WORKING PAPERS

N° 1220

May 2021

"Information Loss over the Business Cycle"

Patrick Coen

Toulouse School of Economics

Information Loss over the Business Cycle*

Patrick Coen†

May 2021

Abstract

The composition of the mutual fund industry changes through entry and exit over the business cycle. Entrants may on average be higher quality than exiting funds (a cleansing effect that improves welfare), but they have no returns history and so investors have less precise beliefs about their ability (an information loss effect that harms welfare). I find that the net effect of this firm turnover is negative in the short-term but turns positive as the effect of information loss decays over time. I show that older funds should optimally be subsidized during recessions to preserve their socially valuable returns history.

*I am particularly grateful to Alessandro Gavazza and Jamie Coen for many helpful discussions. I am also grateful for comments by Ulrich Hege, Christian Julliard, Martin Oehmke and Mungo Wilson, as well as seminar participants at the London School of Economics, the Toulouse School of Economics and Compass Lexecon. I acknowledge the financial support of the Economic and Social Research Council, the French National Research Agency (ANR) under the Investments for the Future program (Investissements d'Avenir, grant ANR-17-EURE-0010), and also of the TSE-Partnership foundation, a research organization under the aegis of Toulouse School of Economics, sponsored by several partners (the list of which is available at www.tse-fr.eu/tsepartnership).

†Toulouse School of Economics, University of Toulouse Capitole, Toulouse, France. patrick.coen@tse-fr.eu.

Conflict of interest disclosure statement

Patrick Coen

- **Employment**: From 2019 to 2020 I was employed by the Financial Conduct Authority, United Kingdom.
- **Research funding**: I acknowledge financial support from the TSE-Partnership foundation, a research organization under the aegis of Toulouse School of Economics, sponsored by several partners (the list of which is available at www.tse-fr.eu/tsepartnership). I also acknowledge the financial support of the Economic and Social Research Council and funding from the French National Research Agency (ANR) under the Investments for the Future program (Investissements d'Avenir, grant ANR-17-EURE-0010).
- **Other**: I have been an academic vistor at the Bank of England since 2016.

I have no other potential conflicts to disclose.

1 Introduction

The business cycle induces firm turnover: firms exit in recessions, and enter in recoveries. What impact does this firm turnover have on outcomes post-recovery? How persistent is this impact? These questions then naturally give rise to questions about policy: which firms, if any, should a social planner support through recessions so they do not exit?

I study a key trade-off that underpins these questions. On the one hand, the business cycle can improve outcomes by "cleansing" the market of low quality firms: replacing low quality firms that exit during the recession with higher quality firms that enter during the subsequent recovery. On the other hand, the firms that exit have a track record of performance, whereas the entrants that replace them do not. To the extent that this information was valuable and had an impact on outcomes, this "information loss" over the business cycle could harm outcomes.

This trade-off between cleansing and information loss is important in the wide class of markets in which unobserved quality is important for outcomes and past performance is informative about quality. The mutual fund industry is such a market, and is a natural setting in which to study this trade-off because there is a broad literature exploring whether quality or ability is important for outcomes in this industry and there is clear evidence that investors in mutual funds respond to past returns.

I evaluate this trade-off by estimating a structural equilibrium model of investor and mutual fund behaviour. I estimate this model, and I use the results to run counterfactuals in which I simulate business cycles of varying types and quantify the impact of the resulting firm turnover. This allows me to draw novel conclusions about the size and persistence of business cycle shocks and derive policy implications. This paper is the first, to my knowledge, to structurally estimate the impact of cleansing and information loss over the business cycle.

The model consists of two parts. On the demand side, rational investors invest in mutual funds based on their beliefs about the heterogeneous abilities of funds to generate excess returns, and update those beliefs over time as they observe fund performance, following Berk and Green (2004). The ability of a given mutual fund to generate excess returns is decreasing in the total size of the mutual fund industry (capturing competition between funds, in the spirit of Pástor and Stambaugh (2012)) and also varies with a macroeconomic factor. The aggregate surplus generated by a fund is the total payout to the fund managers and to investors. This aggregate surplus is increasing in both fund ability and the precision of investor beliefs about fund ability: if these beliefs are imprecise, then there is misallocation (over-investment in low ability funds or under-investment in high ability funds) that harms surplus.

On the supply side, funds make dynamic decisions to exit and enter based on a macroeconomic factor and their type, including their age and their expected ability to generate excess returns. If a fund enters it incurs a fixed entry cost, sets its fee rate and randomly draws ability from the population distribution. After entry, the fund pays a fixed cost every period representing direct fixed costs and the opportunity cost of remaining in business. I allow this fixed cost to vary across the macroeconomic state and the type of fund.

I take the size of the mutual fund market as given, and instead focus on *compositional inefficiencies* regarding the types of funds that make up the market: incumbent funds do not take into account that if they exit then new funds could enter, which may improve aggregate surplus depending on their relative characteristics. A business cycle (which I model as a negative shock to the macroeconomic factor, followed by a recovery) results in exactly this exchange of funds: during the recession funds exit, which reduces competition and allows funds to enter during the subsequent recovery.

The impact of this firm turnover depends on two countervailing effects. Low ability funds are smaller and are more likely to exit during the recession, whereas the firms that replace them in a subsequent recovery are of average, and therefore higher, size and ability: in other words, the recession exchanges a large number of small funds for a smaller number of larger funds. This exchange may reduce total fixed costs incurred, depending on how fixed costs vary across fund types.[1] This is the cleansing effect. The entering funds, however, have no returns history, meaning that investors have less precise beliefs about their ability. This results in more misallocation in equilibrium, which has a negative effect on aggregate surplus. This is the information loss effect. Cleansing is about the *first moment* in ability (entrants are higher ability on average), whereas information loss is about the *second moment* (there is greater uncertainty about the ability of entrants).

The model allows me to formalise the key parameters that determine the relative strength of these two countervailing effects. The strength of the cleansing effect depends on the dispersion in the distribution of fund abilities, the differing extents to which low and high

[1]If fixed costs are homogenous across fund types, then exchanging small funds for a smaller number of large funds will always reduce total fixed costs. When fixed costs are heterogeneous across fund types, then whether this exchange reduces total fixed costs depends on the relative difference between size and fixed costs across fund types.

ability funds exit and, most importantly, the relative sensitivity of a fund's size and fixed costs to its ability. The strength of the information loss effect depends on the informational content of returns and the age of exiting funds.

I estimate both the demand-side and the supply-side of this model using data on US Equity mutual funds. I fit the demand-side to data on mutual fund size, taking into account fund returns. I do not identify the ability of funds directly, but I do identify the beliefs of investors about ability from the size of the fund: the model implies that bigger funds, all else being equal, are believed to be higher ability. I identify the value of information from the rate at which investors adjust their holdings in response to past performance: a returns history is valuable if investors are responsive to returns.

I fit the supply-side to data on fund entry and exit. I identify the fixed cost incurred by a particular type of fund by comparing its size to its empirical exit probability: for example, a fund type that is large (and therefore profitable) but exits with high probability has higher fixed costs.

The model fits well on both the demand- and supply-side. Investors are relatively fast to respond to past returns: the estimated signal-to-noise ratio implies that investors consider the informational content in their priors to be roughly equivalent to 20 months of returns data. Fixed costs vary in intuitive ways with the state and type of the fund: funds have higher fixed costs (possibly through their opportunity cost of remaining as a mutual fund) when the macroeconomic factor is good and when they are believed to be high ability. This means that high ability funds are bigger, but also have higher costs.

I use my results to undertake two sets of counterfactual simulations. First, I counterfactually simulate a single recession and recovery at a given point in time of varying depth, where a deeper business cycle results in more firm turnover. I then compare the surplus generated by the exiting funds and the entering funds at various points after the recovery.

My primary finding is that the effect of a single business cycle on aggregate surplus is negative in the short-run, positive in the medium-term, and decays to 0 in the long-term. In the short-term, the information loss effect dominates the cleansing effect: there is significant misallocation because investors have imprecise beliefs about the new funds. Over time, funds age and acquire a returns history: this is true of both the new entering funds and the exiting funds that they counterfactually replaced because of the business cycle. The benefit of this extra information is greater for the entering funds who started with no information, and so over time the information loss effect decays. After 30 months, the information loss effect has decayed to the point where it is dominated by the cleansing effect. From this point onwards, aggregate surplus is higher because of the business cycle, because it has caused better, larger funds to enter. In the long-run the effect of an initial business cycle decays to 0, because ongoing entry and exit causes firm composition to converge regardless of its starting value.

The sizes of these effects depend on the depth of the initial business cycle. The per-firm effect on surplus can be large: for the deepest business cycle I model (which is roughly equivalent to the financial crisis), the aggregate surplus of entering funds is 14% less than the aggregate surplus of the exiting funds in the first month after the recovery. The effect on aggregate surplus in the industry is material and persistent, but small (no more than 1% of aggregate surplus in the industry) as most firms neither exit not enter.

Having modelled the effect of the business cycle on firm exit and thus on outcomes, it is natural to then consider the extent to which policy should mitigate these effects. In my second set of counterfactuals, I consider which mutual funds, if any, should optimally be subsidised during a stylised, predictably temporary recession such as that resulting from the Covid-19 pandemic. I simulate subsidies targeted at particular types of mutual funds, and show that the trade-off between the information loss and cleansing effects varies across fund types in three ways. First, the information loss resulting from the exit of a young fund is relatively low, as that fund does not have an extensive returns history (there is little information to lose, in other words). Second, subsiding the largest funds has little impact because they are unlikely to exit with or without a subsidy. Third, the cleansing benefit from the exit of a medium-sized fund is relatively high, because they have disproportionately large fixed costs relative to small funds. In other words, the contribution of a particular fund to aggregate surplus is not just a function of its size, but also of its costs, and on this basis smaller funds contribute more than medium-sized funds.[2] The policy implication of these three findings collectively is that subsidies targeted at older, smaller funds have the biggest surplus benefits.

There is already an extensive macroeconomic literature on the Covid-19 pandemic (see Brodeur et al. (2020) for a summary), but I make what I believe to be a novel point in a microeconomic context: subsidies in the pandemic are intended to preserve beneficial connections through a temporary recession, and one of those benefits is the *value of information*

[2]As described above, in broad terms I identify fixed costs by comparing a fund type's size to its exit rate. Empirically, the middle size quintile of mutual funds is 18 times bigger than the first quintile, but is only around half as likely to exit. It is this empirical fact that underpins my result that medium-sized funds have disproportionately large fixed costs such that a social planner would choose to cleanse them and keep smaller funds.

that has been built up about an existing firm. On the other side of the trade-off, I show how the firms that should be cleansed and allowed to fail may not be immediately obvious: in my context, what matters is not just a firm's size and ability, but its size and ability *relative to its costs*.

I review the literature below. In Section 2, I introduce the data and set out some guiding empirical facts. In Section 3, I set out my model. In Section 4, I describe my empirical approach. In Section 5, I report my results. In Section 6, I undertake counterfactual analyses. In Section 7, I conclude.

1.1 Related literature

The main contribution of this paper is to develop an under-explored implication of business cycles: the information loss that results from firm turnover. I explore the conditions under which this information loss dominates the cleansing effect, and I quantify how this trade-off changes over time. This paper is related to three broad strands of literature.

First, this paper is related to the literature on cleansing that goes back to Schumpeter et al. (1939), and is featured more recently in Caballero and Hammour (1996) and Castillo-Martınez (2018). In this paper, I document and measure cleansing in the context of mutual funds. I also show how cleansing may bring first-moment benefits but second-moment costs in the form of information loss. This loss of information over the business cycle has not been studied extensively. Relatedly, Hale (2012) sets out reduced form evidence that recessions affect connections between firms and banks which, in a relationship banking context, could have implications for the extent of information asymmetry. Pástor and Veronesi (2009) set out a model of technological progress in which new technologies may be better than existing ones, but with greater uncertainty and therefore with greater return volatility.

Second, this paper is related to the literature on mutual funds generally (Barber et al., 2016; Berk and Green, 2004; Berk and Van Binsbergen, 2015, 2017; Fama and French, 2010; Gil-Bazo and Ruiz-Verdú, 2009; Ibert et al., 2018; Pástor and Stambaugh, 2012) and more specifically the effect of the business cycle (Gil-Bazo et al., 2020; Kosowski, 2011; Glode, 2011; Kacperczyk et al., 2014, 2016) and of fund scale (Pástor et al., 2020, 2015; Pollet and Wilson, 2008; van Binsbergen et al., 2019b) on mutual fund outcomes. There is a smaller literature that estimates structural models related to mutual funds, including Gavazza (2011), Roussanov et al. (2021) and Roussanov et al. (2020). I introduce information loss over the business cycle as a new consideration within this literature, and quantify its importance in a structural econometric context. I also contribute to the debate on how to identify and rank high ability mutual funds (Barras et al., 2010; Berk and Van Binsbergen, 2015; Jiang and Zheng, 2018; Kacperczyk and Seru, 2007) by showing that unobserved fund costs affect this ranking and can be inferred from fund exit behaviour.

Third, this paper contributes to the recent literature on economic and financial responses to Covid-19 (see Brodeur et al. (2020) for a summary). Much of this literature focuses on macroeconomic aspects of the pandemic, including Acemoglu et al. (2020), Bigio et al. (2020) and Eichenbaum et al. (2020). I consider optimal policy in a microeconomic model of a single industry, but make a broader point about preserving socially valuable information during a pandemic.

2 Data

I first describe how I select funds and calculate excess returns. I then describe the key empirical facts that motivate my research question and guide my modelling.

2.1 Sample selection

I obtain data on mutual fund characteristics and their monthly returns and assets from the database maintained by the Center for Research in Security Prices (CRSP), The University of Chicago Booth School of Business. I select data from January 1990 to December 2016. I limit my sample to actively managed US Equity funds that (i) are never smaller than USD 0.2m in size and (ii) have at least 12 months of returns data. This departs from the standard approach in the literature (see for example Berk and Van Binsbergen (2015) for an overview of mutual fund selection), in that I impose lower size and returns history thresholds for inclusion and I do not exclude funds missing data on expense ratios. Each of these is important in my context because propensity to exit is likely to be correlated with size and data availability. In other words, the standard thresholds exclude some of the funds I am seeking to study. Where a fund has multiple share classes, I combine them into a single fund observation by aggregating size and averaging fund characteristics across classes (see Nanda et al. (2009) for an analysis of the role of share classes in mutual fund outcomes). I am left with a sample of 4,446 funds and a total of 582,382 month-fund observations.

2.2 Calculating excess returns

I calculate excess returns following Berk and Van Binsbergen (2015). I regress returns in excess of the risk-free rate (R_{it}) on a set of 11 common factors ($\mathbf{F_t}$) which are the returns to the main index funds operated by Vanguard. The fund's excess return, α_{it} is the residual in this regression:

$$R_{it} = \boldsymbol{\beta}_i \mathbf{F_t} + \alpha_{it} \tag{1}$$

This is a more reasonable benchmark for mutual funds than, for example, a benchmark involving momentum investing returns that would be prohibitively costly to implement in practice. See Berk and Van Binsbergen (2015) for a fuller discussion.

2.3 Empirical facts

I set out four empirical facts:

1. **Heterogeneity in fund size**. Funds vary significantly in size at the point of entry and over their lifetime, as I show in Figure 1. This is true even controlling for the macroeconomic conditions at the time of entry: in other words, this is cross-sectional variation not just inter-temporal variation.

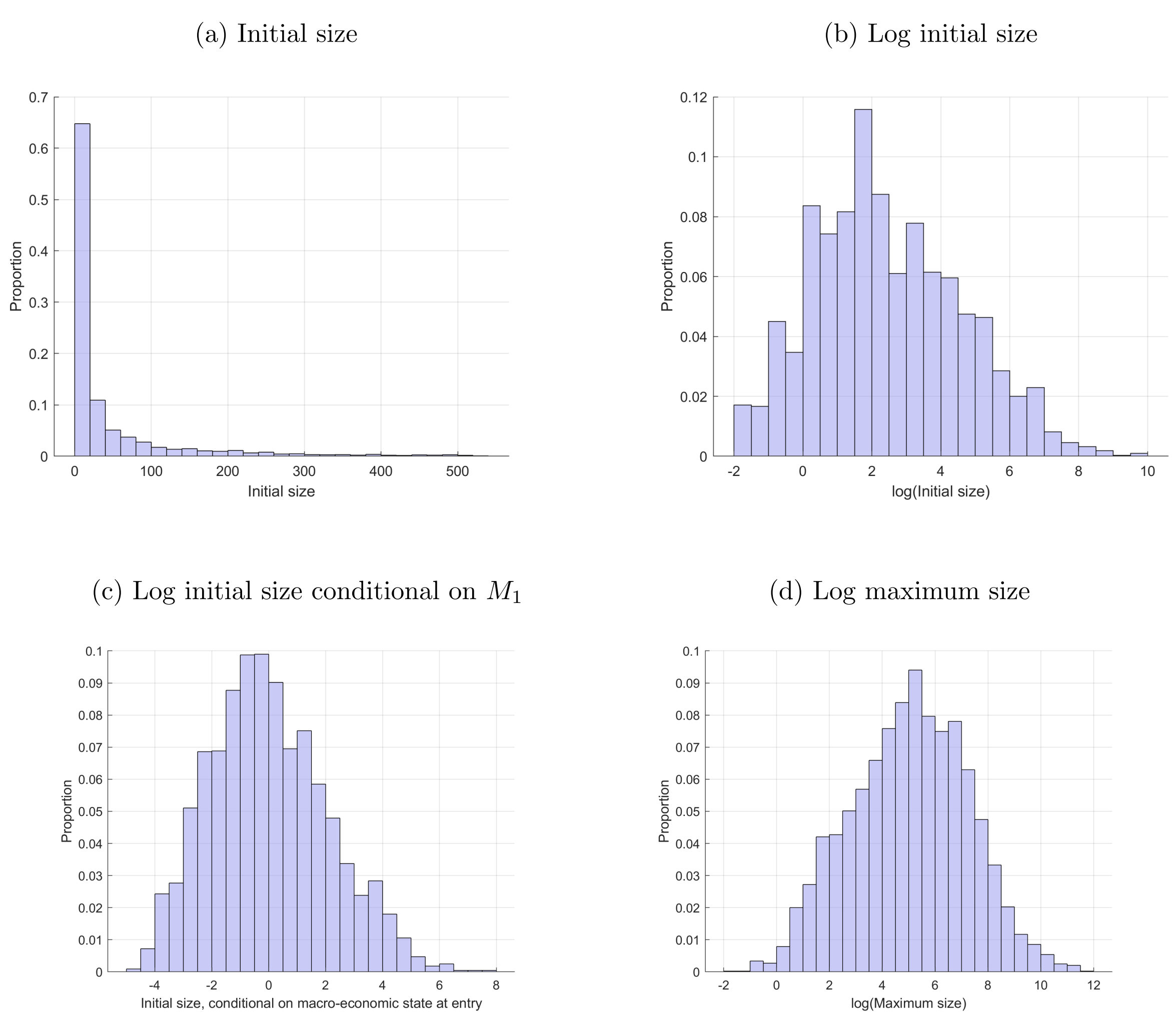
2. **Exit is correlated with size**. Smaller funds are significantly more likely to exit in any given period than bigger funds, as I show in Figure 2, where I define exit as dissolution or being merged into another fund. The relationship between size and exit is convex: for example, the 1st size quintile is 18 times smaller than the 3rd size quintile, but only slightly under twice as likely to exit in any given year.

3. **Exit is counter-cyclical**. Funds are more likely to exit when the S&P500 (which I denote macroeconomic factor M_t) is low than when it is high, as I show in Figure 3.

4. **The size of the mutual fund industry is pro-cyclical**. There is, unsurprisingly, a close relationship between the S&P500 and the aggregate size of the mutual fund industry, which I denote Q_t. I show this graphically in Figure 4 and in the regression results in Table 1. The R^2 of a regression of Q_t on M_t is 0.75, rising to 0.9 if I include a structural break in the financial crisis.

To these empirical facts I add that investors respond to past returns, on which there is a large literature (see, for example, Chevalier and Ellison (1997)). These facts combined naturally give rise to my research question: given that exiting funds are observably different from the average fund, what impact does this exit have on aggregate outcomes? Given that investors clearly attach some value to past returns, what impact does the absence of past returns have on entrants? The macroeconomic factor clearly has an impact on aggregate trends in the mutual fund industry, but what about on its composition?

Figure 1: Heterogeneity in fund size



(a) Initial size

(b) Log initial size

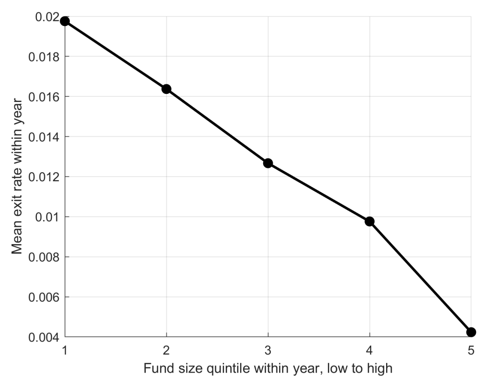
(c) Log initial size conditional on M_1

(d) Log maximum size

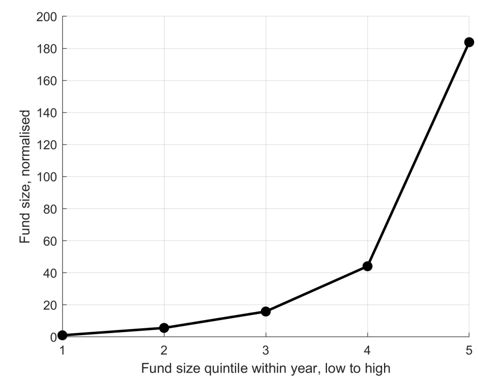
Note: Panel (a) shows the distribution of fund size in the first period of its life, excluding the top 5% of funds by size. Panel (b) shows the distribution of the natural log of initial size. Panel (c) conditions on M_1, the level of the S&P500 in the period in which the fund entered. Panel (d) shows the log of the maximum size the fund attains during my sample.

Figure 2: Variation across fund size

(a) Exit rates



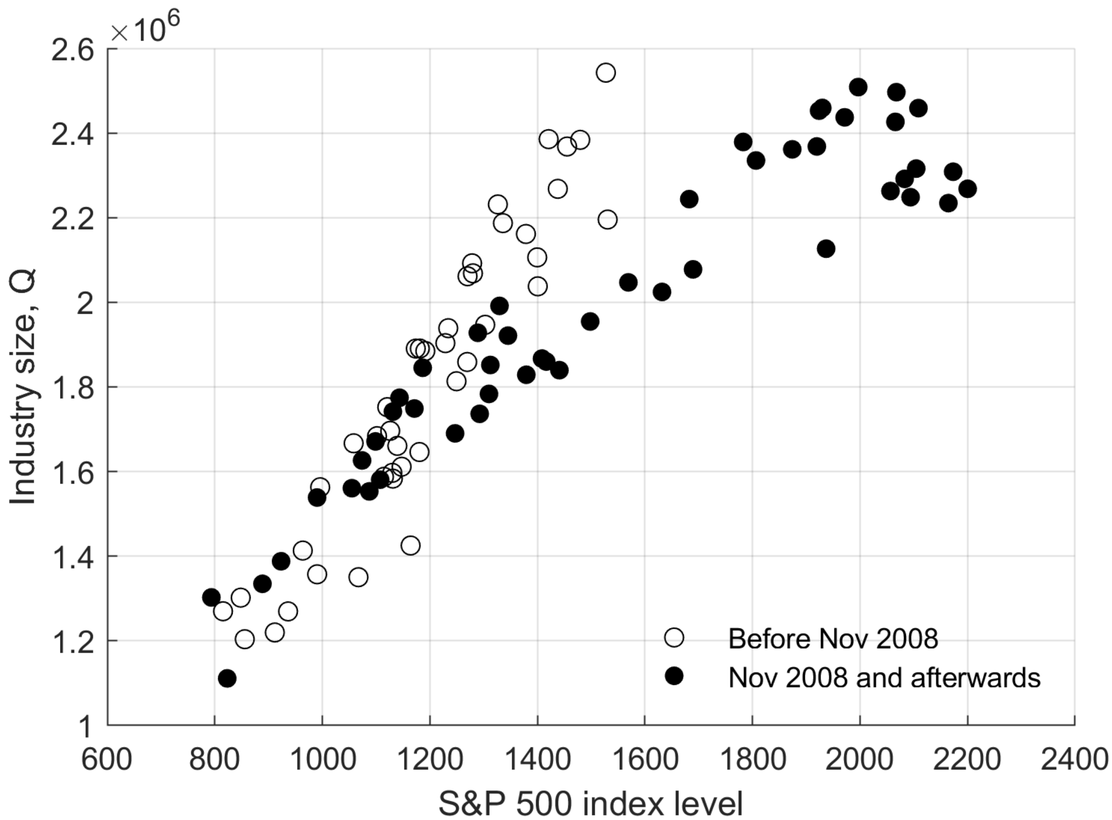
(b) Fund size



Note: Panel (a) shows that smaller funds are more likely to exit than bigger funds. Panel (b) shows that the size distribution is asymmetric: there are some very large funds. In combination the panels show the relationship between size and exit is convex.

Figure 3: Exiting funds and the S&P500

Figure 4: The relationship between Q and S&P500



3 Model

The model consists of two parts: (1) a model of demand by rational investors for mutual funds and (2) a model of supply by mutual funds. I describe each part of the model, before considering the implications of the model for aggregate surplus, efficiency and the role of the business cycle.

3.1 Demand

The model of demand is based on Berk and Green (2004), in that it shares the following two core components. First, there are *decreasing returns to scale* in the ability of funds to earn excess returns. Bigger funds, all else being equal, earn lower returns because their ability to gather and exploit information is diluted or because of price effects or execution costs. Second, *ability is unobserved*, but *investors learn* as they observe past returns. These two core components in combination mean that rational investors form beliefs about the ability of funds and invest up to the point where, given decreasing returns to scale, those returns are competed away. As investors observe past returns of the fund, they update their beliefs about the ability of the fund and adjust their holdings.

To these core components I add the following to suit my research question and to allow the model to be reasonably taken to data. First, following Pástor and Stambaugh (2012), I model *competition between mutual funds* by allowing the returns earned by funds to be decreasing in the total size of the mutual fund industry: a mutual fund earns lower excess returns, all other things being equal, if there are many other mutual funds trying to earn excess returns from the same set of investment opportunities.[3] Second, I include *a role for the business cycle* by allowing the ability of funds to earn excess returns to vary according to a macroeconomic factor that varies exogenously over time.

More formally, I follow Berk and Green (2004) and draw a distinction between the *net* excess return that investors actually earn, and the *gross* excess return the fund would have earned on a single dollar of investment (that is, before the effect of decreasing returns to scale). The total risk-adjusted payout in dollar terms to investors from investing q_{it} in mutual fund i with gross return α^g_{it} and fee rate f_i is:

$$TP_{it} = q_{it}\alpha^g_{it} - C(q_{it}) - q_{it} f_i$$

where $C(q_{it})$ is a cost function representing the decreasing returns to scale in the ability to earn excess returns. I parameterise the cost function as $C(q_{it}) = \phi_i q^2_{it}$. The parameter $\phi_i > 0$ denotes the *scalability* of the fund, where high ϕ_i indicates that returns to scale decrease quickly. The *net* α^n_i excess return is what investors actually earn, and is simply this payout divided by the size of the investment:

$$\alpha^n_{it} = \frac{TP_{t+1}}{q_{it}} = \alpha^g_{it} - \frac{C(q_{it})}{q_{it}} - f_i = \alpha^g_{it} - \phi_i q_{it} - f_i \tag{2}$$

I disaggregate the fund's gross excess return into three components. First, the fund's

[3] The effect of competition depends only the size of the industry, and so is homogenous across funds regardless of their investment style or fund family. See Wahal and Wang (2011) for an assessment of local competition that is heterogeneous across funds depending on their investment strategies. See Gavazza (2011) and Sialm and Tham (2016) for more on the role of fund families.

true ability to generate excess returns α_i, where $\alpha_i \sim N(\mu_i, \tau_{i,\alpha}^{-1})$. Second, a fund-specific iid shock ϵ_{it}, where $\epsilon_{it} \sim N(0, \tau_{i,e}^{-1})$ and $\alpha_i \perp\!\!\!\perp \epsilon_{it}$. Third, common variation in ability across funds δ_t:

$$\alpha_{it}^g = \alpha_i + \epsilon_{it} + \delta_t \tag{3}$$

I then disaggregate the common variation into a further three components: an age effect $\delta_{a(it)}$ (where I denote the age of fund i at time t as $a(it)$), the effect of macroeconomic factor M_t and the effect of industry size Q_t:

$$\delta_t = \delta_{a(it)} + \beta M_t + \theta Q_t \tag{4}$$

I estimate unrestricted age effects $\delta_{a(it)}$, macro-effects β and industry-size-effects θ in my empirical analysis, as described below. A natural interpretation at this stage, however, is that $\beta > 0$ and $\theta < 0$. $\beta > 0$ implies that funds are more able to earn gross excess returns when the macroeconomic factor is good.[4] $\theta < 0$ represents competition, in that a larger mutual fund industry means more competition for the same investment opportunities, reducing excess returns.

Investors choose q_{it} before ϵ_{it} is realised. Furthermore, investors do not know the true ability of the fund α_i, but form expectations based on the information available to them at the point of investment, which I denote I_{t-1}. I define these expectations as $e_{it} \equiv \mathbb{E}[\alpha_i \mid I_{t-1}]$. All other components of the return are known to the investor, including ϕ_i and δ_t.

Investors supply capital with infinite elasticity to any fund with positive expected *net* returns α_{it}^n, taking aggregate investment q_{it} in the fund as given. In equilibrium, q_{it} is then such that $\mathbb{E}[\alpha_{it}^n \mid I_{t-1}] = 0$. Substituting in Equations 2 and 4, this means that:

$$q_{it} = \frac{e_{it} + \delta_t - f_i}{\phi_i} \tag{5}$$

Investor demand for mutual fund i is therefore increasing in its expected ability e_{it}, increasing in its scalability ϕ_i, decreasing in its fee rate f_i and subject to common variation δ_t. Note that for ease of reference I refer to e_{it} as "ability" and ϕ_i as "scalability", but in some sense both are fund-specific measures of the ability to generate excess returns on q_{it}.

[4] $\beta > 0$ implies that unobservable *gross* excess returns and fund size are positively correlated with the macroeconomic factor, but does not imply that observable *net* excess returns are correlated with the macroeconomic factor. See Kacperczyk et al. (2016) for more on how net excess returns vary over the business cycle.

To complete the model of demand, I need to characterise the expectations formation process behind e_{it}. Investors observe past net excess returns, $\alpha^n_{is<t}$ and from this can infer gross returns α^g_{is}. Investors cannot separately identify α_i from ϵ_{is}, but can extract a signal about α_i given their relative distributions.

Given these distributional assumptions, there are simple closed-form expressions for how investors form and update their posterior beliefs about α_i in responses to these signals. Defining the signal-to-noise ratio $\lambda = \frac{\tau_e}{\tau_\alpha}$ and $s(\lambda, t) = 1 + (t-1)\lambda$:

$$q_{it} = \frac{1}{\phi_i}\left[\delta_t - f_i + \frac{\mu_i}{s(\lambda,t)} + \frac{\lambda}{s(\lambda,t)}\sum_{m=1}^{t-1}\alpha^g_{im}\right] \tag{6}$$

I leave implicit the lower bound of zero. I repeatedly substitute in Equation 2 to solve forward for optimal q_{it} in terms of *net* returns (which are observed by the econometrician), instead of *gross* returns (which are not directly observed by the econometrician):

$$q_{it} = \frac{1}{\phi_i}\left[\mu_i - f_i + \delta_t + \lambda\sum_{m=1}^{t-1}\frac{\alpha^n_{im} - f_i}{s(\lambda, m+1)}\right] + e^q_{it} \tag{7}$$

I add an error term, e^q_{it}, that represents shocks to q_{it} beyond this expectations formation process. This could include, for example, noise traders. I leave further discussion of this error term and its distribution to the section below on my empirical analysis. This Equation 7 characterises equilibrium investor demand for fund i. In what follows I define the "observable type" of mutual fund i as $\mathbf{\Theta_i} = (\mu_i, \phi_i, \sigma^a_i, \sigma^e_i, f_i)$ and its "unobservable type" as α_i.

3.2 Supply

On the supply-side, firms make three decisions: (1) they choose to enter or not to enter, (2) they set a single fee at the start of their life and (3) they choose to exit or not to exit. Before modelling these three choices, I describe firm beliefs about the evolution of industry size, which will be important for each choice.

3.2.1 Firm beliefs about industry size

The payoff to a mutual fund depends on macroeconomic factor M_t and competition through the size of the mutual fund industry Q_t, as I set out in equation 4. I set out in Figure 4 and in Table 1 how closely M_t and Q_t co-move, with a R^2 value of 0.75 in a linear regression of Q_t on M_t.

I assume that M_t follows an AR(1) process, with stochastic error $e^M \sim N(0, \sigma^M)$:

$$M_t = \rho^M M_{t-1} + e_t^M$$

The key assumption on the supply-side is that funds take aggregate industry size Q_t as given, and form beliefs about its dynamics based on its linear co-movement with M_t:

$$Q_t = g(M_t) = g_1 + g_2\, M_t \tag{8}$$

I assume the relationship is linear for simplicity and because of the empirical relationship set out in Figure 4. This means that Q_t also follows an AR(1), such that firms have expectations about how industry size will develop over time, regardless of their own decisions or those of their competitors. This assumption has obvious computational benefits: the modelling environment is not a game, but a series of individual decisions by each mutual fund. The remaining complication, which I consider below, is ensuring that the individual decisions result in aggregate dynamics that are consistent with the firm beliefs set out in Equation 8.

I argue that this assumption is reasonable given that there are a very large number of funds, the significant majority of which are a very small proportion of total Q_t. There are admittedly a small number of very large mutual funds for which this assumption may not be reasonable: these, however, are mostly established, older funds that are very unlikely to exit. That is, this is a reasonable assumption to make when studying, as I am, the entry and exit of mutual funds.

3.2.2 Exit

In each period, a mutual fund earns fees based on the size of the fund, where all marginal costs are subsumed into the fee rate they charge. Each period the fund manager incurs a fixed cost, denoted W, representing fixed expenditures and the opportunity cost of remaining in business.[5] The mutual fund can choose to exit and obtain an outside option, the value of which I set to 0.

[5]See Deuskar et al. (2011) for more on mutual fund exit, including the number and type of managers that leave to work in hedge funds.

This decision is dynamic, and depends on the type of the mutual fund and the state, including investor beliefs about the mutual fund e_{it}, the age of the mutual fund a_{it} and the macroeconomic factor M_t:

- Type: $\mathbf{\Theta_i} = (\mu_i, \phi_i, \sigma_i^a, \sigma_i^e, f_i)$
- State: $\mathbf{S_t} = (e_{it}, a_{it}, M_t)$
- Action: $z_{it} = 0$ if exit, $z_{it} = 1$ if do not exit.

Firms take expectations over the development of beliefs about their ability e_{it} and changes in M_t. Their age, and with it the precision of investor beliefs about their ability, updates deterministically. As is standard in the literature (see for example Hotz and Miller (1993)) funds receive an action-specific shock $\eta(z_{it})$ that is distributed Type-1 extreme value. In recursive Bellman form:

$$V_{it}(\mathbf{S_t}; \mathbf{\Theta_i}) = \max_{z_{it}} \; z_{it} f_i q_{it}(\mathbf{S_t}; \mathbf{\Theta_i}) - z_{it} W(\mathbf{S_t}; \mathbf{\Theta_i}) + \eta(z_{it}) + z_{it} \beta \mathbb{E}[V_{it+1}(\mathbf{S_{t+1}}; \mathbf{\Theta_i})] \tag{9}$$

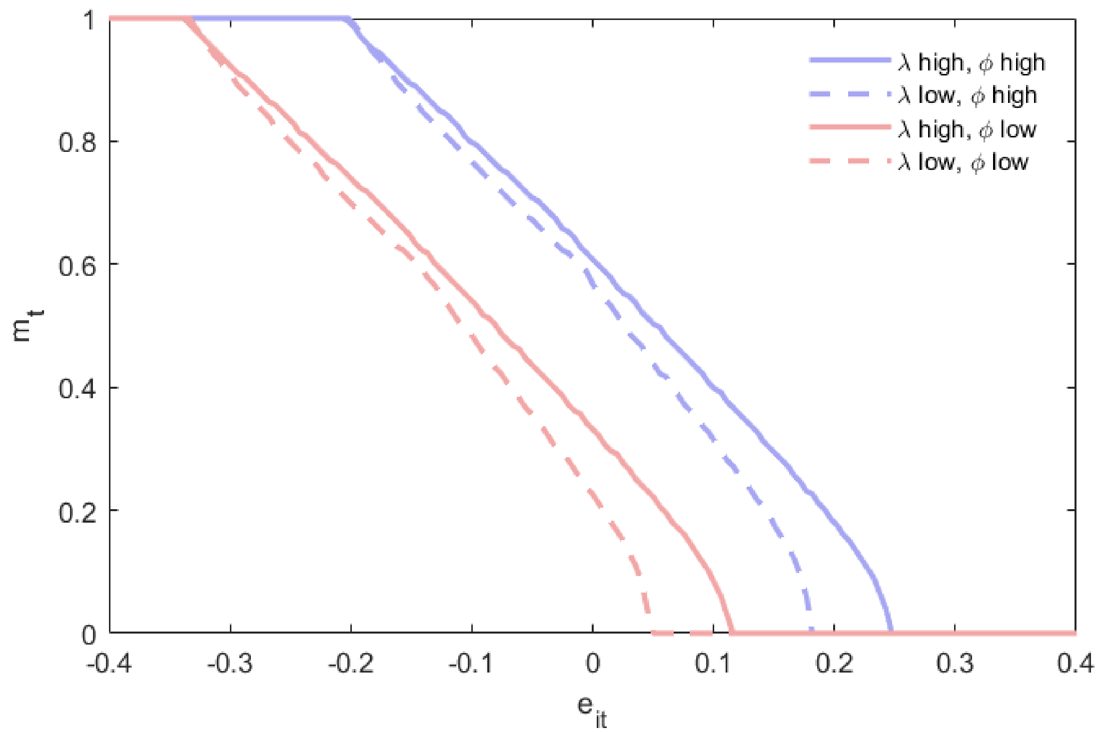
I allow the fixed cost $W(\mathbf{S_t}; \mathbf{\Theta_i})$ to vary with the state and type of the fund, on the basis that the opportunity cost of remaining in business may be greater for better funds and in better macroeconomic states. It also affords me additional flexibility to fit observed empirical exit probabilities. To illustrate the equilibrium exit decisions resulting from this model, I solve for $z^*(\mathbf{S_t}; \mathbf{\Theta_i})$ under various combinations of ability beliefs e_{it}, macroeconomic states m_t and parameter values in Figure 5 below, ignoring the action-specific shock $\eta(z_{it})$ and with homogenous fixed cost W. These numerical results indicate a cutoff rule: funds exit when they are perceived to be bad or when the macroeconomic state is bad, or some convex combination thereof.

I abstract away from complications relating to exit, such as the decision to liquidate or merger with another fund. It would in principle be straightforward to add an exogenous probability (possibly varying with state or type) with which the assets of the exiting fund are merged with another fund rather than liquidated.

3.2.3 Entry

Firms decide to enter and set fees without knowing their observable type $\mathbf{\Theta_i}$, but knowing the type distribution $h_{\mathbf{\Theta_i}}$. Firms choose to enter if the expected value of entry, taking

Figure 5: Exit decisions



Note: The area under the curve shows the combinations of ability belief (e_{it}) and business cycle state (m_t) in which a fund exits: funds exit when they are perceived to be bad or when the macroeconomic state is bad, or some convex combination thereof. Funds are less likely to exit when returns are less informative (λ is low) and/or when their ability scales up easily (ϕ is low). For ease of exposition this figure ignores action-specific shocks and assumes constant fixed cost across types and states.

---

expectations over $\mathbf{\Theta_i}$, exceeds an entry cost that is common across firms but can vary over time:

$$f_t^* = \arg\max_{f_i} \int V_{it}(\mathbf{S_t}; \mathbf{\Theta_i}) \tag{10}$$

$$\int V_{it}(\mathbf{S_t}; \mathbf{\Theta_i})\, d\Theta_i \geq F_t^{entry} \tag{11}$$

This produces inter-temporal variation in the fee rate over time, in that funds charge lower fees at entry when the macroeconomic factor is worse. Fees are homogenous across funds that enter at the same time. This means that there is no relationship between fund ability and market power, such as there would be if, for example, funds chose their fee rate after learning their type. Such a relationship would provide an additional channel for the cleansing effect to work, in that better funds would earn higher profits. I justify this assumption on the basis that empirical fund fees, whilst not homogenous, do not appear to vary systematically.

3.3 Equilibrium

In equilibrium, (1) investors invest in any fund with positive expected excess returns, as per Equation 7; (2) mutual funds choose to enter, set fees and exit optimally, given their type, the state, investor behaviour and their beliefs about future competition, as per Equations 17, 11 and 10; and (3) mutual fund beliefs about the dynamics of future competition are consistent with optimal mutual fund behaviour.

Expanding on the third of these equilibrium requirements: the entry and exit rules, conditional on $g(M_t)$, induce dynamics in Q_t, which I denote $h(M_t, g(M_t))$. Equilibrium is a fixed point such that $g^*(M_t) = h(M_t, g^*(M_t))$. In other words, in equilibrium the entry and exit rules induce fund behaviour that is consistent with the overall dynamics in Q_t given M_t.

I do not solve for this equilibrium function. Instead, in the empirical analysis below, I observe and estimate this equilibrium function and hold it constant in the counterfactuals I run. This clearly places restrictions on the counterfactuals in which this equilibrium function could plausibly be held constant, which I discuss below.

3.4 Aggregate surplus

I define the surplus (or value-added, following Berk and Van Binsbergen (2015) and van Binsbergen et al. (2019a)) generated by a given fund i as the dollar return to fund and investors, less the fixed cost W_{it}:

$$s_{it} = f_i q_{it} + \alpha_{it}^n q_{it} - W_{it} \tag{12}$$

Aggregate surplus is then the sum of individual fund surplus: $AS_t = \sum_i s_{it}$. The model I set out above has two important implications for how s_{it} varies across funds.

First, the surplus generated by a given mutual fund industry of size Q_t is increasing in the ability of its constituent funds. This follows simply from the presence of fixed costs: the model of demand above implies that better funds are larger, meaning that an industry of given size is made up of fewer funds and incurs fewer fixed costs in total. There are no cleansing effects, holding industry size constant, via fees (which do not vary with ability) or excess returns (which are competed away in expectation).

Second, conditional on true α_i, the surplus generated by a given fund is typically increasing in the precision of investor beliefs. The general intuition for this is straightforward: investor beliefs are correct in expectation, but in particular realisations investors can think a particular fund is good when it is bad, and vice versa. This uncertainty results in mis-allocation (investing too much in bad funds or too little in good funds) that harms surplus. More formally, substitute Equations 2 and 5 into Equation 12 for surplus and assume for ease of exposition that $f_i = W_{it} = \delta_t = 0$ and $\phi = 1$:

$$s_{it} = (\alpha^g_{it} - e_{it})e_{it} \tag{13}$$

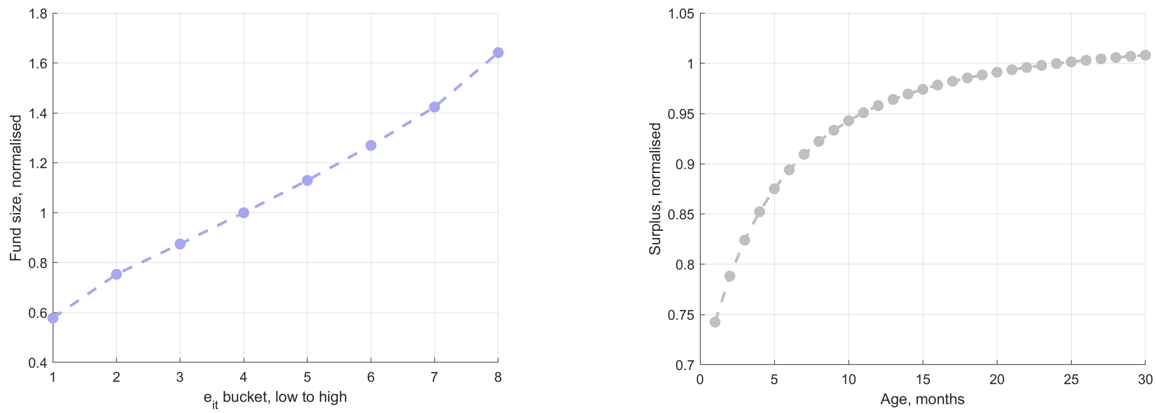
Surplus is, under these assumptions, a function only of the net return to investors. This is positive if $\alpha^g_{it} - e_{it} > 0$, meaning that the fund is better than investors believe, such that they under-invest and some net excess return is not competed away. Conversely it is negative if investors believe that the fund is better than it is, and so over-invest. Let $e_{it} = \alpha_i + \epsilon^e_{it}$, where ϵ^e_{it} denotes the error in investor beliefs for fund i, with mean zero and variance σ^e_{it}. Substitute in Equation 6 for α^g_{it} and take expectations conditional on unknown true ability α_i, and it follows that s_{it} is decreasing in the variance of this error: mis-allocation is harmful to surplus.

The primary determinant of the precision of investor beliefs is the age of the mutual fund. Older mutual funds have a returns history that is a signal of their ability, and so allows investors to form more precise beliefs. As the age of a fund goes to infinity, the error term ϵ^e_{it} and its variance σ^e_{it} go to zero. In Figure 6, I set out an example of how the surplus of a given fund is typically increasing in its age.

Additional information as a fund ages *typically* increases surplus. Whether this is always the case depends on the age of the fund and the fee rate f_i: when f_i is set too low or too

Figure 6: The effect of age and ability on surplus

(a) The effect of ability on surplus (b) The effect of age on surplus



Note: Panel (a) shows how the size of a mutual fund varies with investor beliefs about its ability e_{it}. Panel (b) shows how surplus is typically increasing in age, because as funds age investor beliefs become more precise as they observe returns.

---

high relative to the fund's true ability, then this introduces a distortion. This distortion can interact with the effect of aging in a way that means that, beyond a certain age, surplus is no longer increasing in age. I discuss this in more detail in Figure 16 in the appendix. For the purposes of my research question, it suffices to say that funds with no returns history typically have lower surplus than those that have a returns history, all other things being equal.

3.4.1 Efficiency

I consider the choices of a social planner without additional information: that is, the social planner does not know true fund ability α_i or have any more information than investors or funds. I am interested in the *composition* of the mutual fund industry, rather than its size. That is, I consider the second-best problem of optimising the composition of mutual funds, whilst taking as given the aggregate size of the mutual fund industry Q_t and its dynamics.

I illustrate a *compositional inefficiency* by considering the mutual fund industry in equilibrium, in which no incumbent fund wishes to exit and and no potential entrant wishes to enter, given the prevailing macroeconomic state M_t and the size of the industry Q_t. Consider a simple example with two types of fund, type 1 and type 2, in equal numbers and of homogenous age A, where $q_1 > q_2$, $f_1 = f_2$ and $W_1 > W_2$. To focus, for now, only on the cleansing effect, suppose that returns have no informational value ($\lambda = 0$, such that the information loss channel is shut down). I ask whether the social planner would be willing to swap n funds of type 2 for a single randomly drawn entrant of expected size $\bar{q} = \frac{q_1+q_2}{2}$ and fixed cost $\bar{W} = \frac{W_1+W_2}{2}$. To focus on composition only, suppose that n, q_1 and q_2 are such that the overall mutual fund industry size Q_t does not change. That is, $nq_2 = \bar{q}$. These assumptions mean that the only part of surplus that is affected by such an exchange is aggregate fixed costs. The exiting funds incur fixed costs nW_2 and in expectation the entering funds incur $\bar{W}$, such that the social planner would make this exchange if $q_2/W_2 < \bar{q}/\bar{W}$. This exchange need not occur in a decentralised market, in which type 2 funds have no incentive to exit so that the funds that replace them can earn greater profits.

The ratio of size to fixed cost determines the value of such a swap. Importantly, the fund type with the highest such ratio need not be the largest or highest ability. If fixed costs are homogenous then $W_2 = \bar{W}$ and the social planner would seek to preserve the larger funds. If, however, fixed costs are heterogeneous, then the smaller, lower ability fund may be the fund type that would be preserved by the social planner.

Now suppose that returns are informative. In this case, then the exchange of funds being considered affects fixed costs, as in the preceding example, but also the expected returns to investors. The exiting funds have age A whereas the entering funds have age 1, meaning that investors have more precise beliefs about the exiting funds because of their returns history. As set out in Figure 6, this means that expected surplus is greater for older funds, all other things being equal. The social planners' decision about swapping fund types therefore depends on the trade-off between ability and information.

3.4.2 The role of the business cycle

I discuss above whether the social planner would be willing to swap certain funds for an average entrant. This type of swap is exactly what results from a business cycle in my model, where I define a business cycle as a recession followed by an offsetting recovery. As I set out in Figure 5, firms are more likely to exit if they are small, if they have high fixed cost and if the macroeconomic factor is bad. The recession therefore has a cleansing effect (in that bad funds are more likely to exit) and an information loss effect (in that they are replaced by funds with no returns history).

The model allows me to formalise exactly which parameters determine the relative sizes of these two countervailing effects. The strength of the cleansing effect depends on (i) the dispersion of the distribution in abilities and (ii) the extent to which exit rates are greater for low ability funds than for high ability funds. The strength of the information loss effect depends on the age of exiting funds and the value of the information contained in past returns, as measured by the signal-to-noise ratio λ.

If, for example, returns are not particularly informative about ability, fund abilities are highly dispersed and low ability funds are significantly more likely to exit, then the cleansing effect is more likely to dominate the information loss effect and the business cycle has a positive effect on aggregate surplus. If instead returns are extremely informative, then the reverse is true. The model, therefore, cannot provide a general answer about the effect of the business cycle on outcomes: it is an empirical question.

4 Empirical approach

There are three aspects to my empirical approach: (1) I estimate some exogenous processes that are outside the model, (2) I calibrate some parameters and (3) I estimate the remaining parameters by matching observed quantities, entry and exit decisions. I discuss each of these in turn, before considering identification.

4.1 Exogenous processes

I model two exogenous processes outside the model. The first is the dynamics of macroeconomic factor M_t, which in my empirical analysis is the S&P500 index. As described above, I assume that M_t follows an AR(1) process, which in estimation I augment with a time trend:

$$M_t = \rho^M M_{t-1} + \tau^M t + e_t^M \tag{14}$$

where $e_t^M \sim N(0, \sigma^M)$. I estimate this equation and recover an estimate of σ^M from the time-series of M_t. In my counterfactual simulations I ignore the time trend component.

The second exogenous process is the relationship between Q_t and M_t, as set out in Equation 8. I use the results set out in column 3 of Table 1, which is a simple linear regression of Q_t on M_t.

4.2 Calibration

On the supply-side, I set the discount factor to 0.99. On the demand-side, all of the parameters in Equation 7 are separately identifiable, including fund-specific scalability ϕ_i and prior beliefs μ_i. In practice, to keep the number of parameters to be estimated down, I calibrate ϕ_i and μ_i based on how q_{it} evolves over time.

I set scalability parameter ϕ_i to be the inverse of the maximum size that fund i reaches in my sample: $\phi = \frac{1}{q_{i,max}}$, where $q_{i,max} = \max_t q_{it}$. This is effectively a fund-specific normalisation such that the product $q_{it}\phi_i \in [0, 1]$ for any i. This means that I do not use the cross-sectional variation in the size of the funds to identify the other parameters, but only the variation over time. In other words, I assume that Vanguard's largest funds are not large relative to other funds because they earned very large returns early in their life, they are large for fund-specific reasons that I effectively encode and leave fixed in ϕ_i.

I infer prior belief μ_i from the size of fund i in the first period of its life. Setting $t = 1$ in Equation 7 and re-arranging: $\mu_i = q_{i1} - \delta_{i1}$. This results in computational benefits, relative to simply estimating μ_i as a fixed effect, as it can be done outside of the main estimation loop. It also better matches the interpretation of μ_i as an initial prior belief about fund ability at the start of its life.

Implementing these calibrations in Equation 7 for demand, re-arranging and defining the within-style transformation $\tilde{\delta}_t = \delta_t - \delta_{i1}$:

$$\frac{q_{it} - q_{i1}}{q_{i,max}} = \tilde{\delta}_{it} + \lambda \sum_{m=1}^{t-1} \frac{\alpha^n_{im} - f_i}{s(\lambda, m+1)} + e^q_{it} \tag{15}$$

4.3 Estimation

I estimate the demand-side and the supply-side separately for tractability. From the supply-side I need to estimate the entry cost F^{entry}_t, the fixed costs $W(\mathbf{S_t}; \mathbf{\Theta_i})$ and the variance of exit shocks σ^η. I estimate all remaining parameters from the demand side.

On the demand-side, I run non-linear least squares on Equation 15, where the only non-linear parameter is λ, which governs the responsiveness of fund size to past returns. I choose 2 months as the unit of time. I winsorize excess returns data at 1% and 99% across my sample so as to avoid (potentially unreliable) extreme values. The specification is sensitive to the units of excess returns chosen: I divide all excess returns by the largest single excess return in my sample, thus normalising excess returns to be weakly less than 1. In some cases funds may experience one-off shocks to their size because of, for example, mergers. I account for this in estimation by including a time- and fund-specific dummy variable wherever a fund that is more than 2 years old grows by more than 100% within a 2 month period. This happened 520 times within my sample of 291,191 month-period pairs.

Given estimates of the parameters in Equation 15, it is then straightforward to infer estimates of unobserved gross α_{it}^{g} from Equation 2, and from that $\sigma_{ie}^{2} = var(\alpha_{it}^{g})$ and $\sigma_{i\alpha} = \lambda\sigma_{ie}$.

On the supply-side, I undertake a nested-fixed point estimation in which I match observed probabilities of exit with model-implied probabilities. I discretise the state-space into 8 buckets for expected ability e_{it}, 8 buckets for macroeconomic factor M_t and 3 buckets for the fund's age. I do this for 2 types of scalability ϕ_i, meaning I have a total of 384 state-type combinations. The estimates of the demand-side and the first-stage estimates relating to the evolution of M_t allow me to model transition probabilities between each of these buckets. I show in Figure 13 in the appendix the exit rules implied by this coarser state space: it matches the key characteristics of the exit rules implied by the finer state space in Figure 17. For each state-type bucket, I calculate the observed annual exit probabilities over 8 years between 2004 and 2011, $\hat{Pr}(z = 1 \mid \mathbf{S_t}; \mathbf{\Theta_i})$. To calculate model-implied probabilities, I first set out the following mean choice-specific utilities, averaging across funds in the same state-type buckets:

$$v_{1t}(z = 1, \mathbf{S_t}; \mathbf{\Theta_i}) = f_i q_{it}(\mathbf{S_t}; \mathbf{\Theta_i}) - W(\mathbf{S_t}; \mathbf{\Theta_i}) + \beta\mathbb{E}[V_{it+1}(\mathbf{S_{t+1}}; \mathbf{\Theta_i})]$$
$$v_{0t}(z = 0, \mathbf{S_t}; \mathbf{\Theta_i}) = 0$$

Given the assumed distribution of exit shock $\eta(z_{it})$, the probability of exit is then a function of these mean utilities:

$$Pr(z = 1 \mid \mathbf{S_t}; \mathbf{\Theta_i}) = \frac{1}{exp(v_{1t}(\mathbf{S_t}; \mathbf{\Theta_i})) + 1}$$

I parameterise fixed costs in order to then simulate over counterfactual combinations of state and type that did not occur during my sample. I find that the following specification of $\sqrt{W_{it}}$ (so as to ensure it predicts weakly positive fixed costs) on a quadratic specification in fund type and state (comprising expected ability e_{it}, scalability ϕ_i, age and macroeconomic factor M_t) works well:

$$\sqrt{W_{it}} = w_0 + w_1 e_{it} + w_2 age_{it} + w_3 \phi_i + w_4 M_t + w_5 e_{it}^2 + w_6 age_{it}^2 + w_7 \phi_i^2 + w_8 M_t^2 + w_9 \phi_i e_{it} + w_{10} age_{it} e_{it} + w_{11} age_{it} \phi_i + w_{12} t + w_{13} t^2 \tag{16}$$

I choose the parameters in this model, along with the standard deviation of the exit shock η, to minimise the distance between model-implied exit rates by state-type bucket and empirical exit rates by nested fixed point estimation. I exclude any state-type buckets for which I have less than 10 observations from my estimation, leaving me with 291 empirical probabilities to fit. The model cannot easily produce an exit probability of 0, so I replace any empirical exit probability of 0 with the smallest strictly positive exit probability in my sample, which is 0.0028.

4.4 Identification

The primary challenge in identification of the demand-side is the role of unobserved shocks to mutual fund size. In the context of the model, the error term e_{it}^q represents investment in the fund that is unrelated to beliefs of investors about the ability of the fund: noise traders, in other words. Correlations in noise trading across funds and across time create challenges in identification in two ways.

First, aggregate industry size Q_t is endogenous in the presence of unobserved shocks that are common across funds. If, for example, the mutual fund industry is popular with noise traders in time t, then both q and Q would be large: this would likely bias our estimate of the effect of Q on q away from zero. Second, α_{it-1}^n is a function of q_{it-1} and so of e_{it-1}^q: this means that historical returns are endogenous in the presence of serially correlated unobserved noise trading. If, for example, firm i is popular among noise traders for two consecutive periods, then returns are low and the fund is big: this would bias our estimate of the responsiveness of rational investors to past returns λ downwards. This should not, however, happen if investors are rational, as they would adjust their investment to account for any predictable inter-temporal variation in noise trading.

Nevertheless, to mitigate the second of these concerns I implement a two-stage estimation process. In the first stage, I estimate the demand equation including a time dummy, δ_t. This insulates my estimation of λ from unidentified common shocks that are correlated across time, such as noise traders. In the second stage, I regress these estimated time dummies on the common, time-varying components of demand, Q_t and M_t. Separately identifying these effects using only time series variation is challenging, in the absence of some exogenous shock to fund size. I include a time trend and I instrument for Q_t using the number of active mutual funds, on the basis that this is likely to be less sensitive to unobserved contemporaneous shocks. Importantly, in my simulations it is not necessary to separately identify the effect of Q_t from M_t, as all that matters is their net effect. I estimate them separately on the demand-side only as support for my implicit assumption that mutual funds compete.

On the supply-side, the primary challenge is the identification of the role of M_t on exit, via its effect on fixed costs $W(\mathbf{S_t}; \mathbf{\Theta_i})$. It is not straightforward to separate the role of the business cycle from other things that changed over the course of my sample period, such as the general trend towards passive investment. I include a time trend and a squared time trend to account for such general trends. I also limit my estimation window on the supply side to 2004 to 2011 so as to minimise the effect of unobserved longer-term changes in the industry, such as the trend away from active and towards passive investment management.

5 Results

I set out the results of my estimation in Tables 2 to 4 and Figures 7 to 15. Model fit is good on both the demand-side (R^2 of 0.73) and supply-side (0.69). Good model fit helps ensure that the beliefs of mutual funds are consistent: firms believe that the size of the aggregate industry Q_t develops linearly with the macroeconomic factor M_t, as it does empirically. The fact that the model fits well means that the model predicts optimal behaviour that will then result in such a linear relationship in aggregate.

On the demand-side, the results have the following implications:

1. **The role of past returns**: I estimate λ to be 0.099. This means that investors respond to returns relatively quickly. It implies, for example, that the investor's priors about a fund are as important to the investor as 20 months of returns history.

2. **The role of competition**. The coefficient on industry size Q_t is negative and significant, indicating that competition between mutual funds plays a role. Furthermore, this parameter estimate is sensitive to the use of an IV approach in the direction one would expect: failing to control for common shocks understates the importance of competition.

3. **The role of the business cycle**. The coefficient on macroeconomic M_t is positive and significant: funds are larger when the macroeconomic factor is good. As well as this direct effect, M_t has an indirect effect on q_{it} via Q_t. The net effect of M_t, taking into account both the direct and indirect effect, is positive: when the macroeconomic factor is good, Q_t is higher (which has a negative impact on q_{it} because of the impact of competition), but not to the extent that it dominates the direct effect.

On the supply-side, the key implications of the results are as follows:

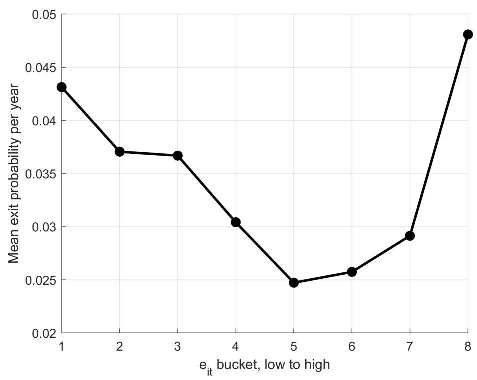
1. **Variation in exit rates**. Based on the results from the demand-side, I allocate each fund to the state-type buckets described above, and calculate the exit rates in those buckets. As set out in Figure 7, I find that funds are more likely to exit when my model indicates that they are low expected ability (e_{it} is low) or do not scale up well (ϕ_i is high).

2. **Variation in fixed cost**. I estimate state-type specific fixed costs, and show their estimated distribution in Figure 17. In Figure 8 I show that these fixed costs vary across types and states in an intuitive way. Fixed cost co-moves closely with the expected ability of the fund e_{it} and with its scalability ϕ_i: funds have greater opportunity costs if they are higher ability and/ or are able to scale that ability up easily.

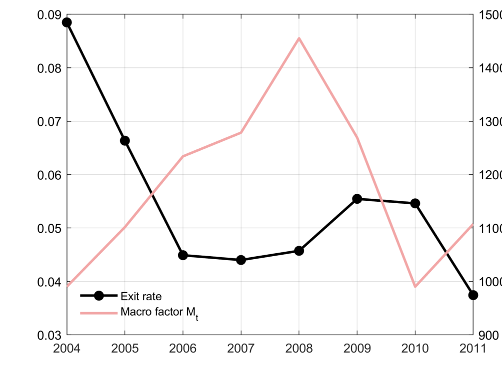
3. **Variation in entry cost**. The expected value of entering is greater when the macroeconomic factor M_t is good. Given that I assume that new entrants are always indifferent between entering or not, this means that the fixed cost of entry F_t^{entry} is also increasing in M_t, as I set out in Figure 15.

Figure 7: Observed firm exit by state and type

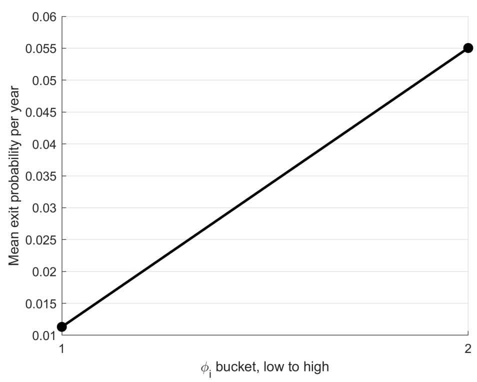
(a) Ability, e_{it}



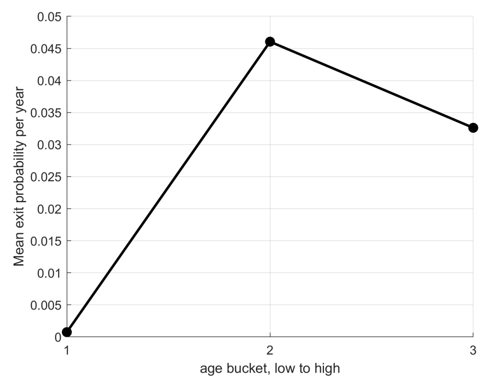
(b) Macroeconomic factor, M_t



(c) Fund scalability, ϕ_i

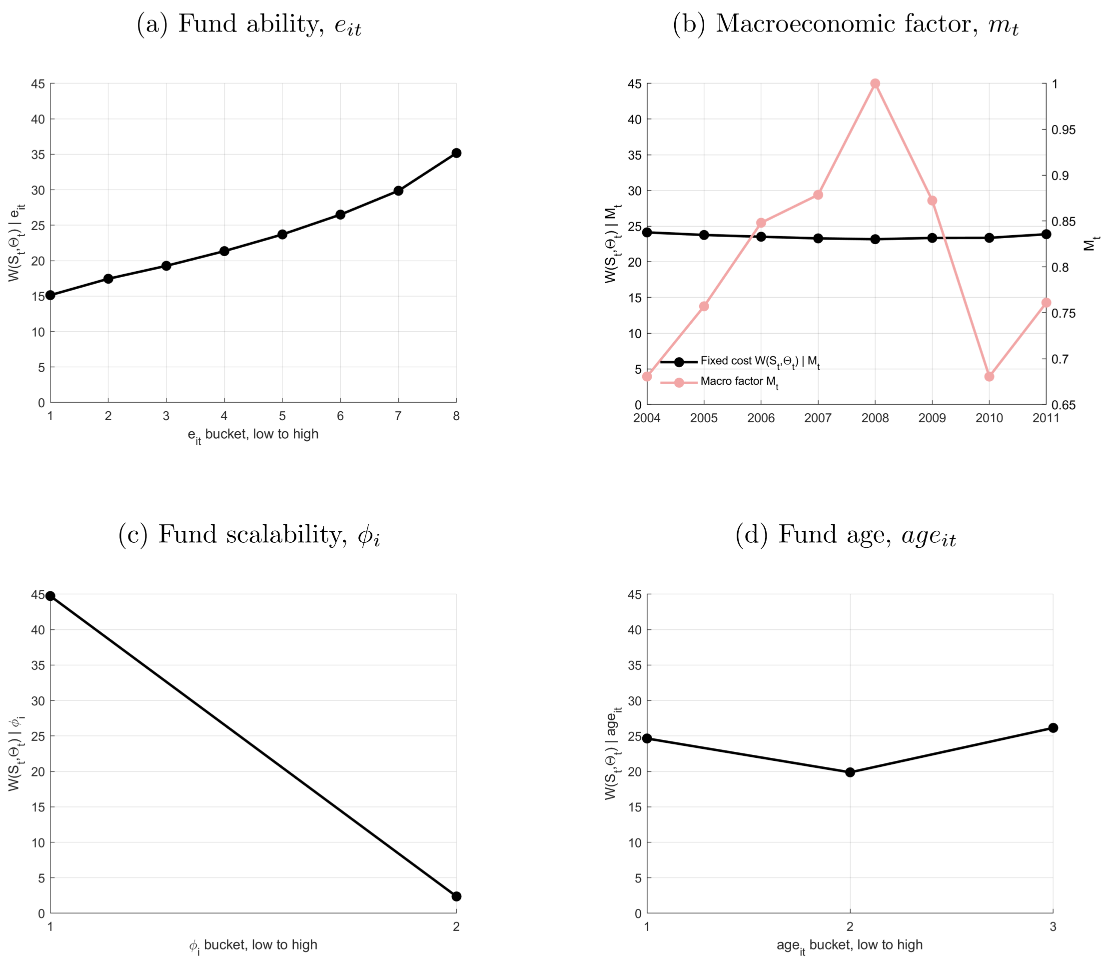


(d) Age, age_{it}



Note: I calculate the observed probability of exit in a year for each state-type bucket. In this figure I show how these observed probabilities vary on average with states and types. Note that the ability of a fund to scale up in size is decreasing in ϕ_i.

Figure 8: State-type-specific fixed costs



Note: I estimate fixed costs that vary according to state and type. In this figure I show how these fixed costs vary according to the state and type of the fund. Funds have higher fixed costs when they have higher expected ability (panel (a)) and when they scale well (panel (c), where a fund that scales well has low ϕ_i). which may represent higher opportunity costs. There is also minor variation in fixed costs with the macroeconomic factor (panel (b)) and age (panel (d)).

6 Counterfactual analysis

I run two sets of counterfactuals. First, I simulate business cycles of various depths to model their impact. Second, I consider optimal subsidies during a pandemic.

6.1 Business cycle depth

I am interested in the effect of the *depth* of the business cycle on outcomes *post-recovery*. To assess this, I simulate a single business cycle (that is, a recession, followed by a recovery) in the macroeconomic factor M_t of varying depths, model the resulting counterfactual firm turnover, and then set out the effect on aggregate surplus. I describe my approach to modelling this counterfactual in detail in Appendix C.

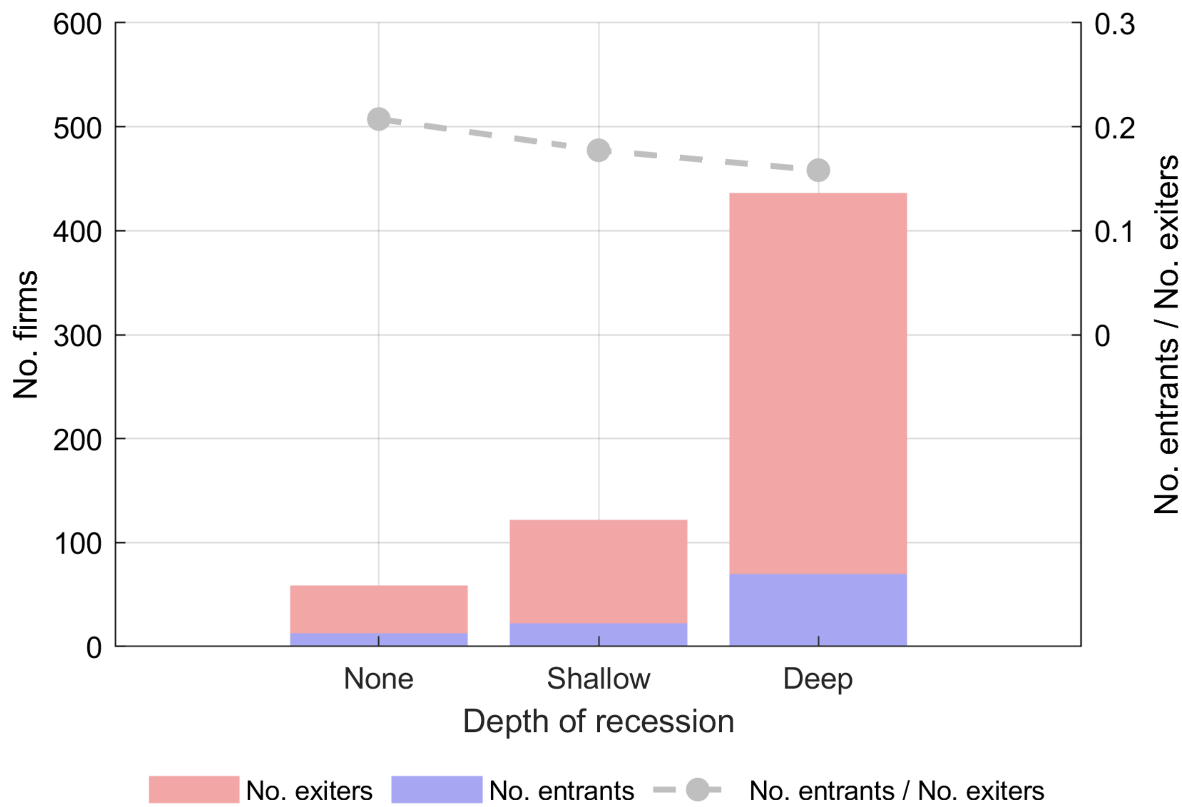
Based on this counterfactual analysis, I draw two main conclusions:

1. **The business cycle harms surplus in the short-term, improves surplus in the medium-term and has no effect in the long-term**.
2. **Deeper business cycles have bigger effects in the short-term and medium-term**.

I set out the impact on firm turnover in Figure 9. The number of exiting firms and the number of entering firms increasing in the depth of the recession. In Figure 10, I show the net effect of firm turnover on aggregate surplus per-period. It is initially negative, indicating that the information loss effect dominates the cleansing effect. Over time, as the funds age, the information loss effect decays, such that there is a "switching point" at around month 30 after the business cycle when the effect of the firm turnover is reversed: the cleansing effect dominates the information loss effect, and per-period aggregate surplus is higher. This reversal occurs because as funds age they acquire a longer returns history and the precision of investor beliefs increases. Because the exiting funds are older, however, the marginal improvement in investor precision over time is much smaller than for the new entrant funds. An extra datapoint is more valuable for funds with few datapoints. In other words, the decay of the information loss effect over time is not about the change in the *absolute* precision of investor beliefs about entrants, but instead about the change in their precision *relative* to the precision about exiting funds.

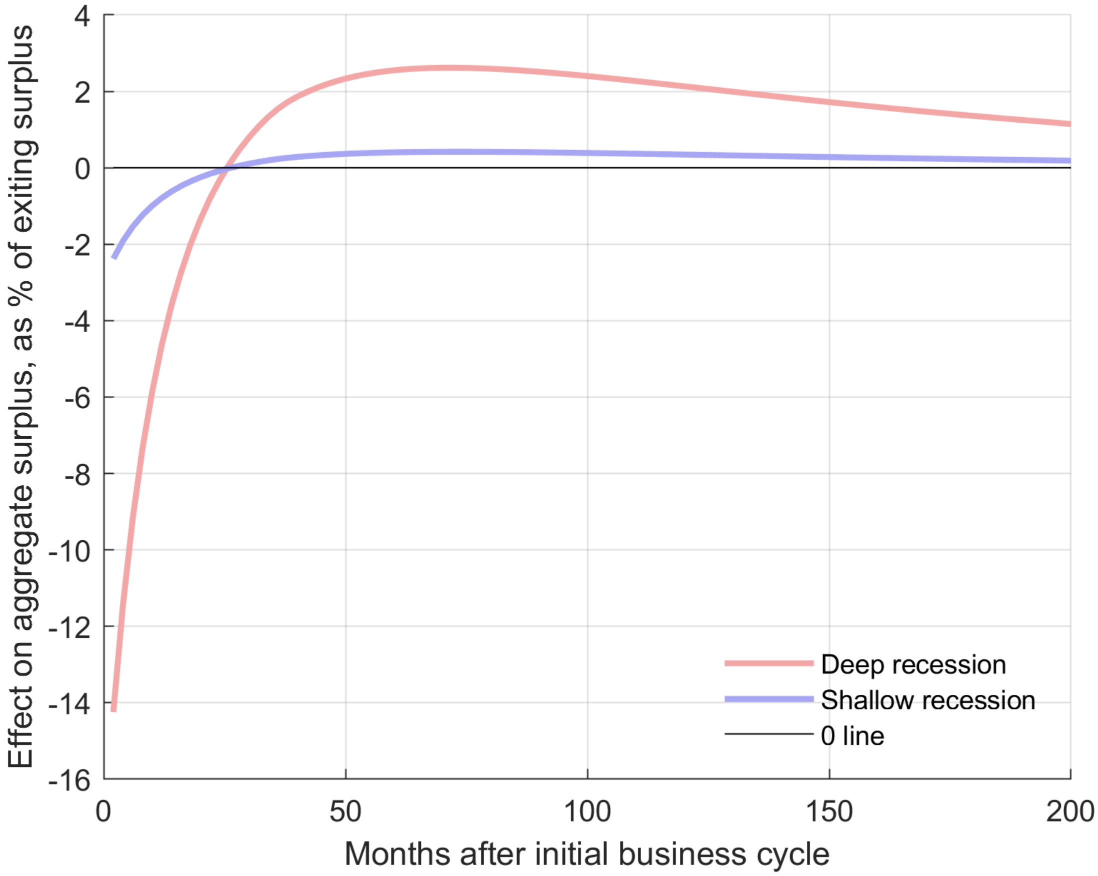
The magnitudes of both the short-term and long-term effects are material relative to aggregate surplus, although not large (as most funds never exit or enter), and are increasing in the depth of the business cycle. For the deepest business cycle I model (which is roughly equivalent to the financial crisis), the aggregate surplus of entering funds is 14% less than the aggregate surplus of the exiting funds in the first month after the recovery. The effects on total surplus in the market (including funds that did not exit) are material but small (ranging between -1.0% and 0.2% of total mutual fund surplus for the deepest business cycles) because the majority of funds neither enter not exit. This also means, however, that the effects are persistent (as can be seen in Figure 10, which simulates 200 months after the initial business cycle), because it is ongoing exit and entry that in the long-term causes the impact of a historical business cycle to decay.

Figure 9: The effect of the business cycle on firm turnover



Note: Deeper recessions and subsequent recoveries result in greater firm turnover. The composition of this firm turnover also varies with the depth of the recession, in that the ratio of the number of exiting funds to the number of entering funds decreases (because small funds are relatively more affected by deep recessions than large funds).

Figure 10: The effect of the business cycle on aggregate surplus



Note: I simulate a recession and recovery at time 0, which results in firm turnover. In this figure I plot the net effect of this firm turnover on per-period aggregate surplus over time. Immediately after the recession, aggregate surplus in the mutual fund industry is lower: the information loss dominates the cleansing effect. As the entrants age, investors obtain a returns history and the information loss effect decays: 30 months after the recovery the cleansing effect dominates the information loss effect and the firm turnover improves aggregate surplus. The sizes of both effects are increasing in the depth of the business cycle. In the long-term, ongoing exit and entry mean that the impact of a business cycle at time 0 decays to 0.

6.2 Optimal subsidies in a pandemic

Having modelled the effect of the business cycle on firm exit and thus on outcomes, it is natural to then consider the extent to which policy should mitigate these effects. Specifically, I consider which mutual funds, if any, should optimally be subsidised during a predictably temporary recession such as that resulting from the Covid-19 pandemic.

To do this, I add predictably temporary shocks to the dynamics of the macroeconomic factor M_t set out in my baseline model above. In any given period, there is a probability p^C that there is a "pandemic", which is a negative shock to M_t of size Δ^C that lasts for a single period only. If there is no pandemic, then M_t evolves according to the dynamics set out above in Equation 14. Being in a pandemic, or not, is then an additional element of the state-space that firms face when making their decisions to exit or not.

For simplicity, I assume that $p_C = 0$, meaning that I do not need to change my estimation of the dynamics of M_t, plus it means that the long-term expectations process of the firms is unaffected by the possibility of a pandemic. In the short-term, of course, firms know that the negative shock of Δ^C will be reversed in the following period. In Figures 11, I set out exit probabilities across state buckets in the pandemic state and in the non-pandemic state: naturally, holding everything else equal, most firms are less likely to exit in the pandemic state (when they know that the macroeconomy will improve in the next period) than in the non-pandemic state (when the macroeconomic factor is as likely to go down as up).

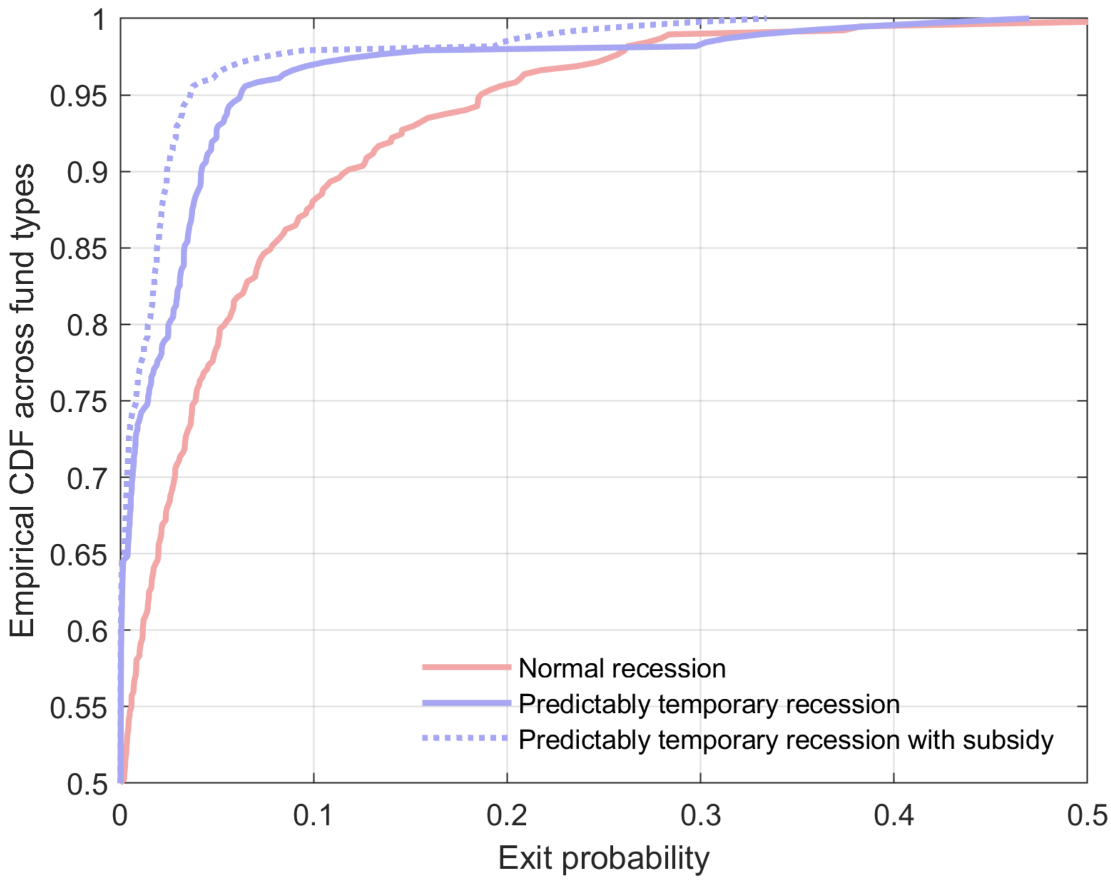
In making these assumptions I abstract away from various issues that are important in the broader context of the pandemic, such as whether the pandemic is likely to have persistent direct economic effects, whether it was anticipated by firms and whether it (or a related pandemic) is likely to reoccur in the future. These issues, and the dynamics of the macroeconomic factor M_t more generally, are out of scope of my research question. The only important thing in my context is that the pandemic induces a recession in which things are predictably going to improve in the future.

I consider public, unanticipated, one-off subsidies τ to specific types of fund, contingent on not exiting:

$$V_{it}(\mathbf{S_t}; \mathbf{\Theta_i}) = \max_{z_{it}} \; z_{it} f_i q_{it}(\mathbf{S_t}; \mathbf{\Theta_i}) + z_{it}\tau_i - z_{it} W(\mathbf{S_t}; \mathbf{\Theta_i}) + \eta(z_{it}) + z_{it}\beta\mathbb{E}[V_{it+1}(\mathbf{S_{t+1}}; \mathbf{\Theta_i})] \quad (17)$$

A subsidy to state-type-bucket i reduces the probability of exit in that period, which then reduces firm entry in the following period. The aggregate size of the industry Q_t and its relationship with M_t is unchanged, the only impact of the subsidy is on composition: it preserves funds of a particular type, and prevents entry of random type. I illustrate this by setting out the impact of a generic subsidy to all funds in Figure 11: unsurprisingly subsidies make firms less likely to exit.

Figure 11: The effect of recession type and subsidies on exit probabilities



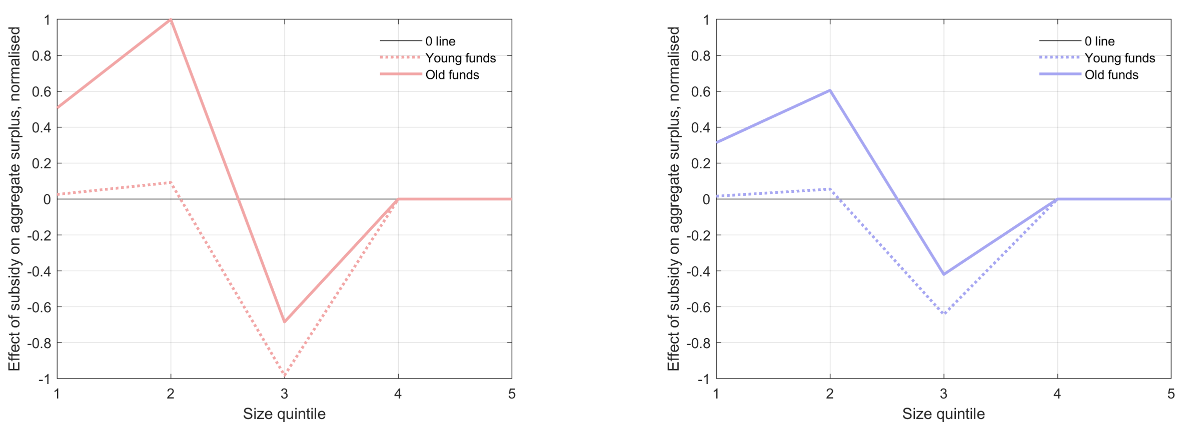
Note: I plot the empirical distribution of exit probabilities across my state-type buckets in a normal recession (where future changes are not known), in a predictably temporary recession and in a predictably temporary recession in which firms have been subsidised.

I describe in the preceding sub-section how the information loss and cleansing effects trade off against each other on aggregate across the industry. The key insight of this sub-section is that this trade-off varies at the fund-level: for example, the information loss effect resulting from the exit of a very young fund is relatively low, as that fund does not have an extensive returns history (there is little information to lose, in other words). Similarly, there is no positive cleansing effect resulting from the exit of a very high ability fund, because a randomly drawn new entrant is likely to be lower ability.

I demonstrate the fund-specific nature of the trade-off by assuming that the social planner can subsidise only a single type of fund (one of the 384 state-type buckets I describe in my empirical approach). A given subsidy has a persistent effect on aggregate surplus in the following periods, as modelled in the previous sub-section. I calculate the net present value of this effect on aggregate surplus, assuming an annualised discount rate of 0.03.[6] I then search over possible subsidy levels to calculate the optimal subsidy that induces the biggest increase in the net present value of aggregate surplus. Finally, I then compare this maximal effect on surplus across fund types, to ask which fund type would be subsidised by a social planner if it could subsidise only one. I abstract away from questions around the funding or implementation of the subsidy. I set out my results in Figure 17, where for clarity I aggregate fund-type buckets by their effect on size.

Figure 12: Optimal subsidies by type

(a) Deep recession (b) Shallow recession



Note: These figures show how subsidies should optimally be targeted at smaller, older funds. I set out the returns to subsidising 10 distinct types of fund: size quintiles for older (age greater than 40 months) and younger funds (age between 6 and 40 months). I do this during predictably temporary deep (panel (a)) and shallow recessions (panel (b)).

[6]My results do not change substantively if I choose some other reasonable discount rate.

The funds with the greatest returns to being subsidised are smaller, older funds. The intuition on the age effect is straightforward: older funds have a socially valuable returns history that the social planner wants to preserve. The intuition on the size effect is more nuanced. Very large funds do not exit with or without the subsidy (meaning the subsidy has little effect on aggregate surplus). Medium-sized funds (the middle quintile in the figure above) are relatively large, but also have large fixed costs, meaning that although they are relatively high ability their net effect on surplus is negative because of their costs. In other words, they are the funds that a social planner would wish to cleanse. The benefits (and potential costs, if the wrong funds are subsidised) are greater the deeper the recession.

For clarity I revisit the empirical facts that drive these counterfactual results. It is more efficient to subsidise older funds because I find that returns are moderately informative, based on the empirical relationship between fund size and past returns. This age effect would be lower if the empirical relationship between fund size and past returns were much stronger (if returns are very informative, then investors have precise beliefs even about relatively young funds) or much weaker (if returns are not informative, then investors beliefs are no more precise about older funds). The size effect I find in optimal subsidies follows from how size and empirical exit rates vary across fund types: medium-sized funds are much larger than small funds, but not much less likely to exit, which leads my estimation to conclude they have high costs. This leads the social planner to choose to subsidise smaller funds rather than medium-sized funds.

7 Conclusion

The persistent effects of the business cycle have been extensively studied in macroeconomic contexts, but less so in market-specific microeconomic contexts. The main contribution of this paper is to develop an under-explored implication of business cycles: the information loss that results from firm turnover. My quantitative results are for the mutual fund industry, but a similar trade-off between cleansing and information loss applies in any industry in which unobserved quality is important for outcomes and past performance is informative about quality.

I draw two primary conclusions that may be important in broader contexts, and have particular relevance for policy during the Covid-19 pandemic. First, the information loss effect can be large, and indeed in my context it dominates the cleansing effect in the short-term. This has what I believe to be a novel implication for policy during the pandemic: subsidies in the pandemic are intended to preserve beneficial connections through a temporary recession, and one of those benefits is the *value of information* that has been built up about an existing firm. Second, the cleansing effect can also be large, but the firm types that should be cleansed and allowed to fail may not be immediately obvious: in my context, what matters is not just a firm's size and ability, but its size and ability *relative to its costs*. These costs may not be observable, but can be inferred from firm behaviour relating to pricing or, as in this paper, exit.

References

Acemoglu, D., Chernozhukov, V., Werning, I., and Whinston, M. D. (2020). A multi-risk sir model with optimally targeted lockdown. *NBER Working Paper No. 27102*.

Barber, B. M., Huang, X., and Odean, T. (2016). Which factors matter to investors? Evidence from mutual fund flows. *The Review of Financial Studies*, 29(10):2600–2642.

Barras, L., Scaillet, O., and Wermers, R. (2010). False discoveries in mutual fund performance: Measuring luck in estimated alphas. *The Journal of Finance*, 65(1):179–216.

Berk, J. B. and Green, R. C. (2004). Mutual fund flows and performance in rational markets. *Journal of Political Economy*, 112(6):1269–1295.

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

Berk, J. B. and Van Binsbergen, J. H. (2017). Regulation of charlatans in high-skill professions. Technical report, National Bureau of Economic Research.

Bigio, S., Zhang, M., and Zilberman, E. (2020). Transfers vs credit policy: Macroeconomic policy trade-offs during Covid-19. *NBER Working Paper No. 27118*.

Brodeur, A., Gray, D. M., Islam, A., and Bhuiyan, S. (2020). A literature review of the economics of Covid-19.

Caballero, R. J. and Hammour, M. L. (1996). On the timing and efficiency of creative destruction. *The Quarterly Journal of Economics*, 111(3):805–852.

Castillo-Martınez, L. (2018). Sudden stops, productivity, and the exchange rate. *LSE manuscript*.

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

Deuskar, P., Pollet, J. M., Wang, Z. J., and Zheng, L. (2011). The good or the bad? Which mutual fund managers join hedge funds? *The Review of Financial Studies*, 24(9):3008–3024.

Eichenbaum, M. S., Rebelo, S., and Trabandt, M. (2020). The macroeconomics of epidemics. *NBER Working Paper No. 26882*.

Fama, E. F. and French, K. R. (2010). Luck versus skill in the cross-section of mutual fund returns. *The Journal of Finance*, 65(5):1915–1947.

Gavazza, A. (2011). Demand spillovers and market outcomes in the mutual fund industry. *The RAND Journal of Economics*, 42(4):776–804.

Gil-Bazo, J., Hoffmann, P., and Mayordomo, S. (2020). Mutual funding. *The Review of Financial Studies*, 33(10):4883–4915.

Gil-Bazo, J. and Ruiz-Verdú, P. (2009). The relation between price and performance in the mutual fund industry. *The Journal of Finance*, 64(5):2153–2183.

Glode, V. (2011). Why mutual funds "underperform". *Journal of Financial Economics*, 99(3):546–559.

Hale, G. (2012). Bank relationships, business cycles, and financial crises. *Journal of International Economics*, 88(2):312–325.

Hotz, V. J. and Miller, R. A. (1993). Conditional choice probabilities and the estimation of dynamic models. *The Review of Economic Studies*, 60(3):497–529.

Ibert, M., Kaniel, R., Van Nieuwerburgh, S., and Vestman, R. (2018). Are mutual fund managers paid for investment skill? *The Review of Financial Studies*, 31(2):715–772.

Jiang, H. and Zheng, L. (2018). Active fundamental performance. *The Review of Financial Studies*, 31(12):4688–4719.

Kacperczyk, M., Nieuwerburgh, S. V., and Veldkamp, L. (2014). Time-varying fund manager skill. *The Journal of Finance*, 69(4):1455–1484.

Kacperczyk, M. and Seru, A. (2007). Fund manager use of public information: New evidence on managerial skills. *The Journal of Finance*, 62(2):485–528.

Kacperczyk, M., Van Nieuwerburgh, S., and Veldkamp, L. (2016). A rational theory of mutual funds' attention allocation. *Econometrica*, 84(2):571–626.

Kosowski, R. (2011). Do mutual funds perform when it matters most to investors? US mutual fund performance and risk in recessions and expansions. *The Quarterly Journal of Finance*, 1(03):607–664.

Nanda, V. K., Wang, Z. J., and Zheng, L. (2009). The abcs of mutual funds: On the introduction of multiple share classes. *Journal of Financial Intermediation*, 18(3):329–361.

Pástor, L. and Stambaugh, R. F. (2012). On the size of the active management industry. *Journal of Political Economy*, 120(4):740–781.

Pástor, L., Stambaugh, R. F., and Taylor, L. A. (2015). Scale and skill in active management. *Journal of Financial Economics*, 116(1):23–45.

Pástor, L., Stambaugh, R. F., and Taylor, L. A. (2020). Fund tradeoffs. *Journal of Financial Economics*, 138(3):614–634.

Pástor, L. and Veronesi, P. (2009). Technological revolutions and stock prices. *American Economic Review*, 99(4):1451–83.

Pollet, J. M. and Wilson, M. (2008). How does size affect mutual fund behavior? *The Journal of Finance*, 63(6):2941–2969.

Roussanov, N., Ruan, H., and Wei, Y. (2021). Marketing mutual funds. *The Review of Financial Studies*, 34(6):3045–3094.

Roussanov, N. L., Ruan, H., and Wei, Y. (2020). Mutual fund flows and performance in (imperfectly) rational markets? *Jacobs Levy Equity Management Center for Quantitative Financial Research Paper*.

Schumpeter, J. A. et al. (1939). *Business cycles*, volume 1. McGraw-Hill New York.

Sialm, C. and Tham, T. M. (2016). Spillover effects in mutual fund companies. *Management Science*, 62(5):1472–1486.

van Binsbergen, J. H., Han, J., Ruan, H., and Xing, R. (2019a). A horizon based decomposition of mutual fund value added using transactions. *Available at SSRN 3478745*.

van Binsbergen, J. H., Kim, J. H. J., and Kim, S. (2019b). Capital allocation and the market for mutual funds: Inspecting the mechanism. *Georgia Tech Scheller College of Business Research Paper*, (3462749).

Wahal, S. and Wang, A. Y. (2011). Competition among mutual funds. *Journal of Financial Economics*, 99(1):40–59.

A Tables

Table 1: Relationship between Q_t and M_t

	[1] ΔQ_t	[2] Q_t	[3] Q_t
Intercept	0.001	7.38 x 10^{5***}	-3.62 x 10^{5***}
	(0.007)	(7.36 x 10^4)	(1.22 x 10^5)
$\mathbf{1}_{Post2008}$			1.10 x 10^{6***}
			(1.38 x 10^5)
M_t		843.82***	1826.8***
		(51.35)	(100.94)
M_t $\mathbf{1}_{Post2008}$			-1022.1***
			(108.6)
ΔM_t	0.466***		
	(0.117)		
R^2	0.15	0.75	0.90
No. obs	90	91	91

Note: Figures in parentheses are standard errors. ***, **, * indicate different from 0 at 1%, 5% and 10% significance, respectively. Q_t is the size of the mutual fund industry, M_t is the S&P500 index and $\mathbf{1}_{Post2008}$ is a dummy variable that is one after 2008. The dataset is from 2001 to 2016, at a frequency of 2 months.

Table 2: Demand-side results (first step)

	[1] q_{it}	[2] q_{it}
λ	0.099***	
	(0.00)	
$mean(\mu_i)$	0.311	0.311
$mean(\phi_i)$	0.037	0.037
Age FE	Y	Y
Time FE	Y	Y
R^2	0.73	0.70
No. obs	291,191	291,191

Note: This table summarises the results of the first step of the two-step procedure described in the text. Figures in parentheses are p-values. ***, **, * indicate different from 0 at 1%, 5% and 10% significance, respectively. q_{it} is the size of mutual fund i at time t and λ represents sensitivity to past returns. I calibrate fund-specific priors μ_i and scalability ϕ_i and report the mean across funds here.

Table 3: Demand-side results (second step)

	[1] $\hat{\delta}_t$	[2] $\hat{\delta}_t$
Intercept	-0.53***	-0.53***
	(0.00)	(0.00)
Q_t	9.30***	-7.43***
	(0.00)	(0.047)
M_t	1.11***	3.57***
	(0.00)	(0.00)
R^2	0.92	0.89
No. obs	161	161

Note: This table summarises the results of the second step of the two-step procedure described in the text. Figures in parentheses are p-values. ***, **, * indicate different from 0 at 1%, 5% and 10% significance, respectively. $\hat{\delta}_t$ is the estimated dummy variable for time t from the first step in Table 2, in which fund size q_{it} is regressed on various dummy variables and past returns. Q_t is the size of the mutual fund industry and M_t is the S&P500 index. In Column [2] I instrument for Q_t using the log of the total number of active mutual funds at time t.

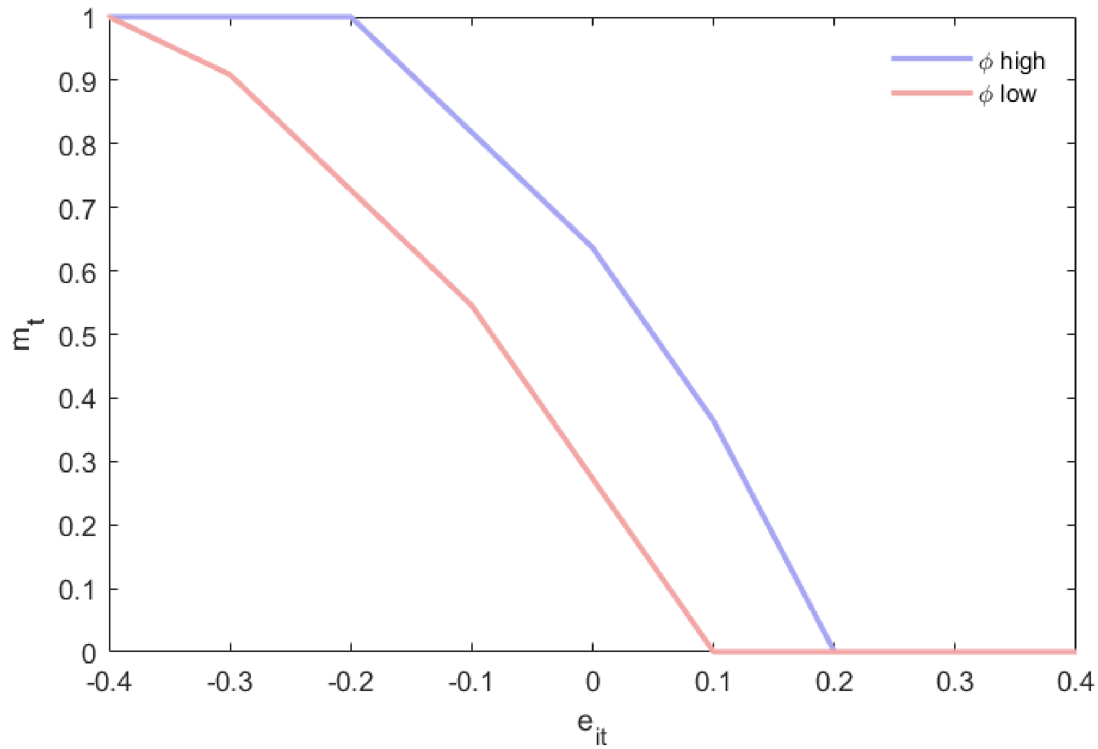
Table 4: Supply-side results

	$\hat{Pr}(Exit \mid \mathbf{S_t}; \mathbf{\Theta_i})$
Intercept	6.37*** (0.00)
e_{it}	1.64 (0.29)
age_{it}	-7.15 (0.29)
ϕ_i	-3.17*** (0.00)
M_t	2.05*** (0.00)
e_{it}^2	0.18*** (0.00)
age_{it}^2	6.84*** (0.00)
ϕ_i^2	-3.77*** (0.00)
M_t^2	-1.19 (0.94)
$e_{it}\phi_i$	-3.27*** (0.00)
$e_{it}age_{it}$	0.82*** (0.00)
$age_{it}\phi_i$	-1.38*** (0.00)
t	-0.11*** (0.00)
t^2	0.01*** (0.00)
R^2	0.699
No. obs	293

Note: Figures in parentheses are p-values. I parameterise state-type fixed costs according to Equation 16 and choose the coefficients to fit the implied exit probabilities to observed exit probabilities. The fixed costs implied by these results are plotted in Figure 8. e_{it}, ϕ_i and M_t represent expected fund ability, scalability and the macroeconomic factor, respectively.

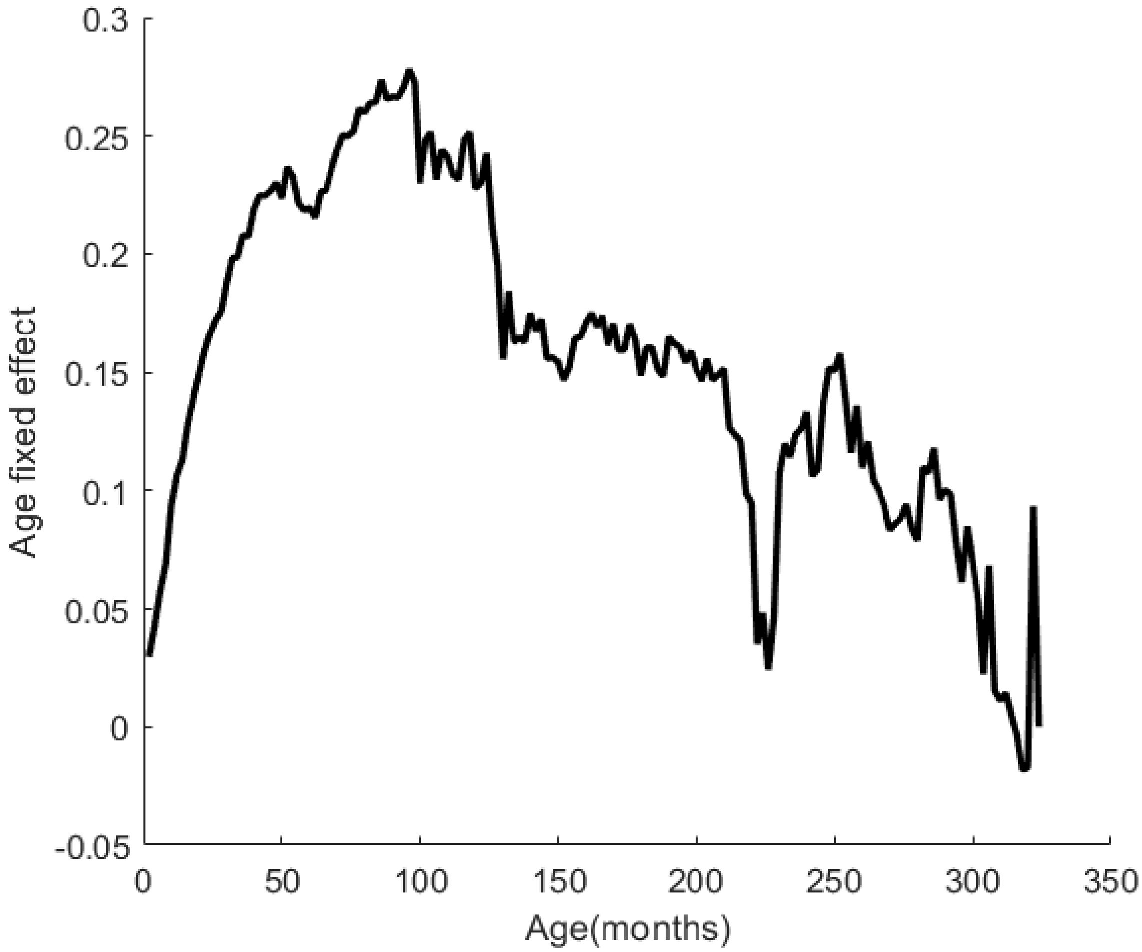
B Additional figures

Figure 13: Exit decisions



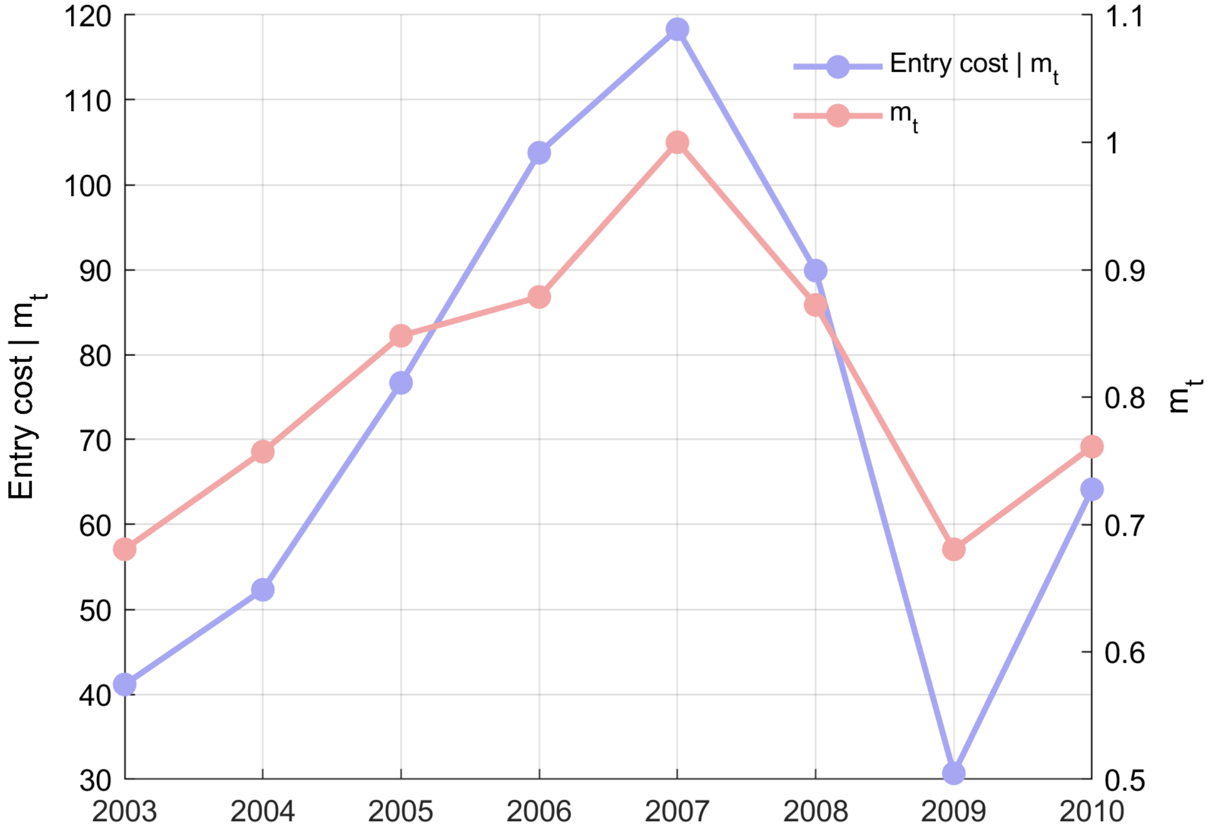
Note: The area under the curve shows the combinations of ability belief (e_{it}) and business cycle state (m_t) in which a fund exits: funds exit when they are perceived to be bad or when the macroeconomic state is bad, or some convex combination thereof. This figure is the same as figure 5, but with a coarser state space.

Figure 14: The effect of fund age on fund size



Note: This figure plots the age dummy that I estimate on the demand-side. On average, young funds grow quickly, peak at age 100 months, and then decline as they age further.

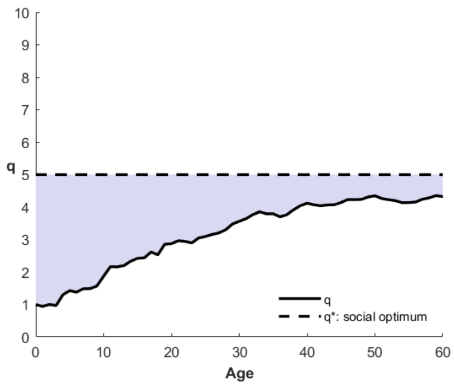
Figure 15: Variation in the fixed cost of entry over time



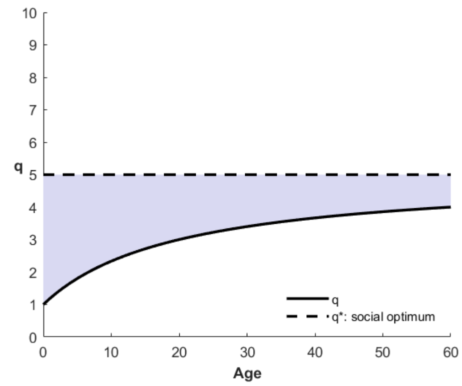
Note: In this figure I plot how the estimated fixed cost of entry F_t^{entry} varies over time (black line). The cost of entry is correlated with the macroeconomic factor M_t (the red line), but the effect is relatively weak: the maximum entry cost is only 6% greater than the minimum entry cost.

Figure 16: The effect of age on value-added

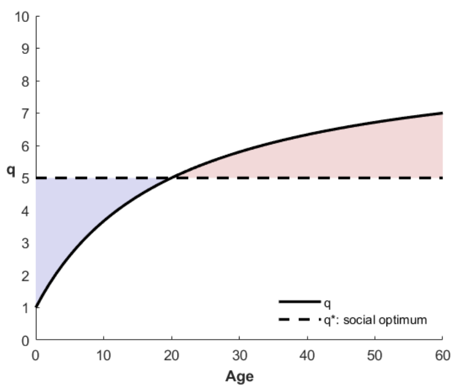
(a) Noisy convergence to optimum



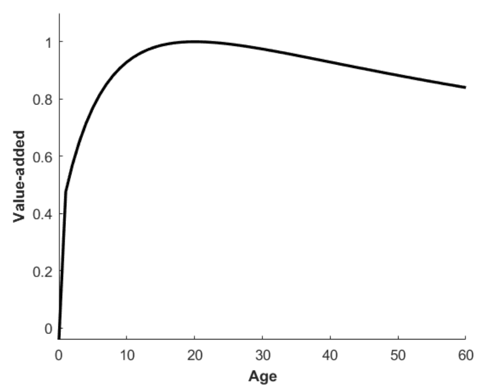
(b) Convergence to optimum



(c) Convergence towards and then away from optimum when fee too low



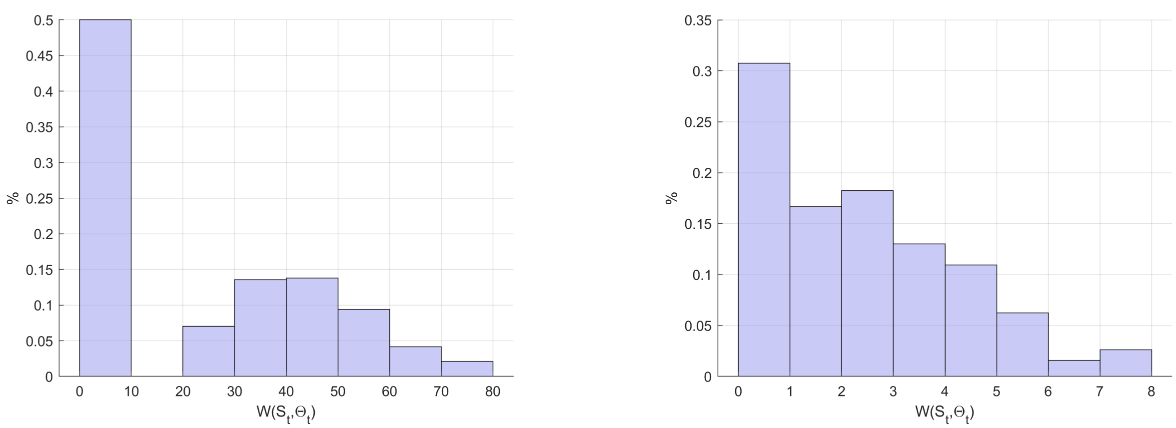
(d) Convergence towards and then away from optimum when fee too low



Note: Suppose that for a given fund true $\alpha = 0.1$ and other parameters are such that optimal fund size $q^* = 5$. If investors priors are incorrect, $\mu \neq \alpha$ and $q \neq q^*$. q is a function of investor beliefs about ability (which converge to true α as the funds ages and investors observe returns) and the fee rate f (which is fixed). This means that q converges to q^* only if f is ex-post optimal, which in this example means $f^* = \alpha/2$. In panels (a) and (b) $\mu < \alpha$ and $f = \alpha/2 = f^*$ such that q converges to q^*, with noise in the signal (panel (a)) and with the noise in the signal turned off (panel (b)). In panel (c) $\mu < \alpha$ and $f = \mu/2 < f^*$, which means that q initially converges to q^* (the blue area), but then overshoots and moves away from q^* (the red area). In other words, fund value-added does not increase monotonically with age, but is n-shaped, as in panel (d). The same is true if $\mu > \alpha$ and so $f > f^*$.

Figure 17: The distribution of state-type-specific fixed costs

(a) Full (b) Small values



Note: In this figure I plot the distribution of estimated fixed costs across states and types, including ability, scalability, age and macroeconomic factor. Panel (a) shows the full distribution. Panel (b) shows the distribution of smaller values only.

C Details on counterfactual

In this section I describe in more detail my approach to the counterfactuals.

For each of the K fund types, I simulate their size (from the demand-side of their model) and their expected surplus. Fees, fund size and fixed costs are known. I simulate the range of possible net excess returns using investor beliefs about fund ability and the precision of those beliefs. To limit age effects to the information loss channel, I remove any variation across fund age in δ_a (age dummies in fund size q_t) or W (fixed costs). I denote the $K \times 1$ vector of fund sizes and surplus across types by $\mathbf{q}$ and $\mathbf{S}$, respectively.

The supply-side of the model shows how exit rates by fund type vary across a grid of possible values of the macroeconomic factor M_t. Prior to time $t = 0$, the industry is in equilibrium at values from 2012. Let $\mathbf{C_0}$ denote the vector of firm numbers by type prior to time $t = 0$.

At time $t = 0$, I simulate a reduction in M_t that is either deep (a 32% decrease), shallow (a 18% decrease) or zero (no decrease in M_t). I use the exit rates to simulate expected exit, given the composition of fund-types. Let $\mathbf{x_m}$ denote the vector of exit rates across types for recession depth $m \in \{Deep, Shallow, Zero\}$, then the expected decrease in aggregate industry size is $\Delta_m = \iota'(\mathbf{x_m}. * \mathbf{C}. * \mathbf{q})$, where $.*$ signifies element-wise multiplication and ι is a vector of ones of length K.

At $t = 1$, I simulate a complete recovery in macroeconomic factor M_t to its value in $t = 0$, which induces entry. Entrants have mean size $\bar{q}$, and I assume that n enter such that $n_m \bar{q} = \Delta_m$. I calculate $\mathbf{C_{1,m}}$ as funds that did not exit from the previous period plus the n new funds split evenly between types. I ignore integer constraints so as to take expectations properly. For example, I allow 0.1 funds of a given type to enter. Importantly, because exit and entry are both functions of the depth of the recession, the composition of funds is now itself a function of m. For example, $\mathbf{C_{1,Deep}}$ has more young funds than $\mathbf{C_{1,Zero}}$ because there was greater firm turnover because of the recession. I calculate the effect of a business cycle of depth m on aggregate surplus as $\Delta S_{1,m} = \iota' * (\mathbf{C_{1,m}}. * \mathbf{S}) - \iota' * (\mathbf{C_{1,Zero}}. * \mathbf{S})$. This is the immediate effect at time $t = 1$.

At time $t = 2$, firms exit and enter again. Exit rates are now the same across m, because the macroeconomic factor at time $t = 2$ does not depend on its level at time $t = 0$. The only thing that varies depending on the initial business cycle is firm composition. I simulate entry and exit for a further 100 periods in an analogous manner to above: applying exit rates to firm numbers, then allowing entry such that in expectation the aggregate size of the industry is unchanged. Two compositional effects occur over time. First, firms age and shift from young types to older types, meaning that the precision of investor priors changes. Second, ongoing exit and entry causes the differences across m, the depth of the time 0 recession, to decay. For example, a deep recession forces out more bad funds at time $t = 0$, relative to a shallow recession. In future periods $t > 1$ this effect is reversed: more bad funds exit in the shallow recession counterfactual than in the deep recession counterfactual, because in the deep recession counterfactual many bad funds had already exited at time $t = 0$. Composition thus converges over time regardless of the depth of the recession at time $t = 0$.