force can both increase investors' awareness of the product and impact their price sensitivity. In our

data we cannot distinguish between these two effects. We thus assume that the marketing expenses are purely informative (rather than persuasive). Christoffersen, Evans, and Musto (2013) find that the broker incentive impacts retail investors' investment decisions. Bergstresser, Chalmers and Tufano (2009) study broker-sold and direct-sold funds and find little tangible benefit of the former to fund investors. Egan, Matvos, and Seru (2016) show that there are potentially severe conflicts of interest between brokers/financial advisors and their retail investor clients, as exemplified by repeat incidence of misconduct in the industry (only about 5 percent of reported misconduct involves mutual funds, however).

Our paper is also related to the literature that aims to understand the observed underperformance of the active funds. Pástor and Stambaugh (2012) develop a tractable model of the active management industry. They explain the popularity of the active funds despite their poor past performances using two components: decreasing returns to scale and slow learning about the true skill level. In our model of the active management industry, we also include decreasing returns to scale and investor learning about unobserved skill (at the fund level). However, our model largely attributes the popularity of active funds to the information friction that prevents investors from easily finding out about index funds.[2]

This paper is related to those studying the role of advertising and media attention in the mutual fund industry. Gallaher, Kaniel and Starks (2006), Reuter and Zitzewitz (2006), and Kaniel and Parham (2016) study the impact of fund family-level advertising expenditures and the resulting media prominence of the funds on fund flows. In our model, we capture some of these effects parsimoniously by allowing fund family size to impact fund's probability of being included in investor's information set.[3]

The remainder of the paper is organized as follows. Section 2 develops our model. Section 3 describes the data used to estimate the model. Section 4 discusses the estimation

[2]Garleanu and Pedersen (2016) [17]incorporate search costs in their model of active management and market equilibrium, but assume that a passive index is freely available to all investors without the need to search.

[3]We follow this simple approach to incorporating advertising since the latter constitutes a very small fraction of fund expenditure, compared to the distribution costs that we focus on. Advertising can be potentially quite important for steering consumers into financial products - e.g., Honka, Hortaçsu and Vitorino (2016) and Gurun, Matvos and Seru (2016).

methods. Section 5 presents the estimation results. Section 6 conducts the counterfactual analysis. Section 7 concludes the paper.

2 Model

Our model combines elements from Berk and Green (2004) and Hortaçsu and Syverson (2004). Every period, investors conduct costly search to sample mutual funds to invest in. Investors care about the fund's expected performance and the expense ratio charged by the fund (i.e. its price). Mutual fund performance is determined by managerial skill as well as the impact of decreasing returns to scale. Mutual funds choose their expense ratios and the marketing expenses. The marketing expenditure can increase the fund's probability of being sampled but decreases its profit margin.

We proceed by first describing the investor's problem and then describe the funds' behavior.

2.1 Fund performance

In a time period t, the realized alpha $r_{j,t}$ for an active fund $j \in \{1, 2, ..., N\}$ is determined by three factors: (i) the fund manager's skill to generate expected returns in excess of those provided by a passive benchmark in that period, denoted by $a_{j,t}$. (ii) the impact of decreasing returns to scale, given by $D(M_t s_{j,t}; \eta)$ where M_t is the total size of the market and $s_{j,t}$ is the market share of the fund j, and $M_t s_{j,t}$ denoting fund size, η is the decreasing returns to scale parameter, and (iii) an idiosyncratic shock $\varepsilon_{j,t} \sim \mathcal{N}(0, \delta^2)$.

$$r_{j,t} = a_{j,t} - D(M_t s_{j,t}; \eta) + \varepsilon_{j,t}, \quad j = 1, ..., N, \tag{1}$$

There are papers discussing issues related to relative size between active funds and passive funds, (e.g., Pástor and Stambaugh 2012). To be able to address this important extensive margin, we include a single index fund $j = 0$ into our model. The alpha of the index fund is assumed to be zero. M_t includes both active funds and the index fund. We treat M_t as an exogenous variable in the model. Our specification is very similar to Berk and Green

(2004) with one exception: the manager's skill is allowed to vary over time. We assume manager's skill follows an AR(1) process:

$$a_{j,t} = (1-\rho)\mu + \rho a_{j,t-1} + \sqrt{1-\rho^2} \cdot v_{j,t}, \tag{2}$$

where $v_{j,t} \sim \mathcal{N}(0, \kappa^2)$. When a fund is born, its first period skill will be drawn from the stationary distribution $\mathcal{N}(\mu, \kappa^2)$. ρ captures the persistence of the skill level. In the limiting case, when $\rho = 1$, skill is fixed over time, which is what Berk and Green (2004) assume.

Following Berk and Green, we assume the manager's skill is not observable to either the investor or fund manager herself: it is treated as a hidden state. Let $\widetilde{a}_{j,t}$ be investor's belief about the manager's skill in that period. Since (2) can be regarded as describing how the hidden state $a_{j,t}$ evolves over time, and (1) says that $r_{j,t} + D(M_t s_{j,t}; \eta)$ is a signal on the hidden state, one can apply Kalman filter to obtain the following recursive formulas:

$$\begin{aligned}
\widetilde{a}_{j,t} &\equiv \boldsymbol{E}\left(a_{j,t} | r_{j,t-1}, s_{j,t-1}, r_{j,t-2}, s_{j,t-2}, \ldots\right) \\
&= \rho\left\{\widetilde{a}_{j,t-1} + \frac{\widetilde{\sigma}^2_{j,t-1}}{\widetilde{\sigma}^2_{j,t-1} + \delta^2}\left[r_{j,t-1} + D(M_{t-1}s_{j,t-1}; \eta) - \widetilde{a}_{j,t-1}\right]\right\} + (1-\rho)\mu, \qquad (3) \\
\widetilde{\sigma}^2_{j,t} &\equiv \boldsymbol{Var}\left(a_{j,t} | r_{j,t-1}, s_{j,t-1}, r_{j,t-2}, s_{j,t-2}, \ldots\right) \\
&= \rho^2\left(1 - \frac{\widetilde{\sigma}^2_{j,t-1}}{\widetilde{\sigma}^2_{j,t-1} + \delta^2}\right)\widetilde{\sigma}^2_{j,t-1} + (1-\rho^2)\kappa^2. \qquad (4)
\end{aligned}$$

and $\widetilde{r}_{j,t} = \mu$, $\widetilde{\sigma}^2_{j,t} = \kappa^2$ for the period t when j was born. When ρ is close to 1, these formulas reduce to what Berk and Green (2004) derived in their proposition 1. The difference between our updating rule and theirs is that in Berk and Green, all the historical signals receive the same weight in determining the investor's belief, whereas in our case, when ρ is smaller than 1 the signals in the more recent periods receive larger weights.

2.2 Investor search

In each time period t, each investor allocates a unit of capital to a single mutual fund as a result of sequential search (conducted during the period). For notational simplicity, the subscript t is suppressed in this subsection. Investor i pays search cost c_i to sample one fund from a distribution of funds. Let $\Psi(u)$ be the probability of sampling a fund with utility smaller or equal to u. (We will explain the investor's utility function in details in the next subsection.) Standard Bellman equation argument implies that it is optimal for investors to follow a cutoff strategy.[4] Let u^* be the currently searched highest utility. The investor continues searching iff $u^* \leq \bar{u}(c_i)$, where the threshold $\bar{u}$ is defined by

$$c_i = \int_{\bar{u}}^{+\infty} (u' - \bar{u}) d\Psi(u').$$

Since we have finite number of funds, the above expression becomes

$$c_i = \sum_{k=0}^{N} \psi_k (u_k - \bar{u}) \cdot \mathbf{1}\{u_k > \bar{u}\},$$

where $\psi_k \equiv \Psi(u_k) - \Psi(u_k^-)$ is the sampling probability of fund $k \in \{0, 1, ..., N\}$. The left hand side is the cost for an additional search, and the right hand side is expected gain. Note that the right hand side is strictly decreasing in $\bar{u}$. So $\bar{u}(c_i)$ is *strictly* decreasing in

[4] Fix an investor. For notational simplicity, we suppress the subscript i. Consider a cutoff strategy that stops at any $u > \bar{u}$. With such a strategy, the value function $V(u^*) = u^*$ for all $u^* > \bar{u}$. On the other hand, the value for $u^* \leq \bar{u}$ should be given by

$$V(u^*) = \sum_{t=0}^{+\infty} \Psi(\bar{u})^t \left[1 - \Psi(\bar{u})\right] \left[\frac{\int_{(\bar{u},\infty)} u d\Psi(u)}{1 - \Psi(\bar{u})} - (t+1)c \right] = \frac{1}{1 - \Psi(\bar{u})} \left[\int_{(\bar{u},\infty)} u d\Psi(u) - c \right].$$

In particular, $\Psi(\bar{u})^t \left[1 - \Psi(\bar{u})\right]$ is the probability that the investor stops exactly at period $t+1$; multiplying this probability is the expectation of the sampled u that triggers the stop minus the incurred search costs of $t+1$ periods.

Notice that $V(u^*)$ for $u^* \leq \bar{u}$ is a constant that does not depend on u^*. In addition, we must have $V(\bar{u}) = \bar{u}$, which gives us the expression for $\bar{u}$: $c = \int_{(\bar{u},\infty)} (u - \bar{u}) d\Psi(u)$. With $\bar{u}$ thus defined, the value function can be written as

$$V(u^*) = \max\{u^*, \bar{u}\}.$$

One can verify that this value function satisfies the Bellman equation:

$$V(u^*) = \max \left\{ u^*, \quad -c + \int_{-\infty}^{+\infty} V\left(\max\{u^*, u\}\right) d\Psi(u) \right\}.$$

c_i. This is intuitive: the bigger is c_i, the smaller is the cut-off $\bar{u}(c_i)$, and the less persistent the investor is in searching.

To facilitate later derivation, here we define a fund-specific cutoff f_j, $j = 0, 1, ..., N$, where

$$f_j = \sum_{k=0}^{N} \psi_k (u_k - u_j) \cdot \mathbf{1}\{u_k > u_j\}.$$

Notice that $u_j = \bar{u}(f_j)$. So, if $c_i > f_j$, then $u_j > \bar{u}(c_i)$. In other words, if an investor's search cost is larger f_j, he will stop searching once he finds fund j.

We assume that search cost c_i is drawn from a continuous distribution with c.d.f. $G(\cdot)$. In our empirical analysis we specify it to be exponential with mean λ. As in Hortaçsu and Syverson (2004), we endow investors with one free search, so that every investor will invest in a fund (regardless of their search cost). Let τ be a permutation that maps $\{0, 1, ..., N\}$ to the same $\{0, 1, ..., N\}$. Let τ be such that $u_{\tau(0)} \leq u_{\tau(1)} \leq ... \leq u_{\tau(N)}$. As a result, $f_{\tau(0)} \geq f_{\tau(1)} \geq ... \geq f_{\tau(N)}$.

Any investor who has a search cost that is higher than $f_{\tau(0)}$ will not make a second search beyond the free search. Then among all of these investors, with $\psi_{\tau(0)}$ probability, they will find fund 0, the worst fund. Nevertheless, they will invest in fund $\tau(0)$. No one else will invest with fund $\tau(0)$. So the market share for fund $\tau(0)$ is

$$s_{\tau(0)} = \psi_{\tau(0)} \left[1 - G(f_{\tau(0)})\right].$$

Two kinds of investors will buy fund $\tau(1)$. The first kind is the investors with $c_i > f_{\tau(0)}$ that find fund $\tau(1)$ in the free search. They have no choice but to invest. The second kind is investors with $f_{\tau(0)} \geq c_i > f_{\tau(1)}$. For these investors to invest in fund $\tau(1)$, they could have found it in the free search, or have found $\tau(0)$ in the free search and $\tau(1)$ in the second search, or have found $\tau(0)$ in the first two searches and $\tau(1)$ in the third search, and so forth... The probability for these events is $\psi_{\tau(1)} + \psi_{\tau(0)}\psi_{\tau(1)} + \psi_{\tau(0)}^2\psi_{\tau(1)} + ... = \frac{\psi_{\tau(1)}}{1-\psi_{\tau(0)}}$.

So the market share for fund $\tau(1)$ is

$$\begin{aligned} s_{\tau(1)} &= \psi_{\tau(1)}\left[1 - G(f_{\tau(0)})\right] + \frac{\psi_{\tau(1)}}{1-\psi_{\tau(0)}}[G(f_{\tau(0)}) - G(f_{\tau(1)})] \\ &= \psi_{\tau(1)}\left[1 + \frac{\psi_{\tau(0)}G(f_{\tau(0)})}{1-\psi_{\tau(0)}} - \frac{G(f_{\tau(1)})}{1-\psi_{\tau(0)}}\right]. \end{aligned}$$

We can follow this line of deduction to obtain the closed-form expressions for the market shares of all funds. For $j \geq 2$,

$$s_{\tau(j)} = \psi_{\tau(j)}\left[1 + \sum_{k=0}^{j-1} \frac{\psi_{\tau(k)}G(f_{\tau(k)})}{\left(1-\psi_{\tau(0)} - ... - \psi_{\tau(k-1)}\right)\left(1-\psi_{\tau(0)} - ... - \psi_{\tau(k)}\right)} - \frac{G(f_{\tau(j)})}{1-\psi_{\tau(0)} - ... - \psi_{\tau(j-1)}}\right].$$

2.3 Market share

In Section 2.2, we take the fund utilities and sampling probabilities as given. Now we specify the utility and sampling probabilities and derive the market shares that are consistent with this specification. Let $p_{j,t}$ be the expense ratio charged by the fund. An investor's utility for the fund is given by

$$u_{j,t} = \gamma \widetilde{r}_{j,t} - p_{j,t}, \tag{5}$$

where

$$\widetilde{r}_{j,t} = \widetilde{a}_{j,t} - \eta \log(M_t s_{j,t}).$$

Recall that $\widetilde{a}_{j,t}$ is the investors' belief on the manager's skill for fund j for this period t. The coefficient in front of the expense ratio is normalized to 1. For the decreasing returns to scale function $D(M_t s_{j,t}; \eta)$, we parameterize it as $\eta \log(M_t s_{j,t})$. For the index fund, $u_{0,t} = -p_{0,t}$ where $p_{0,t}$ is the expense ratio charged by the index fund in period t. The alpha of the index fund is defined to be zero. We assume that G is the exponential distribution with mean denoted by λ.

As to the sampling probabilities, we relate them with fund characteristics as well as

the marketing expenses. Let $\boldsymbol{x}_{j,t}$ be a vector of some characteristics of fund j , $b_{j,t}$ be the marketing expenses paid by fund j, and $\xi_{j,t}$ an unobserved fixed effect that affects the sampling probability. Vector $\boldsymbol{x}_{j,t}$ includes year dummies, fund age, and the number of funds in the same family.

$$\psi_{j,t} = \frac{e^{\boldsymbol{\beta}'\boldsymbol{x}_{j,t}+\theta b_{j,t}+\xi_{j,t}}}{1+\sum_{k=1}^{N} e^{\boldsymbol{\beta}'\boldsymbol{x}_{k,t}+\theta b_{k,t}+\xi_{k,t}}}, \tag{6}$$

$$\psi_{0,t} = 1 - \sum_{k=1}^{N} \psi_{k,t}. \tag{7}$$

Through the specifications in (5) , (6), and (7), the search model in Section 2.2 implies a mapping from $\tilde{\boldsymbol{r}}_t$, $\boldsymbol{p}_t$, $\boldsymbol{b}_t$, $\boldsymbol{x}_t$, and $\boldsymbol{\xi}_t$ to a set of market shares $\boldsymbol{s}_t$. Let us write this mapping as

$$s_{j,t} = F_{j,t}\left(\boldsymbol{p}_t, \boldsymbol{b}_t, \tilde{\boldsymbol{r}}_t, \boldsymbol{x}_t, \boldsymbol{\xi}_t; \Theta\right), \quad j = 0, 1, ..., N, \tag{8}$$

where Θ collects the relevant parameters, which in this case include λ, γ, $\boldsymbol{\beta}$, and θ.

However, $\tilde{\boldsymbol{r}}_t$ depends on the market shares due to decreasing returns to scale, so

$$\boldsymbol{s}_t = \boldsymbol{F}_t\left[\boldsymbol{p}_t, \boldsymbol{b}_t, \tilde{\boldsymbol{a}}_t - \eta \log\left(M_t \boldsymbol{s}_t\right), \boldsymbol{x}_t, \boldsymbol{\xi}_t; \Theta\right]. \tag{9}$$

The above equation pins down the equilibrium size of funds.

$\boldsymbol{s}$ is a fixed point. Assuming that this fixed point is unique, we can write it as a function of the other inputs on the right hand side of (9)

$$s_{j,t} = H_{j,t}(\boldsymbol{p}_t, \boldsymbol{b}_t, \tilde{\boldsymbol{a}}_t, \boldsymbol{x}_t, \boldsymbol{\xi}_t; \Theta), \tag{10}$$

with Θ now also including parameter η. Unlike $F_{j,t}$, $H_{j,t}$ has no obvious closed form expression and requires fixed-point iteration to compute.

2.4 Fund behavior

Recall that M_t is the market size at period t. The profit for fund j is given by

$$\pi_{j,t} := M_t \cdot H_{j,t}(\boldsymbol{p}_t, \boldsymbol{b}_t, \tilde{\boldsymbol{a}}_t, \boldsymbol{x}_t, \boldsymbol{\xi}_t; \Theta) \cdot (p_{j,t} - b_{j,t}). \tag{11}$$

If we assume a Nash equilibrium, each fund chooses $p_{j,t}$ and $b_{j,t}$ to maximize $\pi_{j,t}$, given $p_{-j,t}$ and $b_{-j,t}$. However, while the model should be able to match the patterns in the data, generally speaking, to exactly align the behaviors predicted by a model with the observed behaviors in the data, one must either introduce some unobserved heterogeneous costs or allow some level of bounded rationality.[5] In our model, we allow decision errors as each fund chooses its price and marketing expense. Specifically, the first order condition for the price for fund j at period t is

$$s_{j,t} + \left(\frac{\partial H_{j,t}}{\partial p_{j,t}} \cdot e^{\zeta_{j,t}} \right) (p_{j,t} - b_{j,t}) = 0, \tag{12}$$

where $\zeta_{j,t}$ represents the fund's possible mis-assessment of the slope of the demand curve. We will assume that $\zeta_{j,t}$ has a mean of zero across all periods and funds. In other words, while discrepancies are allowed at the individual fund level, we still ask the average behavior to be consistent with the model.

The first order condition for the marketing expenses is similar, except that it is possible for the fund to choose a corner solution:

$$-s_{j,t} + \left(\frac{\partial H_{j,t}}{\partial b_{j,t}} \cdot e^{\omega_{j,t}} \right) (p_{j,t} - b_{j,t}) \begin{cases} \leq 0, & \text{if } b_{j,t} = 0; \\ \geq 0, & \text{if } b_{j,t} = 0.01; \\ = 0, & \text{otherwise.} \end{cases} \tag{13}$$

The corner solution happens whenever the marketing expense is at either zero or the one percent upper bound set by regulations. Here we again allow a mean zero error $\omega_{j,t}$, which

[5] See Baye and Morgan (2004), which shows that allowing only a small amount of bounded rationality in players' optimization behaviors can be of great use in reconciling the Nash hypothesis with the commonly observed price patterns in the data.

represents the fund's possible mis-assessment of how its demand curve responds to the marketing expense.

There is a connection between the decision errors that we introduce here with the notion of ϵ-equilibrium in game theory, first introduced by Radner (1980). A set of choices constitutes an ϵ-equilibrium if the difference between what a player achieves and what he could optimally achieve is less than ϵ. In other words, it only requires each player to behave near-optimally, which turns out to be the same as what we ask in (12) and (13). Specifically, there is a mapping from $\zeta_{j,t}$ and $\omega_{j,t}$ to the loss that firm j incurs relative to its optimal payoff. When both errors are zero, such loss is zero. More importantly, it can be shown that this mapping is insensitive, in the sense that fairly large errors only lead to a relatively small loss of the optimal profit. In other words, even with some large mis-assessments of the slope of the demand curve, we introduce only a small amount of bounded rationality in terms of the loss on profits.

3 Data

The data come from CRSP and Morningstar. Our sample contains 2,285 well-diversified actively managed domestic equity mutual funds from the United States between 1964 and 2015. Our dataset has 27,621 fund/year observations. In the data appendix, we provide the details about how we construct our sample. We closely follow Berk and Binsbergen (2015) and Pastor, Stambaugh and Taylor (2015)'s data-cleaning procedures.

We now define some of the key variables used in our analysis. For the full list of variable definitions, please refer to Table 1. Summary statistics are provided in Table 2.[6]

[Table 1 about here.]

[Table 2 about here.]

To compute the annual realized alpha $r_{j,t}$, we start with monthly level return data. We first add fund's monthly net return with the fund's monthly expense ratio to get the

[6]We cross check our summary statistics with Pastor, Stambaugh and Taylor (2015)'s summary statistics. We find that our fund size variable's distribution is very similar. Our expense ratio is higher than theirs because we incorporate front load into the expense ratio.

monthly gross return $r^G_{j,t}$. Then we regress the excess gross return (of the 1-month U.S. T-bill rate) on the risk factors over the life of the fund to get the beta for each fund. We multiply beta with factor returns to get the benchmark returns for each fund at each point in time. We subtract the benchmark return from the excess gross return to get the monthly gross alpha. Last, we aggregate the monthly gross alpha to the annual realized alpha $r_{j,t}$. We use 4 different benchmarks: CAPM, Fama-French three-factors model, Fama-French and Carhart four-factor model and Fama-French five-factor model. For our main results, we use the Fama-French five-factor model as the benchmark. But our results are robust to other risk adjustments. In our sample, the average annual realized alpha for Fama-French five-factor model is 54 bp. This result is very close to Pastor, Stambaugh and Taylor (2015)'s estimates, where they find the monthly alpha is 5 bp, which translates to 60 bp of annual alpha.

In the model section, we define the index fund as the "outside" good. Here, we choose all the domestic well-diversified equity *index* funds from Vanguard as the proxy for the "outside" good. We choose Vanguard because, as proposed in Berk and van Binsbergen (2015), index funds from Vanguard are the most accessible index funds to the average investor. Alternatively, we could use all of the index funds offered in the market as the outside good. But due to the fact that Vanguard controls the majority market share in the index fund sector, we believe our results would not change substantially. Since in this paper, our focus is the efficient asset allocation across active funds, we choose to minimize the details related to modeling index fund.[7] We aggregate all index funds from Vanguard to build a single index fund. Specifically, we compute the assets under management (hereafter AUM) by summing AUM across all funds. And we compute the combined fund's expense ratio by asset-weighting across index funds. We count the combined index fund's age from the inception year of Vanguard which is 1975. We fixed this combined index fund's realized alpha at 0. In figure 1 and figure 2, we plot the total asset under management of all the Vanguard equity index funds and the asset-weighted mean expense ratio respectively. We can see that starting from the mid of the 1990s, Vanguard equity

[7]For a detailed study of search frictions *within* the index fund market, see Hortaçsu and Syverson (2004).

index funds start to take off in terms of AUM. Now it manages over 600 billion dollars. Meanwhile, the mean expense ratio keep decreasing from over 60 bp in 1975 to under 10 bp in 2015.

[Figure 1 about here.]

[Figure 2 about here.]

We define the total market M_t as the sum of AUMs of all the active funds and the combined index fund in year t. We define market share $s_{j,t}$ as the ratio between fund j's AUM and the total market. $M_t s_{j,t}$ gives the fund's AUM in millions of dollars. We exclude fund/year observations with fund's AUM below \$15 million in 2015 dollars. A \$15 million minimum is also used by Elton, Gruber, and Blake (2001), Chen et al (2004), Yan (2008), and Pastor, Stambaugh and Taylor (2015). In our dataset, there is huge skewness in fund's AUM. From the summary statistics, we can see the mean of fund's AUM is much larger than the median. The funds at the 99 percentile is over 1,100 times larger than the funds at the 1 percentile. This skewness could potentially affect our estimates. Following Chen at al (2004), for the usual reasons related to scaling, we use the log of a fund's AUM as the proxy of fund size.

Following the literature, we conduct our analysis at the fund level instead of the share class level. To be able to do so, we need to aggregate the share class level expense ratio, 12b-1 fee and front load to the fund level. In mutual fund industry, a single mutual fund may provide more than one share class to investors. Different share classes charge different expense ratios, front loads and 12b-1 fees. Since marketing expense and price are key variables to our study, here we elaborate how we construct the effective marketing expenses and expense ratios at the fund level. We define the marketing expense $b_{j,t}$ as "effective" 12b-1 fee. For fund j in year t , if a C share class exists, we replace all the other share classes expense ratios and 12b-1 fees as the C share class's data. If no C share class exists in the fund, then for all the other share classes, we take the sum of the share class's 12b-1 fee and the annualized front load for that share class and use it as the effective 12b-1 fee. [8] For this case, we also increase the expense ratio by the amount of

[8]Following Sirri and Tufano (1998), we annualize the front load by 7 years.

the annualized front load. Lastly, within a fund, across share classes, we aggregate the effective 12b-1 fee by the AUM of each share class to get the fund level effective 12b-1 fee. We do so for the expense ratio too. In figure 3, we plot the histogram of the effective 12b-1 fee. We can see that there are about 45.7% of the observations are binding at the upper bound 1% which is the cap imposed by SEC. And about 23.7% of the observations are at 0.

In figure 4 we plot the ratio between total marketing expenses and total expense ratios for active funds. We can see that this ratio is relatively stable from 1992 to 2015, at around 41%. Figure 5plots the aggregate time series of the total amount of marketing expenses in dollars. The mean is around 8.5 billion dollars. Marketing expenses is significant both in the absolute term and as a ratio to the industry's revenue.

In our model, marketing works in the same way as advertising. But we didn't use actual advertising data, instead we use the compensation to brokers. The reason is as follows: in U.S., many investors purchase mutual funds through intermediaries such as brokers or financial advisors. Among all the expenses that mutual fund companies categorized as marketing, advertising expenses constitute only a tiny portion (ICI report). The majority is compensation paid to brokers and financial advisors. We, therefore focus on this channel as the dominant method for marketing mutual funds.

[Figure 3 about here.]

[Figure 4 about here.]

[Figure 5 about here.]

The variables defined above are the main variables used in our estimation. The remaining variables from Table 1 will be defined later.

4 Estimation

We first estimate Berk and Green model related parameters μ, κ, δ, ρ, and η using the observed panel of $\{r_{j,t}, s_{j,t}, j = 1, ..., N, t = 1, ..., T\}$. Then we estimate the search model

parameters λ, γ, $\boldsymbol{\beta}$, and θ by relating the observed $s_{j,t}$'s to the fund characteristics, as well as making inferences from the equilibrium behavior of the funds.

4.1 Fund performance

From (1), we can write down the probability of observing $r_{j,t}$ conditional on the information in the data up to t:

$$\Pr\left(r_{j,t}\middle|s_{j,t}, r_{j,t-1}, s_{j,t-1}r_{j,t-2}, s_{j,t-2}, ...\right) \sim \mathcal{N}\left[\tilde{a}_{j,t} - \eta\log(M_t s_{j,t}),\ \tilde{\sigma}_{j,t}^2 + \delta^2\right].$$

In particular, notice that $s_{j,t}$ is a function of $\tilde{a}_{j,t}$ but not $a_{j,t}$, so it does not provide further information towards $a_{j,t}$ beyond $\{r_{j,t-1}, s_{j,t-1}r_{j,t-2}, s_{j,t-2}, ...\}$. Neither does $s_{j,t}$ depend on $\varepsilon_{j,t}$, so it does not provide any information towards $\varepsilon_{j,t}$.

We can use the above conditional probability to construct a partial log likelihood function:

$$\sum_{j=1}^{N}\sum_{t}\log\Pr\left(r_{j,t}\middle|s_{j,t}, r_{j,t-1}, s_{j,t-1}r_{j,t-2}, s_{j,t-2}, ...\right).$$

The first summation is across all the funds. The second summation is across all the periods in which fund j existed. One maximizes this likelihood with respect to μ, κ, δ, ρ, and η to obtain their estimates. The asymptotics rely on $N \to +\infty$.

4.2 Search model

Let $s_{j,t}$ be the observed share for fund j in period t. Given a set of parameters, we can find the $\boldsymbol{\xi}_t$ that matches our model predicted shares with the observed shares:

$$H_{j,t}(\boldsymbol{p}_t, \boldsymbol{b}_t, \tilde{\boldsymbol{a}}_t, \boldsymbol{x}_t, \boldsymbol{\xi}_t; \Theta) = s_{j,t}, \tag{14}$$

where $\tilde{\boldsymbol{a}}_t$ is obtained from the estimation on the fund performance. Solving for $\boldsymbol{\xi}_t$ can be done in a similar fashion as the contraction in Berry et al. (1995). However, because H requires fixed-point iteration to evaluate, this usually takes a lot of time. A shortcut is

instead solving $\boldsymbol{\xi}_t$ from

$$F_{j,t}\left[\boldsymbol{p}_t, \boldsymbol{b}_t, \tilde{\boldsymbol{a}}_t - \eta \log(M_t \boldsymbol{s}_t), \boldsymbol{x}_t, \boldsymbol{\xi}_t; \Theta\right] = s_{j,t}, \tag{15}$$

where η is obtained from the estimation on the fund performance. One simply plugs the observed $\boldsymbol{s}_t$ to the left hand side and then search for the $\boldsymbol{\xi}_t$ that makes $F_{j,t}$ equal to the observed $s_{j,t}$ for each j. Because evaluating $F_{j,t}$ is much faster than evaluating $H_{j,t}$, solving (15) is generally faster than (14).

The first set of moment conditions comes from $\boldsymbol{E}(\xi_{j,t}|\boldsymbol{x}_t, \tilde{a}_{t,j}) = 0$, which is standard in the framework of estimating differentiated product markets. Let $j \in t$ denote that fund j is alive in period t. The sample version of the moment conditions is

$$\sum_{t=1}^{T} \sum_{j \in t} \xi_{j,t} \begin{pmatrix} \boldsymbol{x}_{j,t} \\ \tilde{a}_{t,j} \end{pmatrix} = \mathbf{0}.$$

$\boldsymbol{x}_{j,t}$ contains the variables that affects fund's sampling probability besides marketing expenses. Following Hortaçsu and Syverson (2004), Chen et al (2004), we include both log age and number of funds in the same family to capture the fund level social learning effects and fund family level impacts. It seems reasonable to contain lag fund size information into $\boldsymbol{x}_{j,t}$. But we choose not to do it for the following reasons: 1 our model is static. If we were to include lag fund size into $\boldsymbol{x}_{j,t}$, that could potentially create dynamic incentives for the funds. Then the model and estimation are not perfectly consistent with each other. 2 In the data, fund size is quite persistent, including lag fund size could create "over-fitting" problem. The "over-fitting" problem arises due to the fact that lag size captures information of alpha, price, and marketing, all of which are persistent. So including the lag fund size contaminates the estimates of the coefficients for alpha, price, and marketing. For the above moment conditions to be valid, we need $\boldsymbol{x}_{j,t}$ and $\tilde{a}_{t,j}$ to be uncorrelated with $\xi_{j,t}$. Clearly the number of funds in the same fund family is decided at the fund family level. It should not be correlated with fund level unobservaed shock $\xi_{j,t}$. Second, we take the fund's age as exogenous as assumed in Hortaçsu and Syverson

(2004), so fund age is also uncorrelated with $\xi_{j,t}$. Finally, the fund expected skill $\tilde{a}_{t,j}$ is formed at the beginning of the period, before $\xi_{j,t}$ realized. So it is also uncorrelated with $\xi_{j,t}$.

Notice that we do *not* require $\boldsymbol{E}(\xi_{j,t}|p_{t,j}) = 0$ because $p_{j,t}$ is an endogenous outcome of the model so it is correlated with $\xi_{j,t}$ in general. The same applies to $b_{j,t}$, which is also endogenous in our model. One typical approach that the literature explores to deal with such endogeneity is using instruments that could influence the firm's pricing or marketing choices. Another typical approach is relying on the optimality of the observed firm choices; intuitively, what choices are optimal depend on the properties of the demand function, so reversely, the observed choices, if are optimal, must tell us something about the demand function.

Here we take the second approach, mostly because we believe it is difficult to come up with really good instruments in the context of the mutual fund industry. Nevertheless, later, we will explore a few instrument candidates to compare with our benchmark estimates. For the benchmark estimation, we assume nothing more than that the fund choices of prices and marketing expenses are optimal, *on average*. Given (12) and (13), this amounts to

$$\sum\nolimits_{t=1}^{T} \sum\nolimits_{j \in t} \zeta_{j,t} = 0, \tag{16}$$

$$\sum\nolimits_{t=1}^{T} \sum\nolimits_{j \in t} \omega_{j,t} = 0. \tag{17}$$

The first error, $\zeta_{j,t}$, can be directly backed out from the first order condition given any set of parameter values:

$$\zeta_{j,t} = -\log\left(\frac{-\partial H_{j,t}/\partial p_{j,t}}{s_{j,t}}\right) - \log\left(p_{j,t} - b_{j,t}\right).$$

The other error, $\omega_{j,t}$, can be computed exactly for $0 < b_{j,t} < 0.01$ but unfortunately not

for the boundary cases:

$$\omega_{j,t} \begin{cases} \leq \bar{\omega}_{j,t}, & \text{if } b_{j,t} = 0; \\ \geq \bar{\omega}_{j,t}, & \text{if } b_{j,t} = 0.01; \\ = \bar{\omega}_{j,t}, & \text{otherwise,} \end{cases}$$

where

$$\bar{\omega}_{j,t} \equiv -\log\left(\frac{\partial H_{j,t}/\partial b_{j,t}}{s_{j,t}}\right) - \log\left(p_{j,t} - b_{j,t}\right).$$

Hence, we cannot simply use the average of $\omega_{j,t}$ as an estimate of $E(\omega_{j,t})$. A conventional way to deal with this kind of truncation problem is making an additional distributional assumption and applying an MLE estimator. However, a key issue here is that the truncation varies endogenously across funds. To see this, notice that another way to write the truncated value is:

$$\bar{\omega}_{j,t} + \log\left(\partial H_{j,t}/\partial b_{j,t}\right) = \zeta_{j,t} + \log\left(-\partial H_{j,t}/\partial p_{j,t}\right).$$

As explained in Section 2.4, it is reasonable to expect a high and positive correlation between $\zeta_{j,t}$ and $\omega_{j,t}$. As a result, $\bar{\omega}_{j,t}$ is likely correlated with $\omega_{j,t}$.

There might be methods to estimate $E(\omega_{j,t})$ even with such endogenously truncated values. However, here we take a less technical but more exploratory approach by comparing the estimates based on several subsample versions of (17):

$$\text{(i)} \sum_{0<b_{j,t}<0.01} \omega_{j,t} = 0; \ \text{(ii)} \sum_{b_{j,t}=0} \bar{\omega}_{j,t} = 0; \ \text{(iii)} \sum_{b_{j,t}=0.01} \bar{\omega}_{j,t} = 0; \ \text{(iv)} \sum_{all} \omega_{j,t} = 0.$$

The first version (i) assumes that on average, the funds that choose an interior marketing are right about the effect of marketing on market share. These are the funds for which we can exactly calculate the ω_j. We acknowledge that these funds are a selected sub-sample of all funds; their average do not necessarily reflect the average of all funds. However, these are the funds that choose the less extreme marketing expenses. In addition, they make up a substantial portion (about 30 percent) of the funds in the data, so it is reasonable to believe that their average assessment is not far from the population average. The second

version (ii) uses the truncated values (upper bounds) of the ω_j of the funds that choose zero broker marketing expenses. The third version (iii) uses the truncated values (lower bounds) of the ω_j of the funds that choose the highest possible marketing expenses. The last version (iv) uses all the values for ω_j.We use these three latter cases as robust checks. If the estimates based on these four different assumptions are similar, then we can be confident that estimates based on the full sample moment (17) will be similar too.

4.3 Standard errors

The standard errors can be computed by parametric bootstrapping. The only element that we have to take as exogenous in the simulation is the existence of the funds over time (we do not have a model of entry and exit). The shocks that we need to generate include $\nu_{j,t}$, $\varepsilon_{j,t}$, $\xi_{j,t}$, $\zeta_{j,t}$, and $\omega_{j,t}$. The latter two shocks are highly correlated (as explained in Section 2.4,) and each shows persistence over time. One way to incorporate these is using a VAR process. We can start at year $t = 1$, first take the $\tilde{a}_{j,1}$ as the prior beliefs, then compute the equilibrium prices, marketing expenses, and market shares, given the prior beliefs and a set of randomly drawn $\xi_{j,1}$'s. After this, we can move on to $t = 2$, first compute the belief $\tilde{a}_{j,2}$ based on the simulated $r_{j,1}$ and $q_{j,1}$, then compute the equilibrium given these beliefs and a set of $\xi_{j,2}$'s. Keep moving on til the last period T. This provides us with a panel of simulated data on which we can apply our estimation algorithm. We can run Monte Carlo experiments to verify that our estimator is able to recover "true" parameters.

5 Results

5.1 Fund Performance Estimation Results

Table 3 reports estimates of the fund performance related parameters using our full sample.

[Table 3 about here.]

We want to emphasize that it is important to use all the available information to estimate the Berk and Green model. In Berk and Green model, when a fund was born, it draws an initial skill level from the prior skill distribution. Then investors use *all* the subsequent realized performances to update their belief about the fund's skill level. So it is crucial to include all the realized performances from the beginning of the fund. In our dataset, the first period with non-missing data is the year 1964, so our full sample estimates use the data from 1964 to 2015. If we were to start the sample from some time later, for example, year 1995, then we lose the performance information for a lot of funds who are in operation well before year 1995. And those information are important in terms of pinning down the model's parameter. [9]For demonstration, we also estimate the model using different starting points: 1975, 1985 .[10] We can see that using shorter samples under estimate both decreasing returns to scale and the mean prior skill level of funds. Now let's discuss the estimates.

The magnitude of decreasing returns to scale,η is 0.0048. And it is statistically significant. Since one standard deviation of log fund size is 1.628, a one standard deviation **positive** shock to the fund size is associated with approximately 78 basis point **decrease** in annual alpha. This result is close to Chen el al (2004). This magnitude is economically significant. In comparison, in our sample, the mean gross alpha is 54 basis point. For robustness check, we also estimate the model using linear fund size instead of log fund size and the estimated η is close to Pastor, Stambaugh and Taylor's results.

How skillful are mutual fund managers is an interesting question. Previous literature uses fund level performance persistence as a proof of the existence of skills. Here we take a different approach by estimating the Berk and Green model. We find that the mean of the prior managerial skill distribution is 3.05%. This number is positive and significant, which means that the mutual fund managers on average are skillful. Using the estimated parameters μ and κ, we plot the distribution of management skill in Figure 6. We also add a vertical line of the mean expense ratio. We can see that over 71% of the

[9]To circumvent the truncation problem, we can pick a starting year and keep only the funds which are founded after this year. But this approach would bias the estimates toward newer funds.

[10]Due to computation burden, we didn't provide the standard errors for the estimates starting from 1975 and 1985.

funds have the skill levels higher than the mean expense ratio. Compare our results with Berk and Green's calibrated management skill distribution, we have a lower mean skill level and a higher precision. The difference results from different optimization objectives. They calibrate these two parameters by targeting the empirical survival rates and relation between the flow of funds and performance. We estimate the two parameters in a MLE approach by minimizing the difference between predicted alpha and realized alpha. Our MLE likelihood function is derived from Berk and Green model's prediction. In their model, the investor's belief combined with the impact of decreasing returns to scale is the unbiased predictor for fund's realized performance.

[Figure 6 about here.]

Another parameter of interest is ρ which is the persistence of skill level. Our empirically estimated persistence is 0.94. In words, our results support the existence of skilled or informed mutual fund managers. The logic is as follows: if we were to believe that mutual fund managers have no skills, then all the realized performances are driven by idiosyncratic noises. Those noises have no persistence. As a result, the past beliefs which are formed based on past performances and sizes are not a useful predictor of future performance. This will drive down ρ. In our estimation, we find this is not the case. Indeed past beliefs are very useful in predicting future performances. Our skill persistence result is consistent with Berk and van Binsbergen (2015) where they find that the cross sectional difference in value added are persistent for as long as 10 years. Also, our estimated persistence parameter is very close to Berk and Green's model assumption,1. The reason why we don't get a persistence parameter exactly at 1 is because in reality there is managers turnover at the mutual funds. If we believe the management skill of a mutual fund is partially due to the mutual fund manager, then a change of the manager might affect the skill level of the fund. Fidelity Magellan fund manager Peter Lynch would be an apt illustration. During his tenure from 1977 to 1990, according to our measure of performance, Magellan fund achieved 14 consecutive years of positive alpha. After Peter Lynch's departure, Magellan's performance becomes not that impressive.

5.2 Asset Misallocation

Equipped with estimated parameters, we compute the investors' belief about each fund's skill level at each time. Then we can derive the Berk and Green's model predicted fund size. By comparing the difference between fund size in the data and in the model, we can assess the asset misallocation in the mutual fund industry. Now let's go step by step.

First, we compute the investor's belief about the fund's skill level $\tilde{a}$ using the recursive expression derived in 2.1. At fund's birth, we assign the fund a skill level of μ.Then we use realized return $r_{j,1}$ and fund size $M_1 s_{j,1}$ to get the updated belief, $\tilde{a}_{j,1}$. By iterating forward, we can generate the whole series of fund j's expected skill level. Next, we compute the Berk and Green model implied fund size. Berk and Green's model predicts that fund's size should be determined by the investor's belief and the price charged by the fund and the degree of decreasing returns to scale:

$$\log(q_{j,t}^{BG}) = \frac{\tilde{a}_{j,t} - p_{j,t}}{\eta}, \tag{18}$$

where q is fund size (as in dollars). We denote $\tilde{a}_{j,t} - p_{j,t}$ as net skill of fund j at period t. Equation 18 is intuitive: the higher the skill level $\tilde{a}_{j,t}$, the larger the fund's size. The higher the price and the larger the decreasing returns to scale effect, the smaller the fund's size.

To compare the model predicted fund sizes with data, we conduct the following exercise. We construct ten portfolios of mutual funds base on the deciles of net skill. We then compute the portfolio's mean of log size in the data and in the model.[11] Figure 7 presents the result. First, we can see that in the data, the mean fund size monotonically increases with net skill. This result support the Berk and Green's prediction. But we also witness discrepancy between the data and the model. On the higher end, BG predicts the mean size of funds in portfolio 10 to be 7.3 billion. In the data, the mean size of funds in portfolio 10 is 936 million. On the lower end, according to BG, the mean size of funds in portfolio 1 is 0.7 million. And in the data, it is 134 million. From this figure, we can

[11] We winsorize the belief $\tilde{a}$ at 1% and 99% level because there are some outliers in the belief. And we don't want our results to be driven by the tail.

draw the conclusion that asset misallocation exists in both bad funds and good funds in the data.

[Figure 7 about here.]

To quantitatively assess the amount of asset misallocation, we use the value-added measure proposed by Berk and van Binsbergen (2015). Value added is the product of realized performance and fund's size, $V_{j,t} = q_{j,t-1} r_{jt}$. For the data, we have both realized performance and fund's size, but for the Berk and Green model, we don't have the fund's realized performance in the model. In other words, we don't know in the counter factual world, what will be each fund's realized performance. To be able to make the comparison, we generate the model implied realized performance: it is the return in the data adjusted by the impact of decreasing returns to scale causing by the difference between sizes:

$$r_{j,t}^{BG} = r_{j,t} + \eta \log(q_{j,t}) - \eta \log(q_{j,t}^{BG}).$$

If BG predicted size is larger than the fund size in the data, then $r_{j,t}^{BG}$ is smaller than $r_{j,t}$ due to the decreasing returns to scale effect. Table 4 summarizes our findings. In Panel A, the total value added generated by the BG model is about 20 times larger than that in the data. After we adjust for the number of observations to get the mean value added, we find in the model it is 12 million dollars per fund per year. As in the data, it is only 0.6 million dollars per fund per year. The difference between the model implied value added and the value added in the data is our measure of inefficiency. In total, because of capital misallocation, we lose 311 billion dollars of value added. In table 4 Panel B, we also compute the value added difference between Berk and Green model and the data for 10 different portfolios. We find that according to Berk and Green model, most of the value added are generated by the top portfolio. This is natural because according to Berk and Green theory, the funds who have the largest net skill should be the largest funds. In the data, portfolio 1 to 8 all generated negative value added. This result is very similar to Berk and Binsbergen (2015) table 3. They find that across the whole distribution of average value added, 57.01% of funds generate negative value added. In their sample,

they include funds who invest both domestically and internationally whereas in our case, we focus on the domestic funds. The largest difference between BG value added and data value added are created by the difference between portfolio 10. The main message is that unskilled funds are not small enough while skilled funds are not large enough.

[Table 4 about here.]

5.3 Search Model Estimation Results

In the previous section, we demonstrate that there is asset misallocation in the U.S. mutual fund industry. To be able to explain the misallocation, we resort to informational friction. More specifically, we use search friction to explain it. Previous literature both in economics and finance demonstrates that costly search can distort capital allocation. In the mutual fund literature, Sirri and Tufano (1998) find that the performance-flow relationship is most pronounced among funds with higher marketing expenses. Hortaçsu and Syverson (2004) find even in index fund market, sizable search costs exist. We follow Hortaçsu and Syverson (2004) and model the impediments to investor ability to allocate capital optimally across mutual funds as a search friction. This approach is intuitive at least when applied to retail investors: the task of choosing among thousands of funds can be daunting even for the most sophisticated investors. Since mutual funds spend considerable resources on marketing, it is reasonable to assume that by doing so they are able to influence the likelihood of being picked by investors.

[Table 5 about here.]

Table 5 reports the estimates of parameters of the structural search model. Since later we want to conduct counterfactual analysis, we require our estimated parameters to be more relevant for the recent pattern. We pick the starting point as 2001. But our estimation results are robust to various starting points. As described in the estimation section, we estimate the model using three versions of moment conditions in 17. First thing to notice is that besides θ, all the other parameters are quite stable across the three

sets of estimates. This partially assures us that even though our identification for θ relies on some subsample, it will not affect other parameters drastically.

λ, the mean of search cost is 39 basis point. Hortaçsu and Syverson (2004) find that the mean search cost for the S&P 500 index fund market is from 11 bp to 20 bp across different specifications. [12] Our estimated search cost is higher than theirs. We attribute it to the following reason, the investors in their sample is a sub-sample of the investors in our sample. They specifically study the investors who decide to invest in S&P 500 index funds in the late 90s. As in our case, we study all the equity funds investors, both active fund investors and index fund investors. The index fund investors, in general know more about the mutual fund industry. And they can differentiate good funds from bad funds. They are the low search cost investors.

The magnitude of the mean search cost is quite significant. For the average investors, if they conduct another sampling, it cost them 39 basis points. λ is almost comparable to the mean alpha in our sample. The large magnitude of estimated search cost is a reflection of the active funds under performance puzzle. In the mutual fund literature, numerous papers documented the under performance of active funds. But in the data, those under performing funds still enjoy sizable market shares. For our model to rationalize those facts, it will imply high search cost. Then in our model, a lot of the high search costs investors will find it not optimal to conduct more searches to find better funds. In the counter factual case, if the search cost is low, then index funds should be much larger than what we observed in the data. Our search cost based explanation for the under performance puzzle could be empirically tested if we have the investor level survey data.

The next parameter of interest is θ, the coefficient in front of marketing expenses. First, we notice that the estimated θ is the smallest when we use the moment condition of the funds who choose 0, and the largest for the funds who choose the upper bound 1%. For the funds who choose the interior levels, θ is in the middle. This is intuitive because

[12] In Hortaçsu and Syverson, they estimated two types of search model. In the first type, the sampling probabilities across funds are different whereas in the second type, the sampling probabilities are the same. They estimate search cost for both types of model. We view our model closer to the first type. The estimation results for the first type of model is reported in Table III in their paper. The log mean search cost is around -6.17 to -6.78. So the mean search costs ranging from 11 bp to 21 bp in the S&P 500 index fund market.

θ measures the effectiveness of marketing. The funds who choose the upper bound must believe it is very useful in terms of increasing the fund's awareness.

To put the estimates into economic perspective, we conduct the following experiments. We compute the percentage changes in funds' size for various groups of funds if marketing expense increases by 1 bp. Table 6 provides the results. One thing to remember is that this experiment is a comparative statics. When we change fund j 's marketing, we fix all the other funds' prices and marketing expenses and fund j's price. So this is not a counter factual analysis. In panel A, we sort funds by their sizes. We find that as fund's size decreases, 1 bp increase in marketing leads to more increase in fund's size. This is intuitive because as a prior, marketing investment should be much more effective for smaller funds because they have smaller probabilities of being known. Investing in marketing is a good way for small funds to attract more of investor's attention. In panel B, we sort funds by their skill level $\tilde{a}$. Interestingly we find that marketing are much more useful for high skill funds. This indicates that if high skill funds can get into the consideration sets of more investors, they will be picked by more investors. But for the low skill funds, even if they are known to more investors, their size will not increase much. Lastly in panel C, we sort funds by their original marketing expenses levels. Binding at Lower Bound funds are funds who originally choose 0 marketing expenses. Binding at Upper Bound are funds who originally choose 1% marketing expenses. Non binding funds are the rest of funds. We can find that the additionally 1 bp increase in marketing is not very useful to funds binding at upper bound. This is due to the curvature of our sampling probability function. Basically there is decreasing returns to scale in marketing in our model.

Next we analyze the impact of marketing on funds' profit. One thing to remember is that this experiment is also a comparative statics. In panel A, we sort funds by their sizes. We find that for the small funds if all the other funds' strategies in pricing and marketing stay the same, their profit will increase. In panel B, we find when θ is at a high level of estimates, it is profitable for high skill funds to do more marketing. In panel C, we find, all the funds are worse off if they increase their marketing.

[Table 6 about here.]

[Table 7 about here.]

In the sampling probability function, besides fund's marketing expenses, we include family size, log fund age and year fixed effect. The coefficient of family size is positive and significant confirming the idea that larger fund families are better at informing investors about all of their products. Age coefficient is positive and significant. This result is consistent with Hortaçsu and Syverson (2004). There they study the S&P500 index fund market. Here we confirm that in the active funds market, the older funds also have more visibility than younger funds.

5.4 The Role of Sampling Probabilities

In this section, we quantify the impacts of various components in sampling probabilities in explaining the size distribution. Our method is as follows: we first set one of the components in sampling probability to be 0. Then we recompute the model implied market shares of funds. Notice here we are not recomputing the whole equilibrium. We fix all other variables and parameters. Lastly we regress the log of market share of funds in the data onto model predicted log market shares and report the R squared. In Table 8, we report the results. The lower the R squared, the more important that component is in terms of explaining the size distribution. Among all, ξ , the unobserved characteristics of the fund is the most important one. This is reasonable because we only include limited number of variables in our estimation. A lot of the variables that could potentially affect the size would be subsumed by ξ . The second most important variable is age. This result is consistent with Hortaçsu and Syverson (2004), where they also use fund age to approximate for the awareness of fund. After controlling for fund's age and other variables, the family size doesn't add much explanatory power. We next remove either marketing, price or skill and the R squared drop are approximately the same. This indicates that marketing is as important in terms of explaining size distribution as price or skill.

We are also interested in the question of how does various component contributing to the misallocation, so we compute the correlation between model predicted fund size and

Berk and Green model predicted fund size. We can see that in the data it is 0.0901. If by changing some component, this correlation increases, that means the allocation become more efficient. We see that by removing price and skill, this correlation drops. But if we remove marketing, this correlation actually increase. This means that marketing could potentially account for the misallocation. In the later counter factual section, we discuss how restricting marketing can improve welfare in more details.

[Table 8 about here.]

There is a different way to see the impacts of various components on capital misallocation. We redraw figure 7 but remove the components: ξ, fund age, fund family size and marketing expenses, respectively. The black line is the BG model implied size. The blue line is the data. The purple dash line plots the new fund size as predicted by the the restricted model. Now we compare the restricted model implied fund size with data. In the first figure, all the portfolios parallelly shift upwards. This is due to the Jensen's inequality introduced by the log fund size. In the second figure, all the portfolios parallelly shift downwards because fund's age is useful in informing investors. In the third figure we can see the counterfactual is similar to the data, this means that fund family size is not so important in affecting fund size. The last figure plots the counterfactual fund size when there is no marketing. Interestingly we see the purple line becomes steeper than the data. This means that marketing helps worse funds to preserve their market shares.

[Figure 8 about here.]

6 Counterfactual Analysis

Section 5.2 documents the asset misallocation fact in the mutual fund industry. In this section, we use our model to quantitatively study the importance of marketing expenses and search costs in shaping the equilibrium fund sizes and expense ratios. We also investigate how they affect allocation efficiency and investor's welfare. We present three counterfactual experiments. In the first one, we explore whether tightening regulation on

marketing could lead to worse allocation efficiency in active fund sector. In the second one, we shut down marketing and study the equilibrium market outcome. In the last one we investigate search cost's impact on equilibrium marketing expenses.

These experiments are closely related to the concern of Securities and Exchange Commission (SEC). SEC discussed several times about the impacts of marketing expenses on investor's welfare. In 2010, SEC discussed the proposal to improve the regulation of mutual fund distribution fees. Especially, the proposal includes the item on protecting investors by limiting fund sales charges which is one kind of marketing fees.

We proceed by first defining the welfare measures in our model and then conducting the counterfactual analysis. All the counterfactual analysis are using the data from year 2015. We can interpret all the results as if SEC impose a rule at the beginning of year 2015.

6.1 Welfare measures

6.1.1 Investor's welfare

In our model, investor's utility is consist of two parts, the expected indirect utility provided by the fund that he/she invested in and the expected total search costs he/she incurred before invested in this fund. The expression is as follows: for an investor with search cost c_i, his/her welfare is

$$V(c_i) = \frac{\int_{\bar{u}(c_i)}^{+\infty} u d\Psi(u)}{1 - \Psi\left[\bar{u}(c_i)\right]} - c_i \frac{\Psi\left[\bar{u}(c_i)\right]}{1 - \Psi\left[\bar{u}(c_i)\right]} \tag{19}$$

where $\bar{u}$ is the reservation level of indirect utility for investor i. The detailed derivation for investor's welfare is provided in the appendix. Here we provide the intuition. For a higher level of reservation utility, the investor needs to search more to get the desired fund. We can see that $\frac{\Psi[\bar{u}(c_i)]}{1-\Psi[\bar{u}(c_i)]}$is actually increasing in $\bar{u}$. For the first part, the numerator is the expected indirect utility for the funds with higher than $\bar{u}$ utility level. The denominator is to adjust for the fact that the investor will only pick the funds from this part of the distribution.

And the aggregate measure of utilities in this model is derived by integrating all the investors across the search costs which is given by:

$$U = \int_0^{+\infty} V(c_i) dG(c_i). \tag{20}$$

6.1.2 Fund's profit

If we were to think the mutual funds are owned by the investors as in a general equilibrium setting, fund's profit will also be part of the general welfare. The funds' profits include the profits for both active funds and index funds.

$$P = \sum_{j=1}^{N} (p_j - b_j) s_j + (1 - \sum_{j=1}^{N} s_j) p_0. \tag{21}$$

Here the first part is the total profit for the active funds, the second part is the total profit for the passive funds. In the counterfactual analysis, we assume index fund price is fixed and we resolve the equilibrium for the active funds' prices and marketing expenses. Here we omit the total market size M. In our counterfactual we assume M stay the same.

6.1.3 Marketing expenses

Actually marketing expenses is also a important piece in the welfare analysis:

$$B = \sum_{j=1}^{N} b_j s_j \tag{22}$$

6.1.4 Total welfare

Our measure of the total welfare is just the sum of the above three components.

$$T = U + P + B$$

6.2 Simulation 1: Regulation and active fund sector efficiency

In this section, we show that active fund sector's efficiency and total welfare could decrease as the regulation cap drops under certain parameterization. For the details of the parameters please check table 9. We set the regulation levels to 0bp, 25bp, 50bp, 75bp and current level 100bp. Funds reoptimize under new regulation levels. To assess the active fund sector's efficiency, we construct two measures: 1 correlation between the model-implied and the "efficient" (in the Berk and Green sense) allocation. 2 total value added implied by the model allocation. We provide the results in table 10. The trend is very clear: as relaxing regulation, Correlation between $q^{BG}$and q^{Model}, total value added and total welfare all increase. The intuition is simple: restricting marketing would steer investors from high-skill but "hard to find" funds to lower skill but "easy to find" funds. The correlation between BG implied fund size and model implied fund size captures this effect. A lower correlation indicates more serious misallocation problem. From table 6 we know that smaller funds and high skill funds will be affected more when marketing expenses changed. So for the smaller and high skill funds who were binding at 100 bp in the current regulation, if we tighten the regulation, it will decrease those funds' share a lot. But those funds under the BG model should be larger. This is the reason why we could have a loss in the efficiency when tightening the regulation. The loss of allocation efficiency also shows up in the measure of total value added created by all the funds. As for total welfare, under this parameterization, when tightening the regulation, it also drops. This indicates that the loss of allocation efficiency dominates other gains associated with less marketing expenses (e.g., cheaper funds).

This section is here to show that potentially active fund sector's efficiency and total welfare could decrease as the regulation cap drops. In the next section, we will conduct the counterfactual experiment at our estimated parameters and investigate the impact of different regulations on market outcomes and welfare.

[Table 9 about here.]

[Table 10 about here.]

6.3 Simulation 2

In this simulation, we restrict all funds to choose zero marketing expenses. We use year 2015's data and the parameters from column (4) in table 5. In table 11 we provide the comparison between current equilibrium and the zero marketing equilibrium on some of the key measures. First, the mean price drops by almost 77 basis points. This drop is larger than the mean marketing expenses. It indicates that when funds cannot pay marketing expenses, they not only lower their prices by the marketing expenses but even lower the prices more. We can interpret this result as fiercer price competition when funds cannot do marketing. To further understand the price changes across funds, we split the funds into four groups by their marketing in the current equilibrium: 1 the funds whose marketing is binding at 100 bp, 2 the funds whose is marketing binding at 0, 3 the funds whose marketing is between 1bp to 50 bp and 4 the funds whose marketing is between 50 bp and 99 bp. We first plot the price changes from current equilibrium to the no marketing equilibrium. We find that all the funds lower their prices in the no marketing equilibrium. The amount of change is quite different. Group 1 funds lower their prices around 100bp. The most interesting finding is that the group 2 funds in the new equilibrium also lower their prices by around 30 bp. This is mainly due to the competition effect across funds. And this is the main reason why the mean price drops more than the original amount of marketing expenses.

[Figure 9 about here.]

Second, we find that the total market share of active funds drops from 74% to 68%. This indicates that marketing was useful regarding increasing fund's awareness. When funds cannot do marketing, they lose market shares. But by losing some market share, the average alpha of the industry increases from 37 bp to 41 bp. This is due to the decreasing returns to scale effect. The sampling probability of index funds increases. This is due to the assumption that all the sampling probabilities sum to 1. When active funds cannot do marketing, the index funds are easier to be found. In no marketing equilibrium, active funds' profits drop for 15 basis point. This is resulting from both a shrink of total market

share and a shrink of profit margin. Investor's welfare increases by around 57%. There are three main contributing factors: 1 lower prices. 2 higher alphas, 3 lower search costs. We had already discussed the fund price changes. Here we discuss the alpha changes. In figure 10, we plot the alpha changes between no marketing equilibrium and current equilibrium. We find group 1 funds' alphas increase. This is mainly because in the no marketing equilibrium, those funds cannot do marketing, so their sizes shrink. Then due to decreasing returns to scale effect, their alphas increases. For other groups of funds, some of the alphas increase, some decrease. In total the alpha of the industry increases. This would also increase investor's welfare.

[Figure 10 about here.]

We can compute the total search cost incurred by the investors in the two equilibrium. The total search cost is defined as:

$$U = \int_0^{+\infty} c_i \frac{\Psi\left[\bar{u}(c_i)\right]}{1 - \Psi\left[\bar{u}(c_i)\right]} dG(c_i).$$

We find total search cost is lower in no marketing equilibrium. This is an interesting result since, as a prior, we would expect through marketing, investors get more information so that they would search less. But actually in the current equilibrium, people search more. How to understand it? In section 2.2, we describe how investors search, there is one important feature which is that investor search until they find the funds that satisfied their reservation utility level. Actually another way to interpret the search process is that investors search until the expected benefit of finding better funds is smaller than the unit search cost. If investor i with search cost c_i has already found fund j with utility u_j. Then his incentive to search hinges on the relationship between $\sum_{k=0}^{N} \psi_k (u_k - u_j) \cdot \mathbf{1}\{u_k > u_j\}$ and c_j. If there are not too many better funds out there, then investor's search incentive would be weaker. To show that this is indeed the reason why investors search less in the no marketing equilibrium, we plot the histogram of indirect utilities in the two equilibrium. We find the standard deviation in no marketing equilibrium is 0.0035 while in the current equilibrium it is 0.0055. So in the no marketing equilibrium, investors search less. Through

reducing search cost, investor's welfare increases by 17 basis point.

[Figure 11 about here.]

[Table 11 about here.]

By totally eliminating marketing, active funds total size doesn't drop drastically for two reasons: 1 in the sampling probability function, besides marketing expenses, there are also other variables. Those variables ensure that all active funds sampling probabilities are positive. The second reason is that by lower the size of active funds, it increases the performance of all the active funds. This effect makes active funds more attractive. So investors would find active funds more attractive.

Comparing the total welfare between no marketing equilibrium and current equilibrium, we find the no marketing equilibrium has a higher welfare. This result is mainly due to a significant increase of investor's welfare in the no marketing equilibrium.

6.3.1 Heterogeneous effect across investors

In our model, we assume different investors have different search costs. In this section, we study the impact of this new policy across different search cost groups. We simulate 100,000 investors according to our estimated search cost distribution. The simulated investors conduct costly search as described in section 2.2. Focusing on figure 13, we find that for all the search cost levels, in no marketing equilibrium, investors have higher welfare. But the biggest improvements come from the high search cost investors. Their welfare increases by roughly 100 basis point. For the low search cost investors the increase is not very large. This is because the low search cost investors always find the "best" funds in the market. Let's move to figure 14, there is an interesting non monotonic relationship between unit search cost and total search cost incurred. For the low search cost investors, since their unit search cost is low, even though they search a lot, their total search cost is not very high. For the high search cost investors, most of the time, they find it too costly to conduct any search, so they search infrequently. Consequently, high search cost investor's total search cost is also low. The intermediate search cost investors search

relatively aggressively and their search cost is non trivial. So in total they incur the largest total search cost. Comparing the two equilibrium, we find in the no marketing equilibrium, the intermediate search cost investors incur less total search cost. This is due to the fact that in the no marketing equilibrium, funds qualities improve. So that it will be easier for investors to find funds that satisfy their reservation levels. Focusing on figure 15, 16, and 17we find in general, high search cost investors get lower alpha funds, paid high prices and high marketing expenses. Those are easy to understand because high search cost investors don't search much.

An interesting fact is that for the very low search investors, the funds they invest in have positive net alphas. In Berk and Green's model, since investors have zero search cost so that in equilibrium, all the funds have zero net alphas. But in our model, since investors have positive search cost. The low search investors could find out some funds that are both skillful and cheap. This is an interesting extension to Berk and Green theory.

[Figure 12 about here.]

[Figure 13 about here.]

[Figure 14 about here.]

[Figure 15 about here.]

[Figure 16 about here.]

[Figure 17 about here.]

6.4 Simulation 3

Last, we examine the impact of search cost on equilibrium market outcomes with special attention to marketing expenses. Due to the existence of search cost, competing on marketing could be a potential profitable strategy. But with the emergence of Internet, advancement in search technologies (e.g., Google), and more transparent comparison (e.g.,

Morningstar), we would expect the search frictions to drop in the future. In other words, investors should find it easier to sample mutual funds with the help of new technologies. In this counterfactual, we set the mean search cost to 35bp and 20bp respectively. Given new search cost, funds reoptimize their prices and marketing expenses. We find that as search cost decreases from 39 bp to 35 bp, mean marketing expenses drops from 61 bp to 44 bp. But when search cost further drops to 20 bp, the equilibrium marketing expenses become zero. Notice that the regulation cap is still at 100 bp. The intuition is as follows: low search cost renders marketing less profitable. In the model with high mean search cost, there exists a large fraction of investors who have very high search costs. A subset of funds specifically exploit the high search cost investors. Those funds invest aggressively in marketing so as to enter more of the high search cost investors' choice set. Since high search cost investors will not search much, they will invest with those funds. But when mean search cost drops to sufficiently low level, this strategy is no longer profitable anymore. An interesting finding is that when search costs drop by certain moderate amount, all the funds choose not to invest in marketing. This reveals the fact that due to competition, when search cost are not very high, funds will not invest in marketing.

[Table 12 about here.]

7 Concluding Remarks

The question whether actively-managed mutual funds exhibit skill - i.e., persistent outperformance - has a long history in financial economics, since it is central to the debate about informational efficiency of securities markets in the sense of Fama. While there is still substantial debate about the ability of an "average" fund manager to generate abnormal returns (before or after fees are taken into account), perhaps one of the most robust findings in the literature is that investors' flows are much less sensitive to past bad performance than to outperformance (Sirri and Tufano 1998, Ippolito 1992, Chevalier and Ellison 1997, Carhart 1997, etc.). This evidence hints that the market for mutual funds

may not be efficient at allocating capital across funds because bad funds aren't punished sufficiently for poor performance, and therefore underperforming managers control more assets than justified by their level of skill. Capital misallocation in the mutual fund industry could potentially lead to inefficiencies in capital allocation across firms, distorting real investment (van Binsbergen and Opp 2016). It is therefore important to understand quantitatively how much capital is misallocated in the mutual fund industry. By estimating the Berk and Green model, we find that in the the U.S. equity mutual funds data, from year 1964 to year 2015, all but the best-performing decile of mutual funds are "too large" relative to the optimal scale predicted by the BG model. Overall, comparing model-implied and total value added following Berk and van Binsbergen (2015) implies that 5.8 billion dollars of value added is destroyed by capital misallocation, in an average year. These results indicate that there exist substantial frictions in the mutual fund market.

In our paper, we view mutual fund marketing expenses as purely informative (e.g., Butters 1977). It is possible that a portion of these marketing expenses serves a persuasive function in ways highlighted in the theoretical literature: e.g., firms may find it profitable to steer investors toward non-price attributes (Mullainathan et al 2008, Gabaix and Laibson 2006, Carlin 2009, Ellison and Ellison 2009). But to be able to separate the informative effect from the persuasive effect of marketing would require information about investors' actual choice sets, which is not available. Thus, by making the assumption that all marketing is informative, our welfare analysis results provides an upper bound on the social value of mutual fund marketing.[13] Relaxing this assumption in order to understand the possible welfare loss from "persuasive" marketing is a fruitful venue for future research.

[13]Even if marketing is purely informative, due to the externality of marketing in our model, marketing investment can still be excessive. Fund i's marketing investment could decrease fund j's probability of being known. In a Nash equilibrium, funds will not take the externality into consideration when deciding the marketing investment levels. All of the funds might be better off if they agree on a lower level of marketing investment - But of course, this agreement is fragile since deviation is profitable.

References

[1] Michael R Baye and John Morgan. Price dispersion in the lab and on the internet: Theory and evidence. *RAND Journal of Economics*, pages 449–466, 2004.

[2] Daniel Bergstresser, John MR Chalmers, and Peter Tufano. Assessing the costs and benefits of brokers in the mutual fund industry. *Review of financial studies*, 22(10):4129–4156, 2009.

[3] Jonathan B Berk and Richard C Green. Mutual fund flows and performance in rational markets. *Journal of political economy*, 112(6):1269–1295, 2004.

[4] Jonathan B Berk and Jules H Van Binsbergen. Measuring skill in the mutual fund industry. *Journal of Financial Economics*, 118(1):1–20, 2015.

[5] Steven Berry, James Levinsohn, and Ariel Pakes. Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, pages 841–890, 1995.

[6] Gerard R Butters. Equilibrium distributions of sales and advertising prices. *The Review of Economic Studies*, pages 465–491, 1977.

[7] Mark M Carhart. On persistence in mutual fund performance. *The Journal of finance*, 52(1):57–82, 1997.

[8] Bruce I Carlin. Strategic price complexity in retail financial markets. *Journal of financial Economics*, 91(3):278–287, 2009.

[9] Joseph Chen, Harrison Hong, Ming Huang, and Jeffrey D Kubik. Does fund size erode mutual fund performance? The role of liquidity and organization. *The American Economic Review*, 94(5):1276–1302, 2004.

[10] Judith Chevalier and Glenn Ellison. Risk taking by mutual funds as a response to incentives. *Journal of Political Economy*, 105(6):1167–1200, 1997.

[11] Susan EK Christoffersen, Richard Evans, and David K Musto. What do consumers' fund flows maximize? Evidence from their brokers' incentives. *The Journal of Finance*, 68(1):201–235, 2013.

[12] Mark Egan, Gregor Matvos, and Amit Seru. The market for financial adviser misconduct. Technical report, National Bureau of Economic Research, 2016.

[13] Glenn Ellison and Sara Fisher Ellison. Search, obfuscation, and price elasticities on the internet. *Econometrica*, 77(2):427–452, 2009.

[14] Edwin J Elton, Martin J Gruber, and Christopher R Blake. A first look at the accuracy of the crsp mutual fund database and a comparison of the crsp and morningstar mutual fund databases. *The Journal of Finance*, 56(6):2415–2430, 2001.

[15] Xavier Gabaix and David Laibson. Shrouded attributes, consumer myopia, and information suppression in competitive markets. *The Quarterly Journal of Economics*, 121(2):505–540, 2006.

[16] Steven Gallaher, Ron Kaniel, and Laura T Starks. Madison avenue meets wall street: Mutual fund families, competition and advertising. *Working paper*, 2006.

[17] Nicolae B Garleanu and Lasse H Pedersen. Efficiently inefficient markets for assets and asset management. Technical report, National Bureau of Economic Research, 2015. 2

[18] Umit G Gurun, Gregor Matvos, and Amit Seru. Advertising expensive mortgages. *The Journal of Finance*, 71(5):2371–2416, 2016.

[19] Justine Hastings, Ali Hortaçsu, and Chad Syverson. Advertising and competition in privatized social security: The case of mexico. *Econometrica*, 2016.

[20] Elisabeth Honka, Ali Hortaçsu, and Maria Ana Vitorino. Advertising, consumer awareness, and choice: Evidence from the us banking industry. *RAND Journal of Economics*, 2016.

[21] Ali Hortaçsu and Chad Syverson. Product differentiation, search costs, and competition in the mutual fund industry: A case study of s&p 500 index funds. *The Quarterly Journal of Economics*, 119(2):403–456, 2004.

[22] Richard A Ippolito. Consumer reaction to measures of poor quality: Evidence from the mutual fund industry. *The Journal of Law and Economics*, 35(1):45–70, 1992.

[23] Ron Kaniel and Robert Parham. Wsj category kings–the impact of media attention on consumer and mutual fund investment decisions. *Journal of Financial Economics*, 2016.

[24] Sendhil Mullainathan, Joshua Schwartzstein, and Andrei Shleifer. Coarse thinking and persuasion. *The Quarterly journal of economics*, 123(2):577–619, 2008.

[25] Luboš Pástor and Robert F Stambaugh. On the size of the active management industry. *Journal of Political Economy*, 120(4):740–781, 2012.

[26] L'uboš Pástor, Robert F Stambaugh, and Lucian A Taylor. Scale and skill in active management. *Journal of Financial Economics*, 116(1):23–45, 2015.

[27] Roy Radner. Collusive behavior in noncooperative epsilon-equilibria of oligopolies with long but finite lives. *Journal of economic theory*, 22(2):136–154, 1980.

[28] Jonathan Reuter and Eric Zitzewitz. Do ads influence editors? advertising and bias in the financial media. *The Quarterly Journal of Economics*, 121(1):197–227, 2006.

[29] Erik R Sirri and Peter Tufano. Costly search and mutual fund flows. *The journal of finance*, 53(5):1589–1622, 1998.

[30] Jules van Binsbergen and Christian Opp. Real anomalies: Are financial markets a sideshow. *Manuscript, University of Pennsylvania*, 2016.

[31] Xuemin Yan. Liquidity, investment style, and the relation between fund size and fund performance. *Journal of Financial and Quantitative Analysis*, pages 741–767, 2008.

Appendix

Investor's belief

Here we offer a few details on how to use the Kalman filter to derive investor's belief on manager's skill. Let $y_{j,t} \equiv r_{j,t} + D(q_{j,t}; \eta)$. By (1), we have

$$y_{j,t} = a_{j,t} + \varepsilon_{j,t}.$$

We can treat this as the measurement equation in a state space representation. The state equation is a simple AR(1) process for $a_{j,t}$ as specified in (2). Obtaining Equation (3) and (4) is simply a matter of applying the Kalman filter. In particular, $\tilde{a}_{j,t}$ is the one period ahead prediction of the state, and $\tilde{\sigma}_{j,t}$ is the variance of that prediction.

Proof for Cut-off startegy

Here we provide a few details on how to derive the optimal search strategy for the investors. Fix an investor. For notational simplicity, we suppress the subscript i. The Bellman equation for the dynamic problem is

$$V(u^*) = \max \left\{ u^*, \quad -c + \int_{-\infty}^{+\infty} V\left(\max\{u^*, u\}\right) d\Psi(u) \right\}.$$

Consider a cutoff strategy that stops at any $u > \bar{u}$. With such a strategy, $V(u^*) = u^*$

for all $u^* > \bar{u}$. On the other hand, the value for $u^* \leq \bar{u}$ should be given by

$$\begin{aligned}
V(u^*) &= \sum_{t=0}^{+\infty} \Psi(\bar{u})^t \left[1 - \Psi(\bar{u})\right] \left[\frac{\int_{(\bar{u},\infty)} u d\Psi(u)}{1 - \Psi(\bar{u})} - (t+1)c\right] \\
&= \sum_{t=0}^{+\infty} \Psi(\bar{u})^t \int_{(\bar{u},\infty)} u d\Psi(u) - c\left[1 - \Psi(\bar{u})\right] \sum_{t=0}^{+\infty} \Psi(\bar{u})^t (t+1) \\
&= \frac{1}{1 - \Psi(\bar{u})} \int_{(\bar{u},\infty)} u d\Psi(u) - c\left[1 - \Psi(\bar{u})\right] \left[1 + 2\Psi(\bar{u}) + 3\Psi(\bar{u})^2 + 4\Psi(\bar{u})^3 + ...\right] \\
&= \frac{1}{1 - \Psi(\bar{u})} \int_{(\bar{u},\infty)} u d\Psi(u) - c\left[1 - \Psi(\bar{u})\right] \{\left[1 + \Psi(\bar{u}) + \Psi(\bar{u})^2 + \Psi(\bar{u})^3 + ...\right] + \\
&\qquad\qquad \left[\Psi(\bar{u}) + \Psi(\bar{u})^2 + \Psi(\bar{u})^3 + ...\right] + ...\} \\
&= \frac{1}{1 - \Psi(\bar{u})} \int_{(\bar{u},\infty)} u d\Psi(u) - c\left[1 - \Psi(\bar{u})\right] \{\frac{1}{1 - \Psi(\bar{u})} + \frac{\Psi(\bar{u})}{1 - \Psi(\bar{u})} + ...\} \\
&= \frac{1}{1 - \Psi(\bar{u})} \left[\int_{(\bar{u},\infty)} u d\Psi(u) - c\right]. \qquad (23)
\end{aligned}$$

On the right side of the first line, $\Psi(\bar{u})^t \left[1 - \Psi(\bar{u})\right]$ is the probability that the investor does not stop for t periods and then stops. Multiplying this probability is the expectation of the sampled u that triggers the stop minus the incurred search costs of $t + 1$ periods.

Most importantly, notice that (23) is a constant that does not depend on u^*. In addition, we must have $V(\bar{u}) = \bar{u}$. Equating (23) with $\bar{u}$ gives us the expression for $\bar{u}$ that we gave in the main text:

$$c = \int_{(\bar{u},\infty)} (u - \bar{u}) d\Psi(u).$$

With $\bar{u}$ thus defined, the value function can be written as

$$V(u^*) = \max\{u^*, \bar{u}\}.$$

We can verify that it satisfies the Bellman equation, as for $u^* \leq \bar{u}$,

$$\begin{aligned}
-c + \int_{-\infty}^{+\infty} V\left(\max\{u^*, u\}\right) d\Psi(u) &= -c + \int_{-\infty}^{+\infty} \max\{u, \bar{u}\} d\Psi(u) \\
&= -c + \bar{u} + \int_{(\bar{u},\infty)} (u - \bar{u}) d\Psi(u) \\
&= \bar{u},
\end{aligned}$$

and for $u^* > \bar{u}$,

$$\begin{aligned} -c + \int_{-\infty}^{+\infty} V\left(\max\{u^*, u\}\right) d\Psi(u) &= -c + \int_{-\infty}^{+\infty} \max\{u, u^*\} d\Psi(u) \\ &= -c + u^* + \int_{(u^*,\infty)} (u - u^*) d\Psi(u) \\ &< u^*. \end{aligned}$$

Investor's welfare

Previous section provides the proof that the optimal search strategy is a cutoff strategy. In this section we compute the investor i's welfare for a given search cost c_i. First we denote $\bar{u}(c_i)$ as the reservation level of utility for the investor i. Investor i will only accept the funds which provide utilities higher or equal to $\bar{u}(c_i)$. The expected utility for the potentially accepted funds are $\frac{\int_{\bar{u}(c_i)}^{+\infty} u d\Psi(u)}{1-\Psi[\bar{u}(c_i)]}$. For the expected search costs, we have

$$\begin{aligned} c\left[1-\Psi(\bar{u})\right] \sum_{t=0}^{+\infty} \Psi(\bar{u})^t t &= c\left[1-\Psi(\bar{u})\right] \{\left[\Psi(\bar{u}) + \Psi(\bar{u})^2 + \Psi(\bar{u})^3 + ...\right] + \\ &= \left[\Psi(\bar{u})^2 + \Psi(\bar{u})^3 + ...\right] + ...\} \\ &= c\left[1-\Psi(\bar{u})\right] \{\frac{1}{1-\Psi(\bar{u})} + \frac{\Psi(\bar{u})}{1-\Psi(\bar{u})} + ...\} \\ &= c\frac{\Psi(\bar{u})}{1-\Psi(\bar{u})} \end{aligned}$$

where $\bar{u}$ is $\bar{u}(c_i)$. Combine the two parts together, we have the expression for investor's expected welfare.

The frictionless market

Here we derive the limiting case of our model when the search costs go to zero, $\lambda \to 0$. We will fix a period t and suppress the subscript t throughout the derivation.

First, notice that the active funds must provide the same utility, $u_j = u'$ for some u' for all $j \in \{1, ..., N\}$. To see this, suppose that some j has a utility that is strictly smaller than another fund. Because the investors do no incur search cost, no one will buy j. This

means $q_j \to 0$, which under the log specification of the decreasing return to scale, implies that $u_j \to +\infty$, a contradiction. By the same argument, one can show that $u' \geq -p_0$.

Let us first look at the case that $u' > -p_0$ for all $j \in \{1, ..., N\}$. The outside good will have zero market share. So

$$\sum_{j=1}^{N} q_j = 1.$$

In addition, from the utility specification (5), we have

$$q_j = e^{\frac{1}{\eta}\tilde{a}_j - \frac{1}{\eta\gamma}(p_j + u')}.$$

Putting the two above equations together, we can find the solution for u' and plug it back to the last equation to obtain:

$$q_j = \frac{e^{\frac{1}{\eta}\tilde{a}_j - \frac{1}{\eta\gamma}p_j}}{\sum_{k=1}^{N} e^{\frac{1}{\eta}\tilde{a}_k - \frac{1}{\eta\gamma}p_k}}. \tag{24}$$

Next let us look at the case where $u' = -p_0$. The size of an active fund will be at the point where the decreasing return to scale drives its utility to be the same as the index fund. This is basically the idea of Berk and Green (2004). From the utility specification (5), we have

$$q_j = e^{\frac{1}{\eta}\tilde{a}_j - \frac{1}{\eta\gamma}(p_j - p_0)}. \tag{25}$$

For this case, we must have $\sum_{j=1}^{N} q_j \leq 1$, which translates into

$$-p_0 \geq \eta\gamma \log\left(\sum_{k=1}^{N} e^{\frac{1}{\eta}\tilde{a}_k - \frac{1}{\eta\gamma}p_k}\right). \tag{26}$$

In other words, if this condition on the prices holds, then the market shares are given by (25), otherwise the market shares are given by (24).

We can derive the pricing behavior of funds given these market share equations. Each fund chooses p_j to maximize $q_j(p_j - b_j - mc_j)$. Suppose that condition (26) holds so that q_j is given by (25), then the first order condition implies a uniform markup of $\eta\gamma$ across

the active funds, or

$$p_j = \eta\gamma + b_j + mc_j.$$

If these prices satisfy condition (26), then we have a Nash-Bertrand equilibrium in which the index fund has positive market share.

Uniqueness of the equilibrium

It appears to us that we can prove the uniqueness of the fixed point, using a theorem in Kennan, John (2001), *Review of Economic Dynamics*, Vol.4, p.893-899. What need to be show here are (i) if $\boldsymbol{s}_t$ is a fixed point, then for $0 < z < 1$, we have

$$\boldsymbol{F}_t\left[\boldsymbol{p}_t, \boldsymbol{b}_t, \tilde{\boldsymbol{a}}_t - \eta \log\left(z M_t \boldsymbol{s}_t\right), \boldsymbol{x}_t, \boldsymbol{\xi}_t; \Theta\right] > z\boldsymbol{s}_t,$$

and (ii) for any $\boldsymbol{s}_t$ and $\boldsymbol{s}'_t$ where $s_{j,t} = s'_{j,t}$ but $s_{k,t} \geq s'_{k,t}$ for all $j \neq k$, we have

$$F_{j,t}\left[\boldsymbol{p}_t, \boldsymbol{b}_t, \tilde{\boldsymbol{a}}_t - \eta \log\left(M_t \boldsymbol{s}_t\right), \boldsymbol{x}_t, \boldsymbol{\xi}_t; \Theta\right] \geq F_{j,t}\left[\boldsymbol{p}_t, \boldsymbol{b}_t, \tilde{\boldsymbol{a}}_t - \eta \log\left(M_t \boldsymbol{s}'_t\right), \boldsymbol{x}_t, \boldsymbol{\xi}_t; \Theta\right].$$

Both conditions demand certain properties of the search model in Hortaçsu and Syverson (2004). Intuitively, condition (i) requires that when all active funds' utilities increase by a constant, each active fund's share in the search model does not decrease; condition (ii) requires that when the utilities of all but one active fund increase, that one fund's share in the search model decreases.

Data Appendix

In this appendix, we describe our dataset construction process. Our raw data come from CRSP Survivor-Bias-Free US mutual fund dataset and Morningstar.

CRSP dataset Clean-up

In this step, we follow Berk and Binsbergen's (Hereafter BB) procedure as close as possible.

crsp_fundno and ticker mapping

We merge CRSP and Morningstar datasets based on both tickers and CUSIPs. In CRSP, the unique identifier for each fund's share classes is crsp_fundno. Our *goal* is that after the cleaning procedures: there is a one to one mapping between crsp_fundno and ticker. And there is a one to one mapping between crsp_fundno and CUSIP.

We download the annual fund summary dataset from CRSP through Wharton Reserach Data Service (WRDS). The data span 1961 Jan to Dec 2015. There are 505,073 observations.

1. Out of 505,073 observations, there are 400 observations with same {crsp_fundno, year} as other observations. These duplications all due to multiple reports in the same year. Out of the 400 observations, there are 200 distinct crsp_fundno. We keep the observation with non missing expense ratio information and delete the other one. Now we have 504,808 observations. After this step, we don't have any observations with identify crsp_fundno and year.

2. Out of 504,808 observations left, we have 86,793 obs that missing ticker. We will follow BB's steps to fill those.

3. First we identify all the unique pairs of {crsp_fundno, ticker}. Here we first delete the observations with missing tickers. Then we got 53,278 unique pairs. We find that there are 5,425 pairs of which have the same crsp_fundno but more than one ticker. We follow BB's procedure: we keep the latest ticker which is the ticker with most recent year. Then we back fill all the crsp_fundno with that ticker. That give us 2,595 unique pairs between {crsp_fundno, ticker}. Then add back the non duplicated cases, we have 50,448 unique {crsp_fundno, ticker} pairs.

4. Up to this point, for each crsp_fundno, there is only one ticker. But for each ticker there could be multiple crsp_fundnos. Now we identify the tickers that have multiple crsp_fundnos. There are actually 4,343. We follow BB to leave them as missing. Now we get 42,436 unique pairs of {crsp_fundno, ticker}.

Feature: now our dataset have the one to one mapping between crsp_fundno and ticker.

crsp_fundno and CUSIP mapping

According to Pastor, Stambaugh and Taylor (hereafter PST), CUSIP can match a lot of Morningstar funds to CRSP funds. So we also clean the CUSIP in CRSP. In general we conduct the exact same procedures as we did with the ticker. So in the following, we only report some key statistics.

1. Out of 505,073 observations, there are 120,837 observations with missing CUSIPs. After we do the back fill, observation with missing CUSIP reduced to 29,436.

2. Next we identify the CUSIPs that has been used by multiple crsp_fundno. There are 494 such CUSIPs. We set them to missing.

3. Lastly we have 53,297 unique pairs of {crsp_fundno, cusip}.

Feature: now our dataset have the one to one mapping between crsp_fundno and CUSIPs.

We append the above two dataset together. Now a fund at least have a ticker or a CUSIP. They could have both. This leave us with 54,911 funds.

We merge this dataset to the annual dataset from CRSP.

Front Load

Since we define our C share class funds as funds that charge no front load, we need the information about fund's front load. For the front load data set we downloaded from CRSP. The total observations is 101,848.

1. In CRSP mutual fund front load dataset, for each crsp_fundno there is a pricing schedule for the front load. For each pricing schedule we only keep the maximum front load.

2. Then we delete the observations with front load equal to 0. That leave us with 19,626 observations.

3. We delete obs with front load smaller than 0. There are 30 of them.

4. There are 288 cases that a fund have more than one change in front load in one year. We choose to delete them.

5. We expand the front load dataset to a crsp_fundno year style. Because in the raw dataset, each entry have start year and end year. That gives us 108,818 entries.

6. We merge this back to the dataset generated in the above step.

Rear load

For the rear load data set we first download from CRSP. The total observartion is 151,194.

1. In CRSP mutual fund rear load dataset. For each crsp_fundno there is a pricing schedule for the rear load. For each pricing schedule we only keep the maximum front load.

2. Then we delete the obs with rear load equal 0. That leave us with 28,216 observations.

3. We delete obs with rear load smaller than 0. That delete 33 more obs.

4. There are 370 obs have more than one change in rear load in one year. We choose to delete them.

5. Now we expand the rear load dataset to a crsp_fundno year style. Because in the raw dataset, each entry have start year and end year. That gives us 162,099 obs.

6. We merge this back to the dataset generated in the above step.

Morningstar Clean-up

We start from the fund_ops file that we got from Morningstar. This dataset contains the fund_name, Morningstar category etc. It has 55,571 obs.

We first identify the non domestic well-diversified equity mutual fund. We follow the method provided in PST data appendix.

1. We first identified the observation with duplicated fund_names. And delete them. We also delete the funds with no MS category which is about additional 661 funds.

2. Then we identify the bond fund, international fund, sector fund, target date fund, real estate fund, other non-equity fund. The definition and methods are provided in PST. We attach the relevant pages in PST at the end of this file. Now we are left with 23592 funds.

3. We delete the funds with neither a ticker nor a CUSIP. We left with 21,580 funds.

For Morningstar, we utilize their information on the fund's category and whether a fund is index fund, fund family and portfolio id.

Merge between CRSP and Morningstar

In Morningstar, the unique identifier is secid. The goad is to get one to one mapping between crsp_fundno and secid. For details on secid, please check PST.

We use the CRSP dataset that have unique pairs between crsp_fundno and ticker or CUSIP to merge with Morningstar. First we merge on ticker. We got 12,412 matches.

Small issue in the Morningstar dataset is that the cusip is 9 digit while in CRSP, the CUSIP is 8 digit. Following the instruction on WRDS, we get rid of the last digit of CUSIP in Morningstar dataset.

Then we merge on CUSIP. We got 17,488 matches.

Finally we take the union of the two types of matches. We have 17,658 matches in total. Or 17,658 unique crsp_fundno and secid pairs.

We merge the above data to annual dataset from CRSP and keep the merged observations.

Correcting expense ratio, 12b-1, turn over and management fee

As mentioned in PST, the timing information for expense ratio, 12b-1 fee etc are not accurate in Morningstar. So we use CRSP dataset for those information. CRSP have 12b-1 fee information starting from 1992. We restrict our dataset to 1992 onwards. There are missing values in expense ratio, 12b-1, turn over and management fee. We want to fill in as many as possible.

So for the fund with missing value X, X can be {expense ratio, 12b-1, turn over and management fee}, we use the time series mean of the fund's X to replace the missing value. For example, if the fund miss expense ratio in 1996, we use the fund's lifetime average expense ratio to fill in the missing value in 1996.

Also for the X, we set -99 to missing and get the time series mean of it and replace the missing ones.

Follow the literature, our final dataset is at the fund level not share class level. For a lot of funds, there are many share classes. Different share classes are corresponded to different fee structure. For example, usually A charge front loads and a lower expense ratio. C share charges not loads but higher 12b-1 fee. To make the aggregation of 12b-1 fee reasonable, we make the following treatment. If a fund has a C share class, in a given year, then we set all the other share classes' 12b-1 fee and expense ratio to C share's 12b-1 fee and expense ratio. This treatment is based on the assumption that across different share classes, mutual funds should spend same amount in marketing. An alternative way is to annualized all the loads and add it back to the 12b-1 fee and then do the aggregation. We tried both ways. The results are not very different.

For the funds with no C share classes in a given year, we annualize the front load with 7 years and add it back to 12b-1 fee. In this case, we also increase the expense ratio by the amount of annualized front load.

Finally since there is a cap on 12b-1 fee at 1%, we set the upper bound of 12b-1 fee in our dataset to 1%.

We keep the observation with expense ratio smaller than 10 % and larger than 10 basis point. We also require expense ratio 5 basis point larger than 12b-1 fee.

Correct TNA and Return

As pointed out in PST, before 1993, a lot of the funds in CRSP dataset report their asset under management at quarterly or even annual frequency. But most of the funds report their return at monthly frequency. When we aggregate monthly returns across all the share classes, we need the monthly tna information. So we do the following correction.

1. For the funds who report their tna at quarterly frequency, we replace the missing value of tna with the tna in that quarter. For example, if fund report tna at month 3 for quarter 1, we replace month 1 and month 2 tna as month 3 tna.

2. For the funds who report tna at annual frequency, we replace month 1 to month 11 tna as month 12's tna.

3. If there is no tna information for any month in a year. We delete this year.

After this correction, we have 2,018,242 observations with non missing monthly return and tna.

General Cleaning

We drop the institutional shares which are identified in the following ways: 1 CRSP inst_fund is 'Y'. 2 fund name contains 'Institutional Shares', 'Institutional Class', 'Inst'.

We merge the annual data set developed above with monthly dataset on fund's return, asset under management from CRSP.

Following PST, we exclude fund/month observations with expense ratios below 10 basis point per year, since it is extremely unlikely that any actively managed funds would charge such low fees. We exclude observations with lagged fund size below $15 million.

Identifying Index Funds

In order to identify index funds, we use a simple two steps procedures following PST:

1. If either CRSP or Morningstar indicate an index fund, we label this fund as index fund.

2. If a fund's name contains words 'Index' or 'index', we label it as index fund.

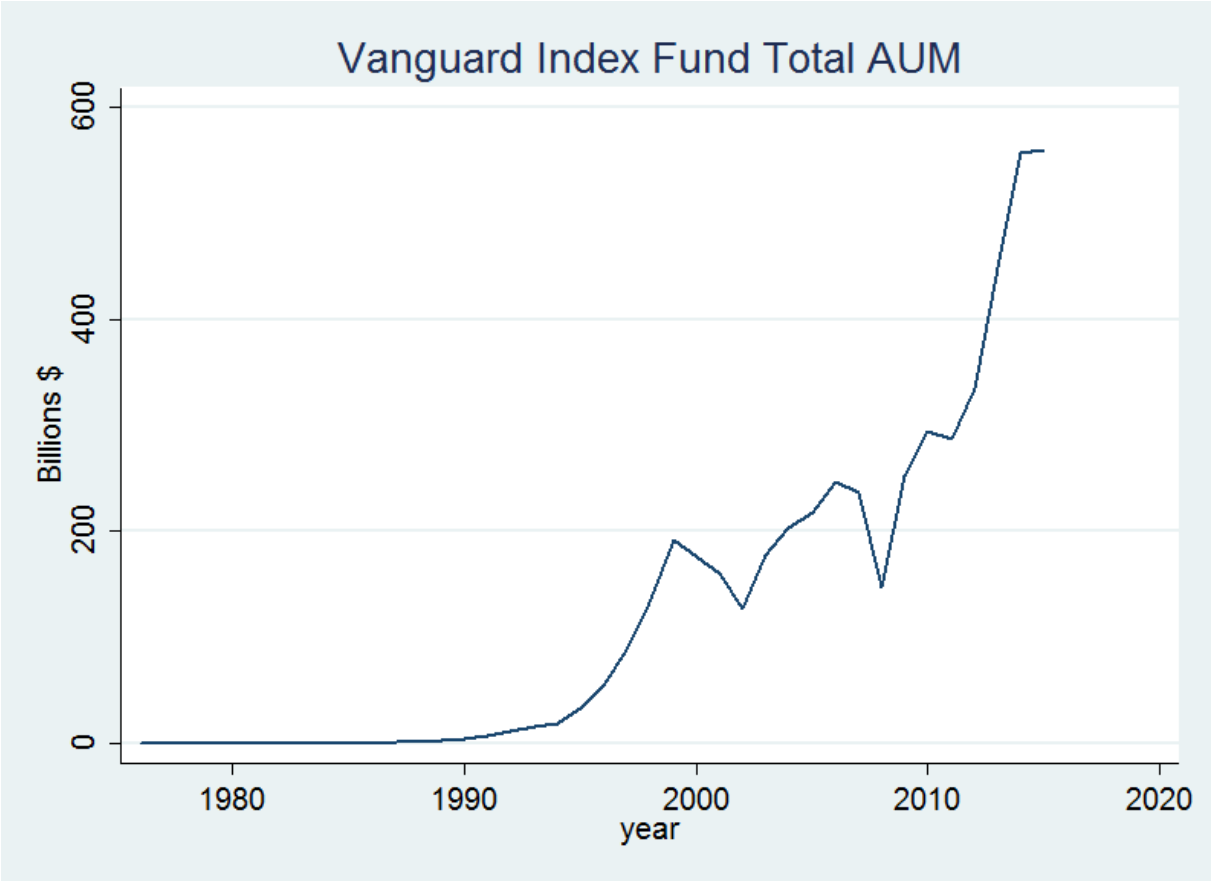
As a result of this procedure, we have 1,480 index funds.

Vanguard Index Fund

As proposed in BB, index funds from Vanguard are the most accessible index funds to the average investors. We further label whether an index fund is from Vanguard by checking with the fund name or the management firm name contains 'Vanguard'. If it does, we label it as Vanguard index fund.

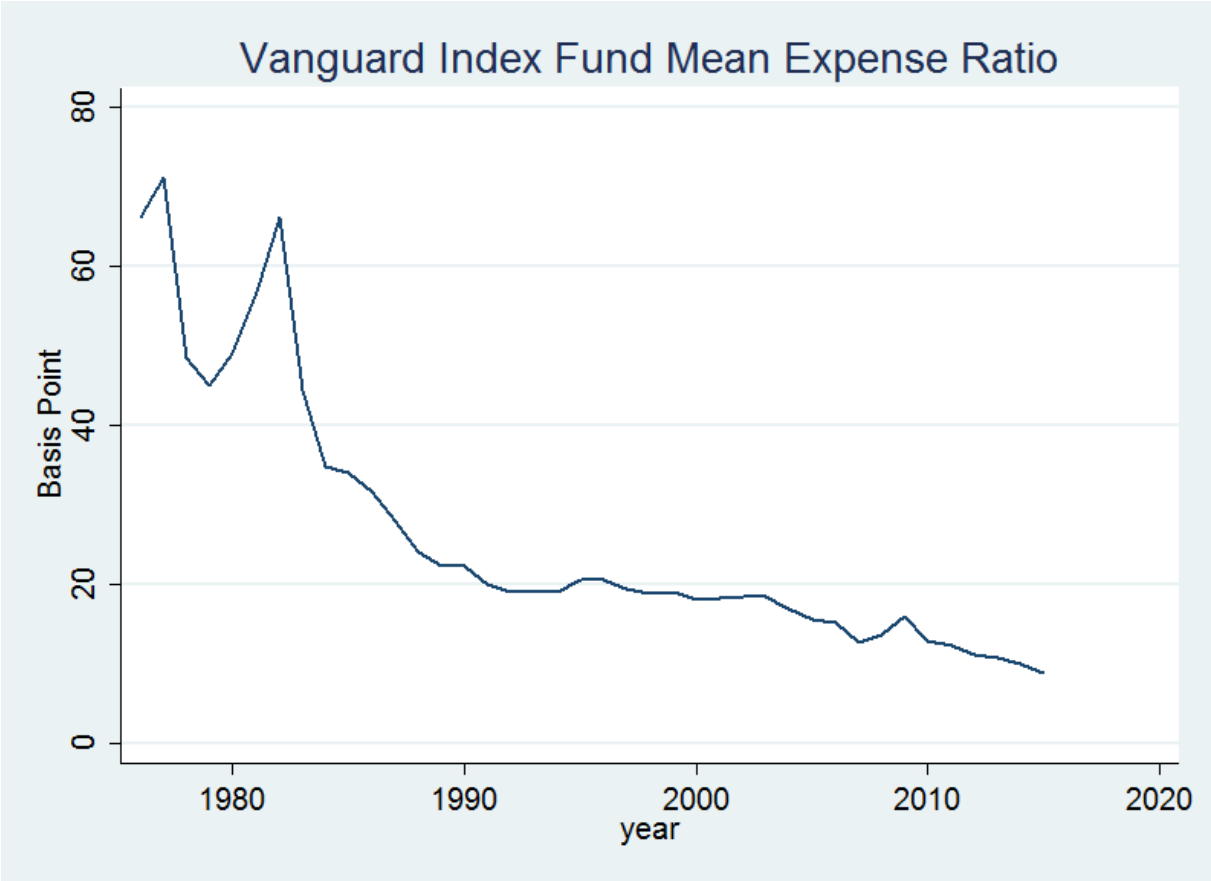
In the final dataset, we use all of the index funds from Vanguard, combined, as the outside good. In each month, we get the total asset under management, asset weighted mean of management fee, returns, expense ratios, turnover ratios, and 12b-1 fees. Then for each month we only keep one observation for the Vanguard index fund.

Figure 1: Total Asset Under Management of all the Vanguard Equity Index Funds



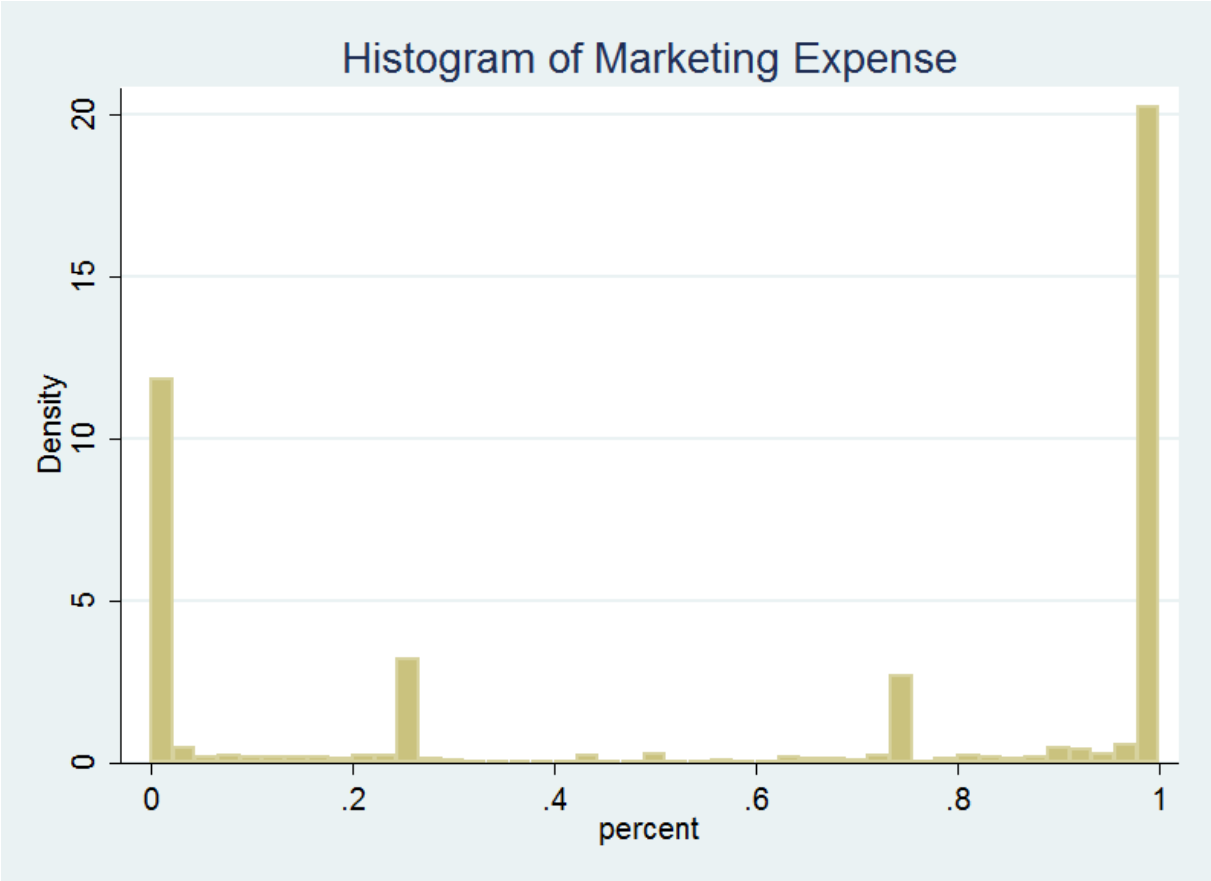
This figure plots the total asset under management of all the Vanguard equity index funds. The data series starts from 1975 when Vanguard launching its S&P 500 index fund. In year 2015, the total AUM reaches over 600 billions of dollars.

Figure 2: Value Weighted Mean Expense Ratio of Vanguard Equity Index Fund



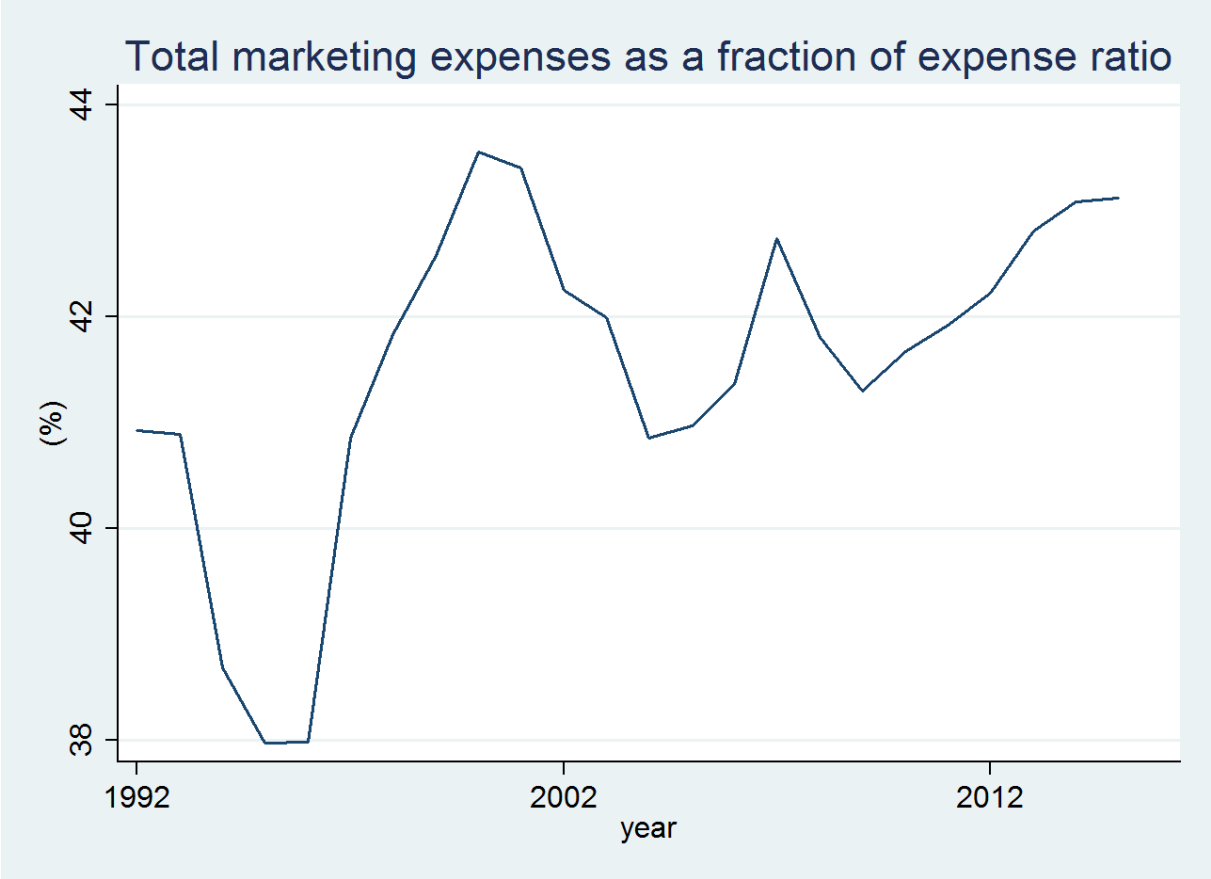
This figure plots the asset weighted mean expense ratio for all the Vanguard equity index funds. We can see a significant drop from over 60 bp in 1975 to under 10 bp in 2015.

Figure 3: Histogram of Effective Marketing Expenses



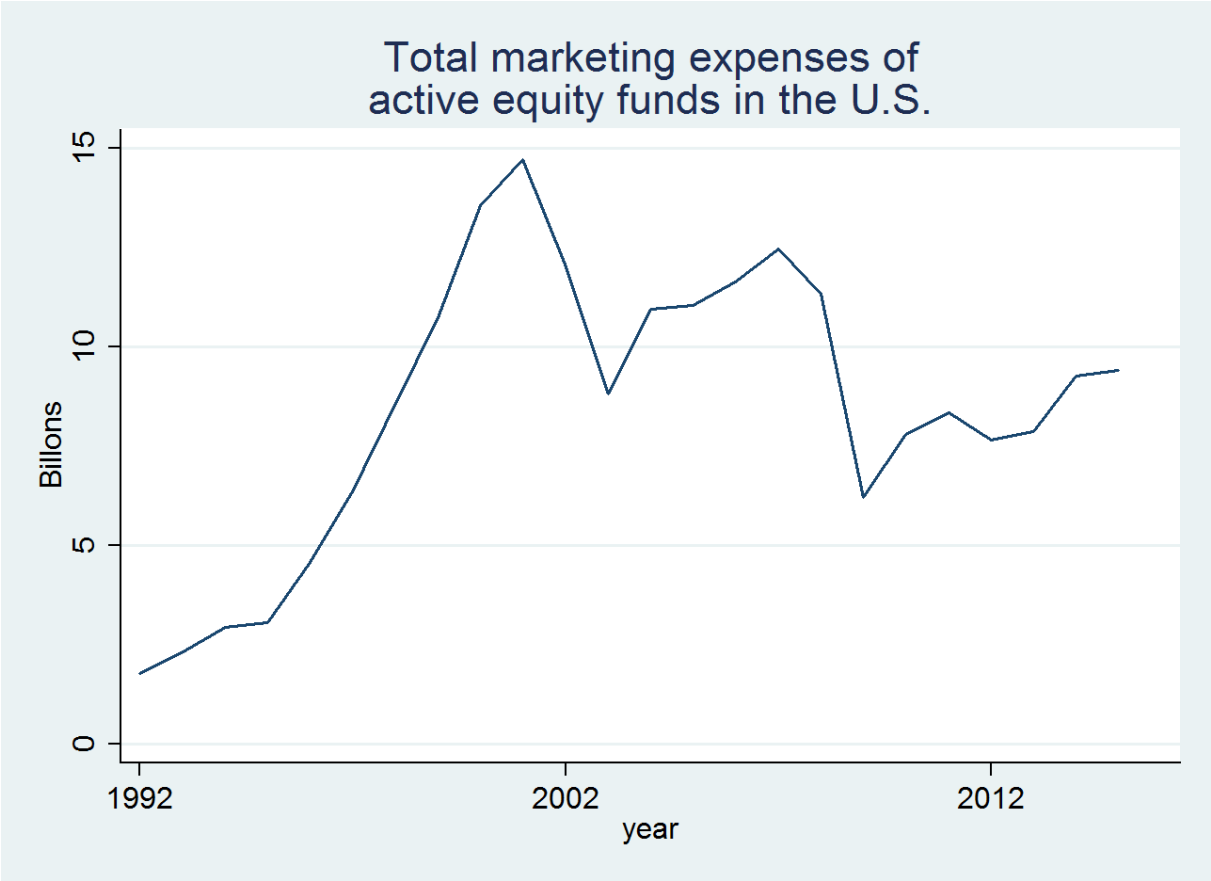
This figure plots the histogram of effective marketing expenses. About 45.7% of the observations are binding at the upper bound, 1% level. And about 23.7% of the observations are binding at 0%.

Figure 4: Total marketing expenses as a fraction of expense ratio for active funds



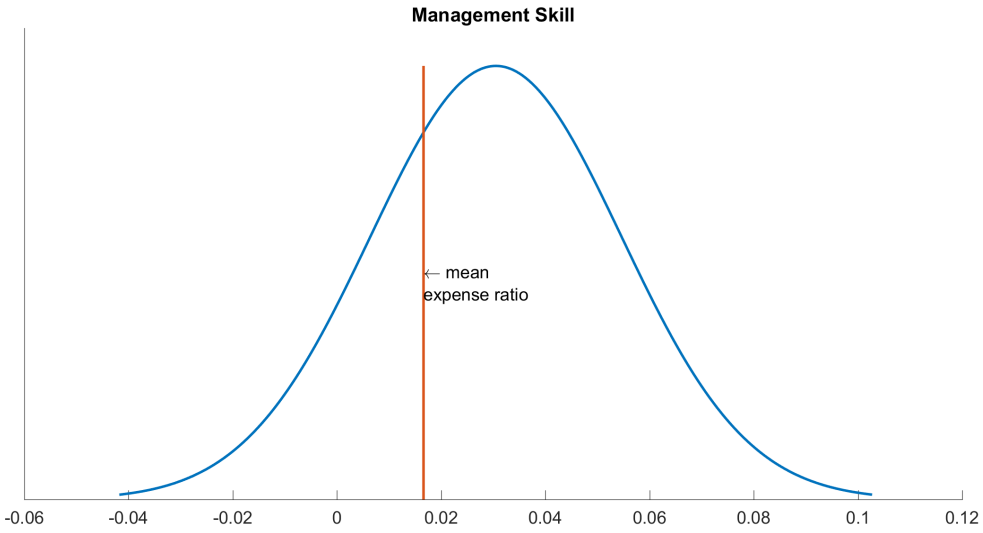
This figure plots the ratio between total marketing expenses and total expense ratios for active funds.

Figure 5: Total marketing expenses for the active funds



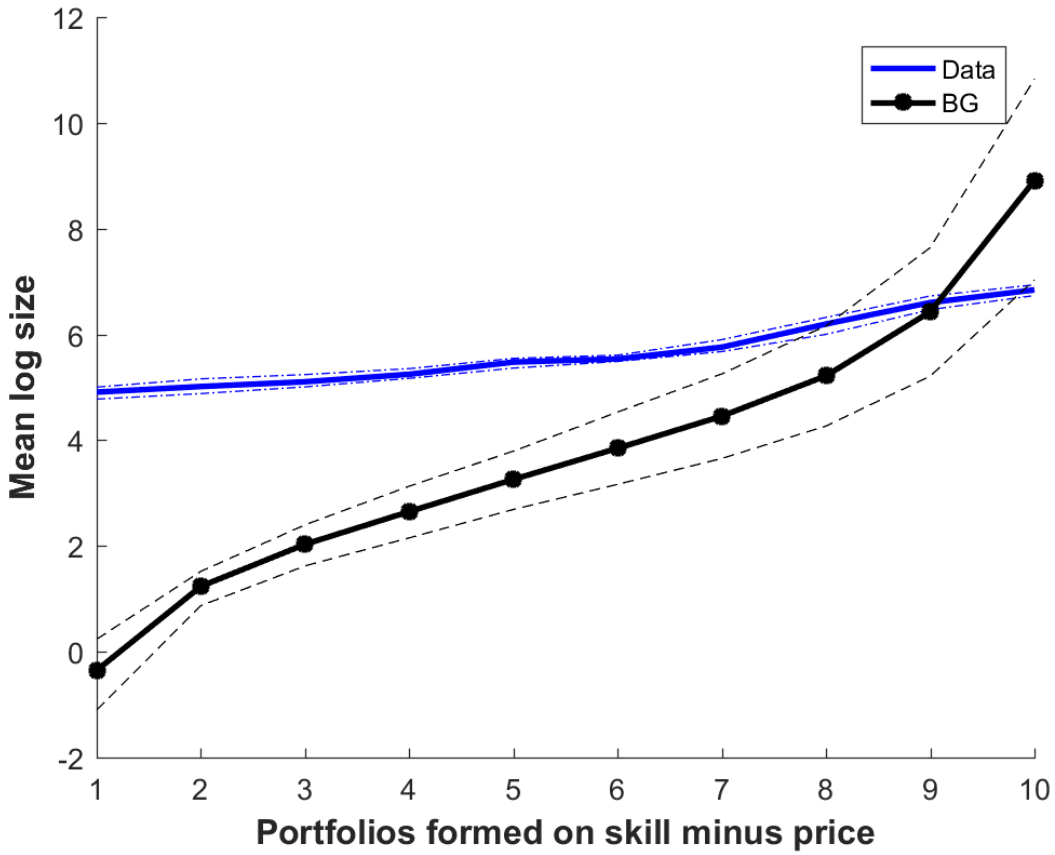
This figure plots the total marketing expenses (in dollars) for all the active equity mutual funds in the U.S. from 1992 to 2015. The funds' asset under management are inflated to 2015 dollars by using Consumer Price Index downloaded from FRED.

Figure 6: Prior Distribution of Manager Skill



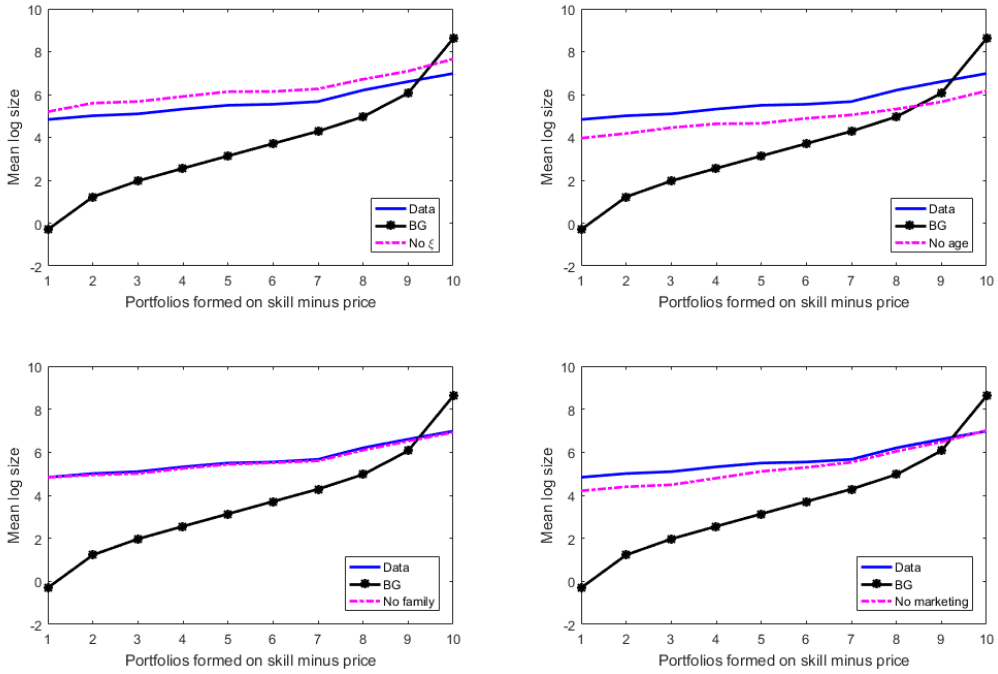
This figure presents the prior distribution of management skill. The vertical line marks the mean of expense ratio in our data. Approximately 71% of funds have management skills higher than the mean expense ratio.

Figure 7: Capital (mis)Allocation in Mutual Funds: Size vs. Skill



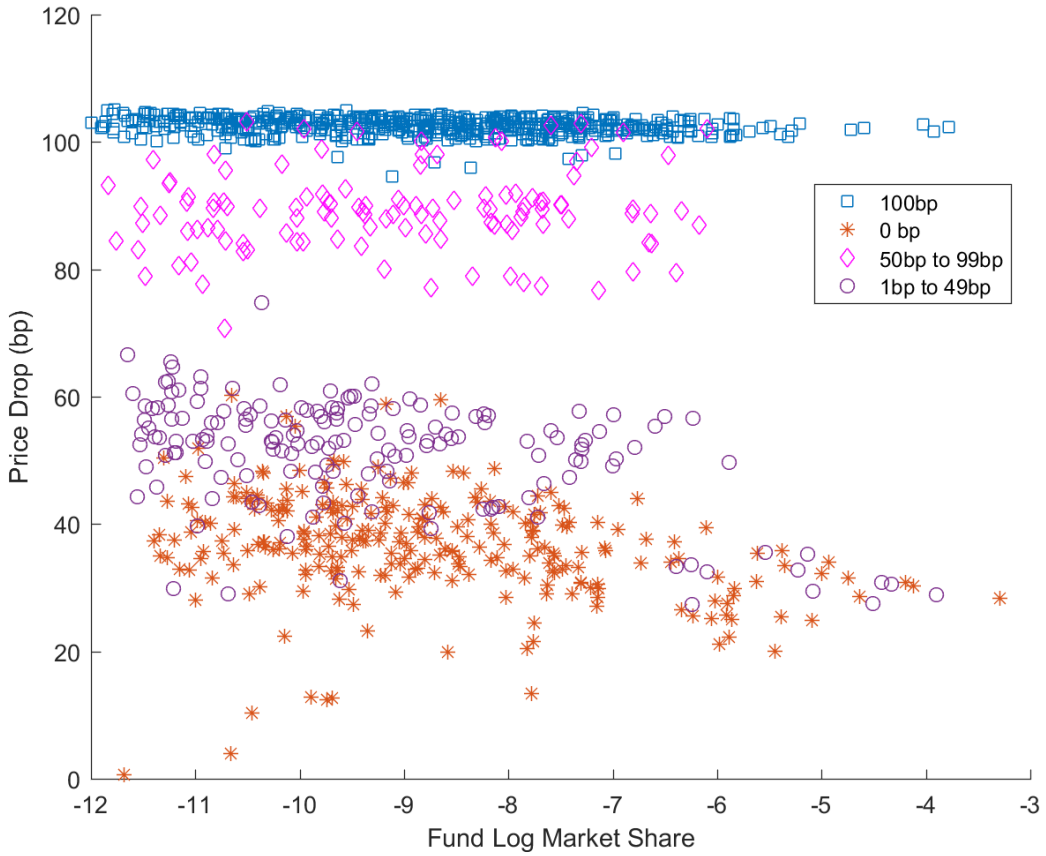
This figure plots the mean log fund size within portfolios formed on skill net of fees. The black line plots the BG predicted funds sizes and the blue line plots the data. We construct ten portfolios of mutual funds base on the deciles of net skill. Portfolio 1 has the lowest skill (net alpha) while portfolio 10 has the highest skill. 95 percentile confidence bound are provided in dash lines.

Figure 8: Counterfactual (restricted model)



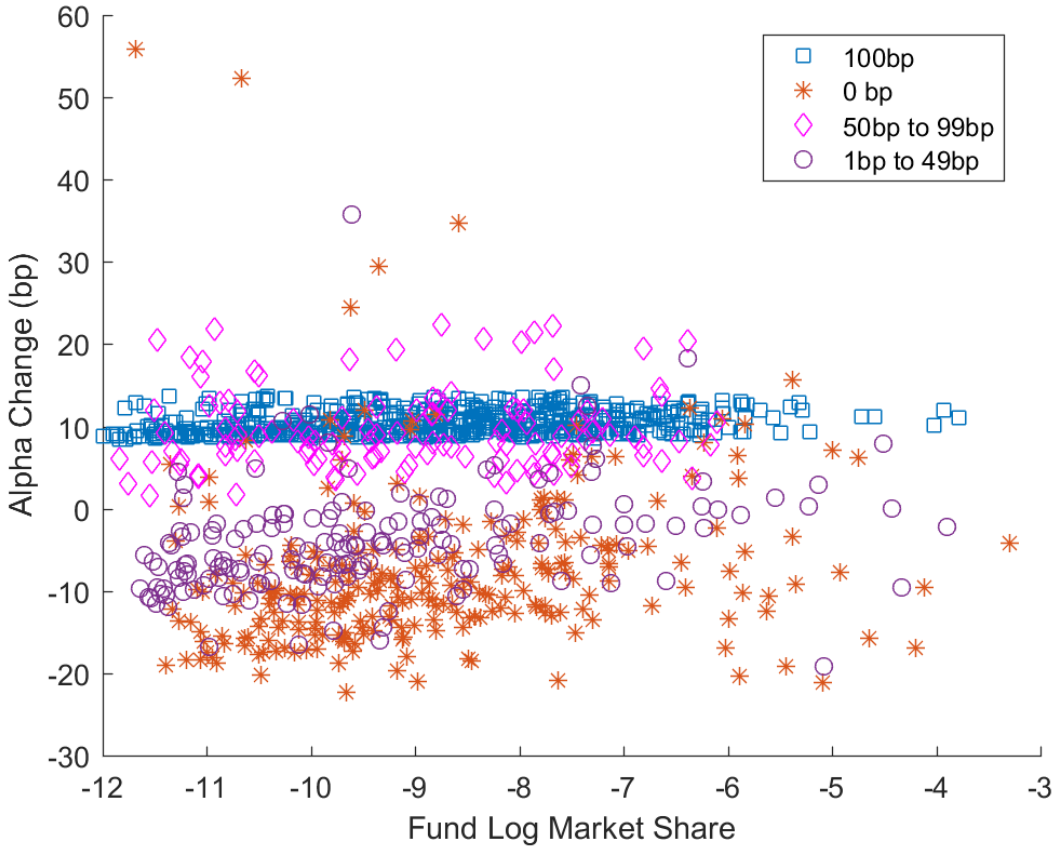
The four figures plot the restricted model's implied fund size. Black line is the BG model implied fund size. Blue line is the data. Purple line is the restricted model implied fund size. In top left figure, purple line plots fund size when there is no ξ. In top right figure, purple line plots fund size when there is no age. In bottom left figure, purple line plots fund size when there is no fund family size. In bottom right figure, purple line plots fund size when there is no marketing. The definition for the portfolios is similar to figure 7.

Figure 9: Price Change from Current Equilibrium to No-Marketing Equilibrium



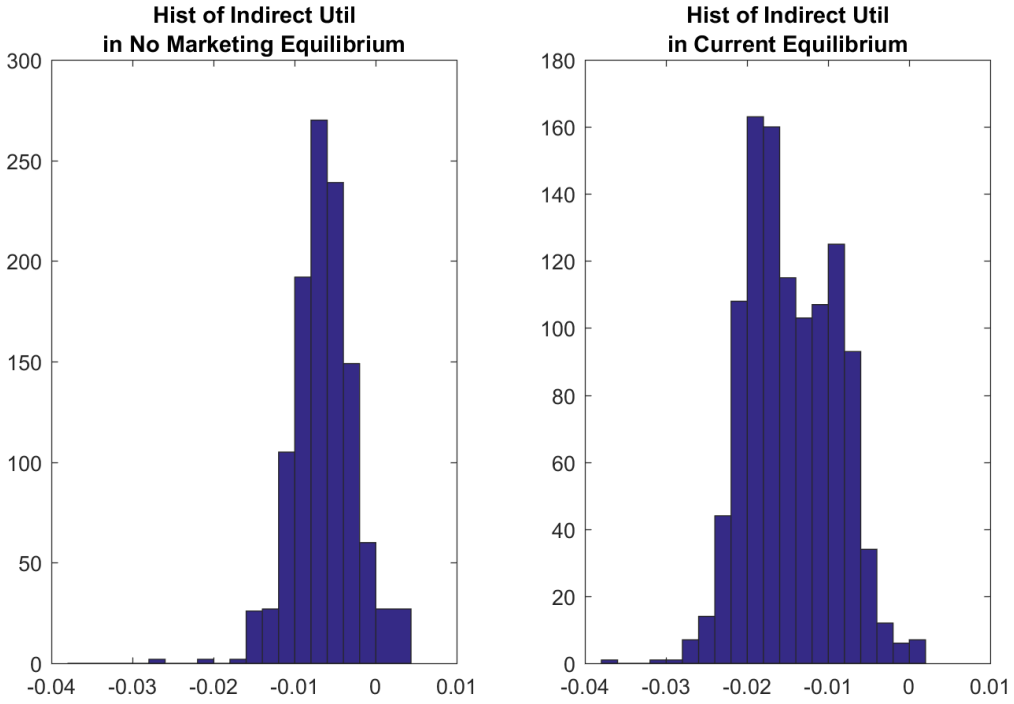
This figure plots the price drop from current equilibrium to no marketing equilibrium.

Figure 10: Alpha: No-Marketing Equilibrium vs. Current Equilibrium



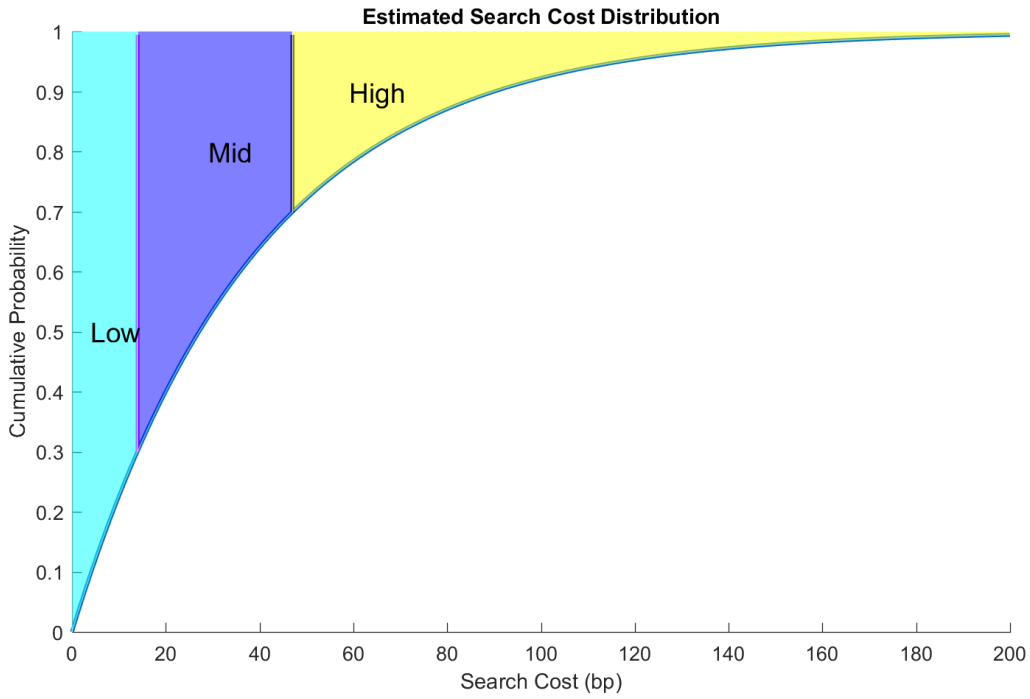
This figure plots the alpha difference between the no-marketing equilibrium and current equilibrium.

Figure 11: Indirect Utilities: Current Equilibrium vs. No-Marketing Equilibrium



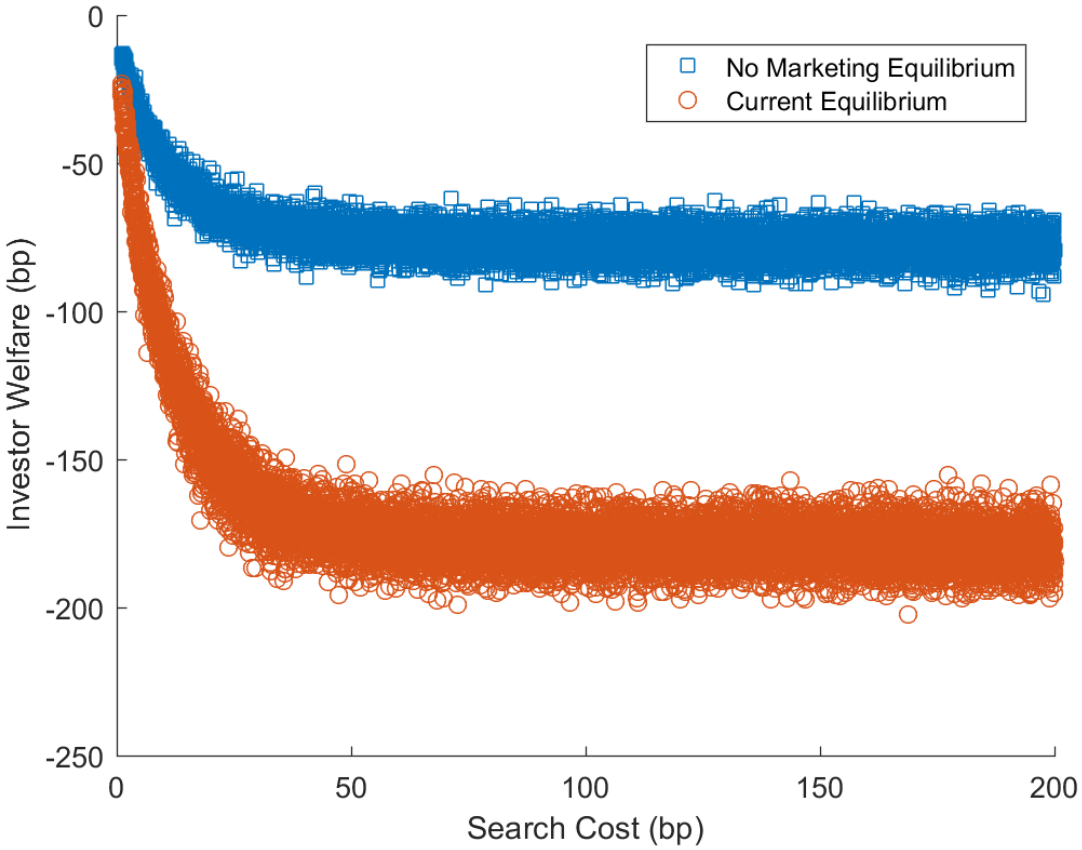
This figure plots the indirect utilities of funds in the no marketing equilibrium as well as the current equilibrium. We can see that according to various measure of dispersion, the current equilibrium is much more dispersed than the no marketing equilibrium.

Figure 12: Estimated Search Cost Distribution



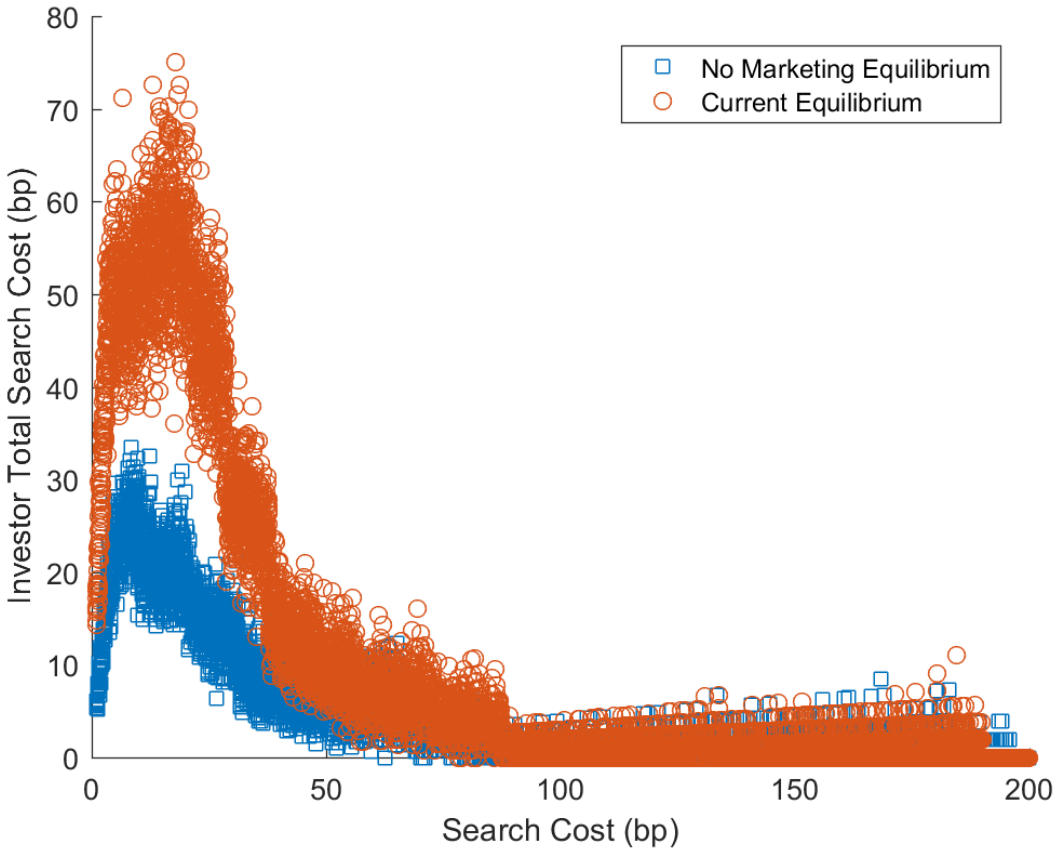
This figure plots the estimated search cost distribution. We assume search cost follows an exponential distribution. Our estimated mean search cost is 39 basis point. We define the search cost lower than 30 percentile as low search cost, search cost higher than 70 percentile as high search cost and the rest as intermediate search cost.

Figure 13: Relationship between Investor's Search Cost and Welfare



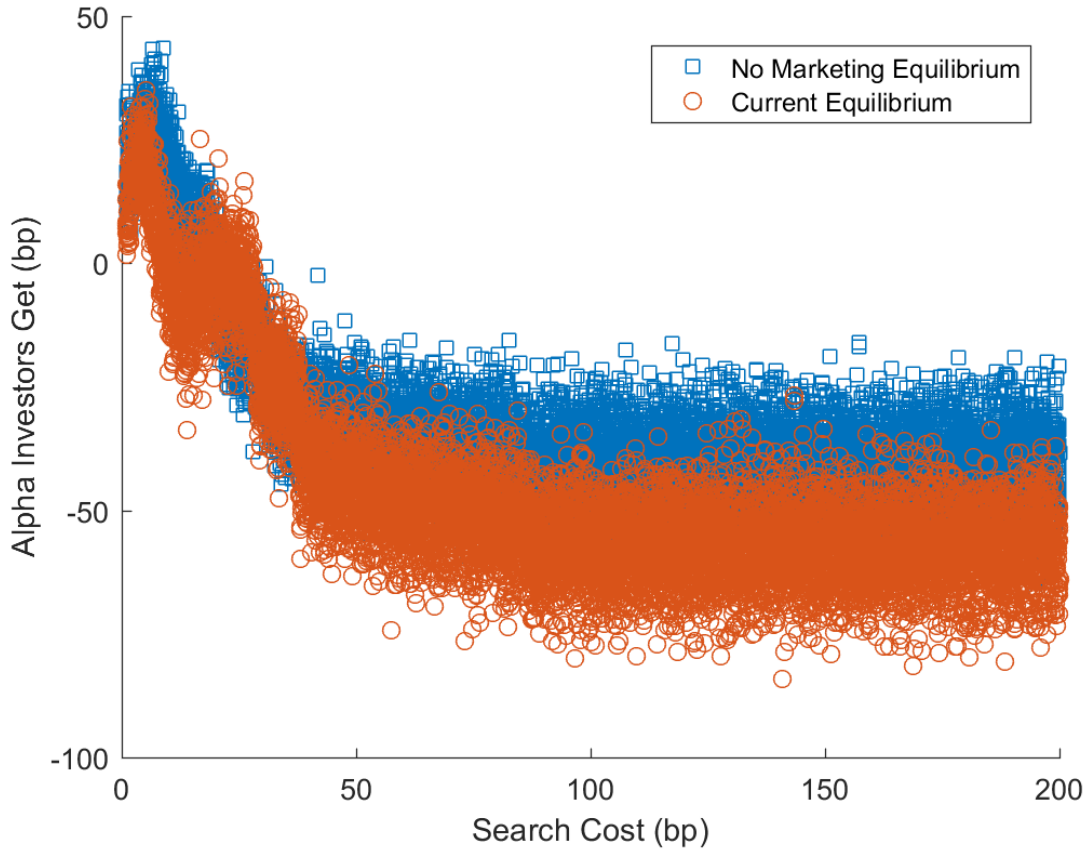
This figure plots the investor's welfare against investor's search cost level. The investor's welfare is in unit of bp.

Figure 14: Relationship between Investor's Search Cost and Total Search Cost



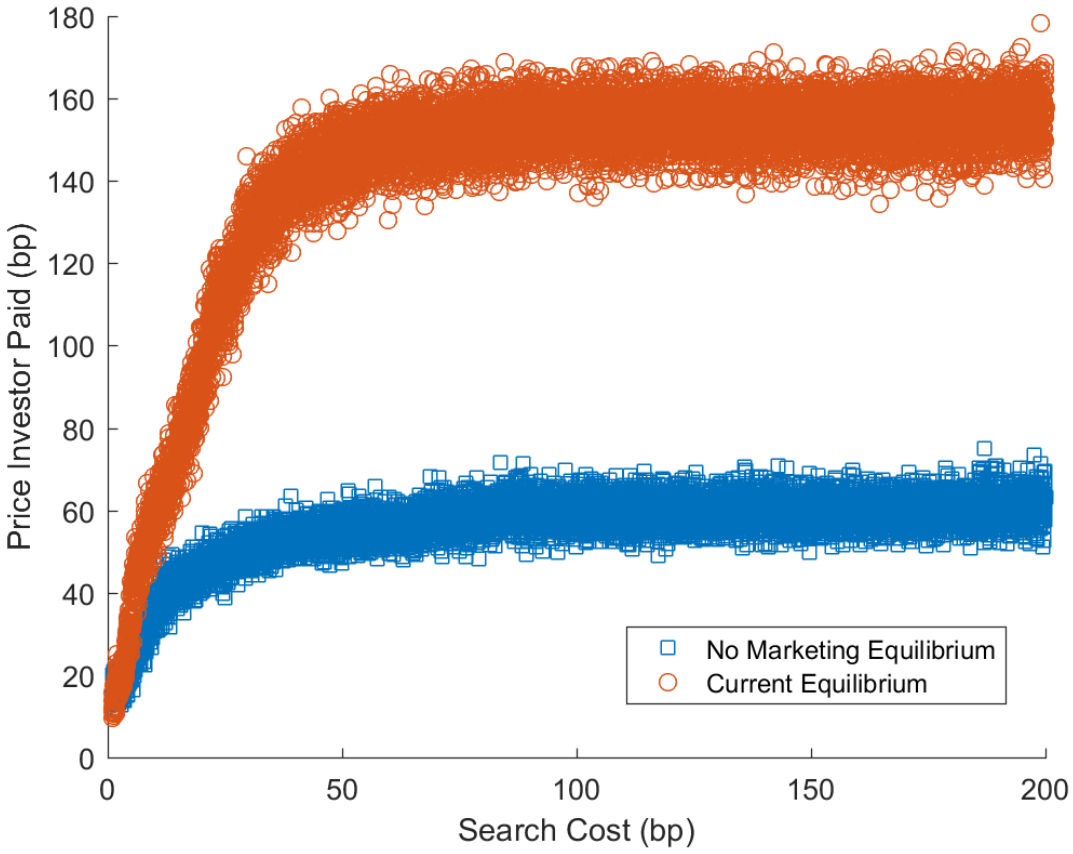
This figure plots the investor's total search cost against investor's unit search cost level.

Figure 15: Relationship between Investor's Search Cost and Alpha



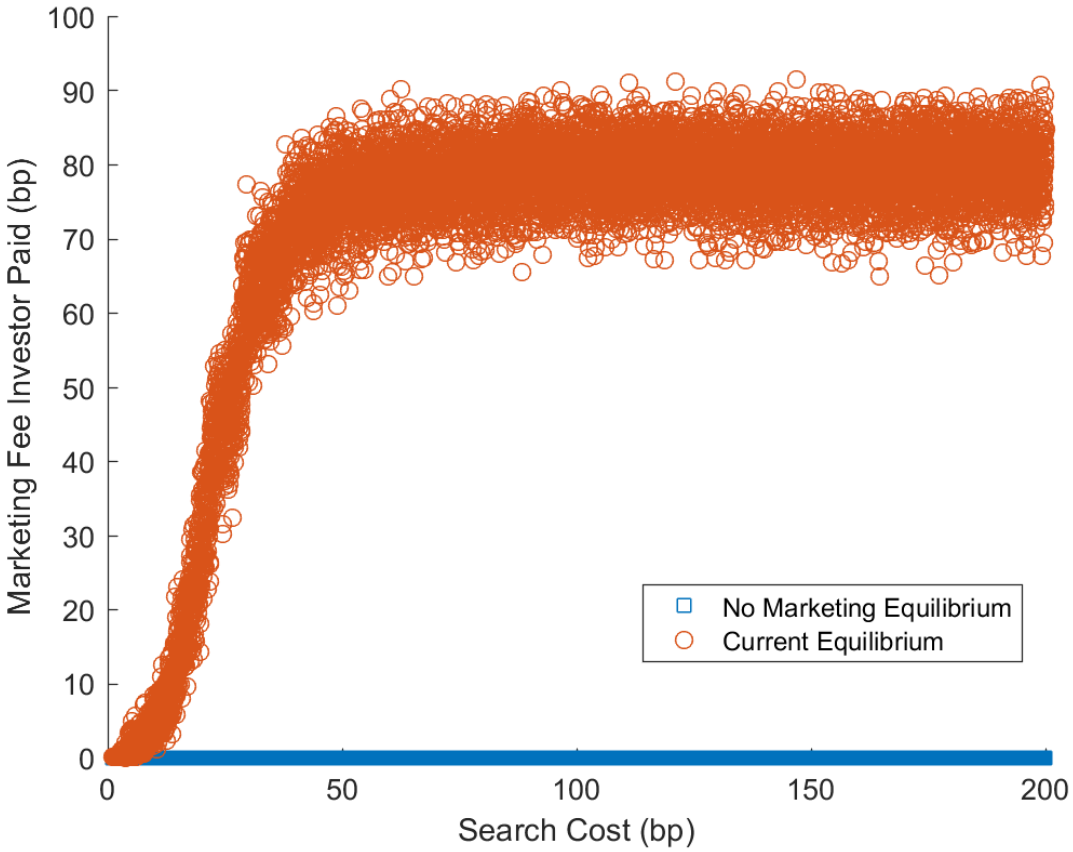
This figure plots the alpha that investors get against investor's search cost level.

Figure 16: Relationship between Investor's



This figure plots the price that investors paid against investor's search cost level.

Figure 17: Search Cost and Marketing Expenses



This figure plots the marketing fees paid by investors against their search costs.

Table 1: Data Definition

Variable	Definition
Fund AUM	Fund's total net asset under management at the beginning of each year, in unit of millions of dollars
Fund expense ratio	The ratio between operating expenses that shareholders pay to the fund and the fund's AUM
Actual 12b1	Reported as the ratio of the AUM attributed to marketing and distribution costs
Management fee	The ratio of the AUM attributed to fund management costs
Fund turnover	Minimum (of aggregated sales or aggregated purchases of securities), divided by the average 12-month AUM of the fund
Total market	Sum of all funds' AUM including both active funds and index fund
Market share	Ratio between fund's AUM and total market in the same year
Age	Approximated by the number of years showed up in the sample
Family size	Number of funds in the same fund family
CAPM α	Outperformance estimated by CAPM
FF3 α	Outperformance estimated by Fama French 3 factor model
FFC α	Outperformance estimated by Fama French and Carhart model
FF5 α	Outperformance estimated by Fama French 5 factor model
New	Dummy which equals 1 if fund is new in the current period
Index fund price	Fund expense ratio of the index fund

This table presents the data definition of all the variables used in the paper. For detailed data construction process, please check the data appendix.

Table 2: Summary Statistics

				Percentiles		
	Num of Obs	Mean	Stdev	25%	50%	75%
FF5 α	27,621	0.0054	0.0798	-0.0347	0.0007	0.0379
Fund AUM	27,621	1339	4791	82	254	886
Fund exp ratio	27,621	0.0166	0.0053	0.0123	0.0175	0.0205
Marketing expense	27,621	0.0061	0.0044	0.0001	0.0089	0.0100
Market share	27,621	0.0018	0.0066	0.0001	0.0002	0.0009
Age	27,621	11.46	10.3	4	8	16
New dummy	27,621	0.0827	0.2755	0	0	0
Family size	27,621	12.08	13.15	3	7	17
index fund price	27,621	0.0017	0.0009	0.0013	0.0017	0.0019
Total market AUM	27,621	1.54E+06	7.49E+05	1.26E+06	1.77E+06	2.13E+06
Family AUM	27,621	27,826	77,700	729	4,920	15,787
FFC α	27,621	0.0055	0.0786	-0.0341	0.0025	0.0397
FF3 α	27,621	0.0065	0.0813	-0.0339	0.0025	0.0409
CAPM α	27,621	0.0097	0.0968	-0.0376	0.0045	0.0492

This table presents summary statistics for our sample of U.S. equity mutual funds. For detailed data defenition, please check table 1. The sample period is from 1964 to 2015. Our unit of observation is fund/year.

Table 3: Investor Beliefs and Manager Skill (BG)

Parameters	Description	1964-2015	1975-2015	1985-2015
η	Decreasing returns to scale	0.0048	0.0048	0.0040
		(0.0004)		
μ	Mean of prior	0.0305	0.0298	0.0245
		(0.0025)		
κ	SD of prior	0.0241	0.0242	0.0218
		(0.0012)		
δ	SD of realized alpha	0.0762	0.0752	0.0752
		(0.0005)		
ρ	Skill persistence	0.9485	0.947	0.9620
		(0.0185)		
Log Likelihood		1.1185	1.1309	1.1368
Number of Obs		27,621	26,682	25,369

This table presents the estimates of the fund performance related parameters. The standard errors are in the parentheses. η is the decreasing returns to scale parameter. μ is the mean of manager's skill distribution. κ is the standard deviation of skill distribution. δ is the standard deviation of the idiosyncratic noise added to the realized alpha. ρ is the persistence of the manager's skill. Due to computational burden, we didn't provide the standard errors for the estimates of the last 2 columns.

Table 4: Value Added: Model vs. Data

	BG	Data	BG-Data
Panel A: Aggregate			
Total Value Added	3.281e5	16,275	3.118e5
Mean Value Added	12.12	0.6012	11.52
Panel B: By Portfolio			
Portfolio 1 (Lowest)	42.88	-11,273	11,315
Portfolio 2	161.0	-10,794	10,956
Portfolio 3	349.7	-7,966	8,316
Portfolio 4	610.7	-18,391	19,002
Portfolio 5	1,077	-12,735	13,812
Portfolio 6	1,856	-5,972	7,829
Portfolio 7	2,972	-3,911	6,884
Portfolio 8	5,917	-25,569	31,487
Portfolio 9	30,339	10,137	20,202
Portfolio 10(Highest)	284,831	102,753	182,077

This table provides the value added for BG model and the data. In Panel A, total value added is the sum across all funds across all years in our sample. Mean value added is just the total value added divided by the number of observations in our sample. The numbers are in millions of 2015 dollars. In Panel B, we provides total value added for each portfolio in BG model and in the data.

Table 5: Search Model Parameters

Parameters	Description	Interior	Lower bound	Upper bound	All
λ	Mean search cost (bp)	39	39	39	39
		(4e-4)	(4e-4)	(4e-4)	(4e-4)
γ	Alpha coef	0.4153	0.4154	0.4164	0.4159
		(0.0302)	(0.0317)	(0.0300)	(0.0303)
θ	Marketing coef	113.11	111.22	133.18	122.56
		(7.3344)	(7.2917)	(8.7978)	(7.39728)
β_1	Number of family funds coef	0.4048	0.4033	0.3811	0.3933
		(0.0262)	(0.0261)	(0.0262)	(0.0260)
β_2	Log fund age coef	1.0324	1.0323	1.0322	1.0323
		(0.0372)	(0.0381)	(0.0370)	(0.0360)
Year FE		Yes	Yes	Yes	Yes

This table presents the estimates of the structural search model. We use the data from 2001 to 2015. The four columns are corresponding to four sets of moment conditions as we described in the estimation section. In columns Interior, we use the funds that are not binding in their marketing expenses to estimate the model. In columns Lower bound, we use the funds that are binding in their marketing expenses at the lower bound to estimate the model. In columns Upper bound, we use the funds that are binding in their marketing expenses at the upper bound to estimate the model. In columns All, we use all the funds to estimate the model.

Table 6: Change in Size when Marketing Expenses Increase by 1 bp

	Lower	Interior	Upper
Panel A: Sort by Size			
Big Funds	0.8735	0.8904	1.043
Intermediate Size Funds	0.8794	0.8965	1.050
Small Funds	0.9085	0.9261	1.085
Panel B: Sort by Skill			
High Skill Funds	0.9670	0.9858	1.155
Intermediate Skill Funds	0.8987	0.9161	1.073
Low Skill Funds	0.8154	0.8311	0.973
Panel C: Sort by Original Marketing Expense			
Binding at Lower Bound	0.9554	0.9739	1.141
Non Binding	0.8990	0.9165	1.073
Binding at Upper Bound	0.8413	0.8575	1.004

This table provides the percentage changes in funds size for various groups of funds if marketing expense increases by 1 bp. Large funds in panel A are funds in the top 10 percentile. Small funds in panel A are funds in the bottom 10 percentile. Intermediate size funds are the rest. High Skill funds in panel B are funds in the top 10 percentile. Low skill funds in panel B are funds in the bottom 10 percentile. Intermediate skill funds are the rest of the funds. In panel C, Binding at Lower Bound funds are funds who originally choose 0 marketing expenses. Binding at Upper Bound are funds who originally choose 1% marketing expenses. Non binding funds are the rest of the funds.

Table 7: Change in Profit when Marketing Expense Increase by 1 bp

	Lower	Interior	Upper
Panel A: Sort by Size			
Big Funds	-0.4317	-0.4150	-0.2645
Intermediate Size Funds	-0.0311	-0.0143	0.1381
Small Funds	0.0850	0.1024	0.2602
Panel B: Sort by Skill			
High Skill Funds	-0.1267	-0.1081	0.0597
Intermediate Skill Funds	-0.3358	-0.3186	-0.1634
Low Skill Funds	-0.2128	-0.1972	-0.0567
Panel C: Sort by Original Marketing Expense			
Binding at Lower Bound	-0.2100	-0.1917	-0.0263
Non Binding	-0.1722	-0.1550	0.0006
Binding at Upper Bound	-0.4180	-0.4019	-0.2569

This table provides the percentage changes in funds' profits for various groups of funds if marketing expense increases by 1 bp. Large funds in panel A are funds in the top 10 percentile. Small funds in panel A are funds in the bottom 10 percentile. Intermediate size funds are the rest. High Skill funds in panel B are funds in the top 10 percentile. Low skill funds in panel B are funds in the bottom 10 percentile. Intermediate skill funds are the rest of the funds. In panel C, Binding at Lower Bound funds are funds who originally choose 0 marketing expenses. Binding at Upper Bound are funds who originally choose 1% marketing expenses. Non binding funds are the rest of funds.

Table 8: Quantifying the Importance of Sampling Probability Components

	ξ	age	num of family funds	marketing	skill	price	R^2	Correlation between q^{BG} and q^{Model}
Model 1		Y	Y	Y	Y	Y	0.5169	0.2988
Model 2			Y	Y	Y	Y	0.2122	0.4988
Model 3				Y	Y	Y	0.1614	0.5698
Model 4					Y	Y	0.1152	0.5947
Model 5	Y	Y	Y		Y	Y	0.9157	0.1456
Model 6	Y	Y	Y	Y		Y	0.9008	0.0301
Model 7	Y	Y	Y	Y	Y		0.9005	0.0776
Data							1	0.0901

We first compute the model predicted market share with some of the components in sampling probability being removed. The column in the table with no "Y" indicates the variable we remove. Then we regress log market share in the data onto log market share as predicted by the model and a constant. We report the R squared of each regression in the table. The data period is from 2001 to 2015. We also compute the correlation between our model implied fund size and Berk and Green model implied fund size.

Table 9: Parameters for simulation 1

η	λ	γ	θ	ω	ζ
0.0049	0.0031	0.4539	250	0	0

This table presents the parameters used for simulation 1. Compare them with the baseline estimation in table 5, we increase the effectiveness of marketing θ from around 100 to 250. This gives funds more incentives to differentiate from other funds. We also decrease the mean search cost from 39 bp to 31 bp. To facilitate the analysis, we also set $\omega$and ζ to be zeros.

Table 10: Simulation 1 results

Regulation cap (bp)	0	25	50	75	100
Correlation between $q^{BG}$and q^{Model}	0.27	0.35	0.45	0.56	0.61
Total value added	-2,010	-1,130	-225.7	325.7	517.7
Total welfare	-0.0016	-0.0012	-0.0009	-0.0008	-0.0007

This table presents the results from counterfactual experiment 1. The trend is very clear: as relaxing regulation, Correlation between $q^{BG}$and q^{Model}, total value added and total welfare all increase.

Table 11: Summary of Outcomes for Current Equilibrium and No-Marketing Equilibrium

	Current	No Marketing
Mean price (bp)	160.27	82.96
Mean marketing (bp)	61.29	0
Mean alpha (bp)	37.24	41.07
Total share of active funds	0.74	0.67
Mean sampling prob (%)	0.085	0.078
Sampling prob for low price funds (%)	0.042	0.14
Sampling prob for index funds (%)	5.91	13.66
Investor welfare (bp)	-140.72	-61.25
Active funds average profit (bp)	57.51	42.19
Passive funds average profit (bp)	2.32	2.86
Total Welfare	-37.37	-16.20
Investor's Search Cost (bp)	29.09	12.15

This table provides various measures of the mutual fund industry under current and no marketing equilibrium.

Table 12: Summary of Outcomes for Different Search Costs

	Low λ 20 bp	Mid λ 35 bp	High λ 39bp
Mean price (bp)	58.52	136.24	160.27
Mean marketing (bp)	0	44.78	61.28
Mean alpha (bp)	38.94	39.00	37.24
Total share of active funds	0.6474	0.7109	0.7412
Mean sampling prob (%)	0.0784	0.0843	0.0854
Sampling prob for low price funds (%)	0.1516	0.0493	0.0420
Sampling prob for index funds (%)	13.66	7.161	5.915
Investor welfare (bp)	-48.42	-118.41	-140.71
Active funds average profit (bp)	31.97	51.46	57.51
Passive funds mean profit (bp)	3.157	2.588	2.317
Total welfare (bp)	-13.98	-33.04	-37.37
Investor's search cost (bp)	9.189	25.75	29.09

This table presents various measures of the mutual fund industry under different search costs distributions.