THE UNIVERSITY OF CHICAGO

COMPETITOR SCALE AND MUTUAL FUND BEHAVIOR

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE UNIVERSITY OF CHICAGO
BOOTH SCHOOL OF BUSINESS
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

BY
LASZLO PAL JAKAB

CHICAGO, ILLINOIS
JUNE 2018

Anyunak, aki sose adja fel.

Apunak, aki ablakot nyitott a világra.

And to Jackie, my companion in adventures old and new.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I thank my advisors Ľuboš Pástor, Samuel Hartzmark, Elisabeth Kempf and Michael Weber for their support. I also thank Douglas Diamond, Eugene Fama, Zhiguo He, Juhani Linnainmaa, Stefan Nagel, John Shim, Willem van Vliet, Taras Zlupko, and seminar participants at Chicago Booth for their helpful comments.

ABSTRACT

I study the effects of competition on the investment behavior and performance of active mutual funds. I find that funds respond to increased competitor scale by curtailing costly active management. To establish causality, I exploit quasi-exogenous variation in fund flows created by a natural experiment—the 2003 mutual fund scandal. Funds whose competitors were the most affected by the scandal expand active management and perform better after the scandal. Interpreting the findings through the lens of models of decreasing returns to scale indicates information asymmetry between fund managers and outside investors.

CHAPTER 1

INTRODUCTION

What is the impact of competition on the investment decisions of active mutual funds? Competition affects both a fund's efficient scale and the optimal composition of its investments. Outside investors allocate capital to the fund based on its perceived ability to generate risk adjusted returns. Fund managers allocate capital across available investment opportunities, taking the size of the fund as given. There is a key tension when funds face decreasing returns to scale. As competition eliminates investment opportunities, the fund's optimal response is to reallocate capital toward passive portfolios. However, if investors are symmetrically informed, they withdraw capital from the fund (Berk and Green, 2004). Decreased fund size lowers the marginal cost of active management, countervailing the shift toward passivity. Models of fund behavior predict the size effect to dominate in equilibrium, implying increased capital allocated to active strategies in response to competition (Pástor et al., 2017b).

I exploit changes in the size of competing funds to identify the effect of competition on funds' capital allocation. I demonstrate that funds respond to competition by reallocating capital from costly active strategies to cheaper, more passive portfolios. I address endogeneity concerns by investigating a natural experiment created by the 2003 mutual fund scandal. Fund families engulfed by the scandal were penalized by investor outflows, which I exploit as quasi-exogenous variation in competitor size. I show that close competitors of affected funds increased active management, and reaped improved performance following the scandal.

Interpreting my findings through models of fund behavior featuring decreasing returns to scale (Berk and Green, 2004; Pástor et al., 2017b) points to information asymmetry between funds and outside investors resulting in mismatch between investment opportunities and capital that is not fully undone by the firm's actions (Berk et al., 2017). This interpretation also provides a potential explanation for the observed relation between measures of activeness such as active share (Cremers and Petajisto, 2009) or industry concentration (Kacperczyk et al., 2005) and future fund performance: if managers have private information about investment opportunities, their actions carry information about expected returns.

Pástor and Stambaugh (2012) argue that each fund's investment opportunities become less lucrative as the size of competing funds increases. Decreasing returns to competitor scale are grounded in liquidity constraints and the associated price impact of other funds' trades. Consider a skilled fund receiving signals of the fundamental value of securities. In the absence of competitors, the limiting factor of the fund's profits is the price impact of its own trades. Introducing another fund that receives correlated signals is detrimental to the fund's profitability. Since both funds chase similar investments, either one might be first to invest in a particular opportunity, pushing up its price. The total impact of the other fund will depend on both the similarity of its signals, which determines the likelihood of being leapfrogged, and the fund's size, which governs the magnitude of price impact. Competitor size is therefore the sum of the product of similarity and fund size across all potential competitors:

$$\text{Competitor Size}_i = \sum_{j \neq i} \text{Similarity}_{i,j} \times \text{Fund Size}_j.$$

If funds receive identical signals, similarity equals one, and competitor size will equal aggregate industry size (Pástor et al., 2015).[1] Unlike industry size, competitor size is specific to each fund, which allows for studying variation in the cross-section. Most importantly, taking fund similarity into account enables me to analyze novel evidence on decreasing returns to competitor scale from a natural experiment.

Berk and Green (2004) posit a relation between active share, fund size, and fees. Pástor et al. (2017b) introduce a richer model of fund behavior, proposing that funds jointly optimize turnover and portfolio liquidity. Portfolio liquidity is a multi-dimensional object that can be

1. I measure fund similarity by the cosine similarity of market capitalization adjusted portfolio weights. I cap adjust portfolio weights by dividing them with market weights, as cross-holdings of small-capitalization stocks are more informative of similarity than cross-holdings of large capitalization stocks (Cohen et al., 2005).

decomposed as a product of stock liquidity and diversification (itself a product of coverage and balance). I make the model-driven argument that if investors do not fully recognize the detrimental impact of competition on future returns, then increases in competitor size will have a measurable impact on fund activeness even conditional on own size, fees, and time fixed effects.

My empirical analysis is based on a sample of actively managed U.S. equity funds spanning 1980-2016, with size and returns information from CRSP linked to Thomson Reuters holdings data. Guided by theory, I relate quarterly changes in competitor scale to changes in active share, turnover to portfolio liquidity ratio, and various components of portfolio liquidity, conditional on own size, expense ratio, and time fixed effects. I find that funds react to increases in competitor scale by decreasing active share and increasing all dimensions of portfolio liquidity.

I bolster the causal link between competitor scale, fund behavior, and performance by providing novel evidence from a natural experiment created by the 2003 mutual fund scandal. In September 2003, the New York State Attorney General announced investigations into illegal trading practices at several prominent mutual fund families. As investigations gained momentum, evidence mounted that families had allowed favored clients to abuse ordinary investors by trading fund shares at stale prices (Zitzewitz, 2006). By October 2004, a total of twenty-five fund families were embroiled in the scandal (Houge and Wellman, 2005). The involved families represented a considerable proportion of the industry, collectively managing over a fifth of assets prior to the scandal. Following the announcement of the investigations, investors abruptly began withdrawing capital from tainted families (Figure 6.1).

I exploit post-scandal outflows at tainted funds as an exogenous shock to the competitor size of funds pursuing similar investment strategies. We would expect the favorable impact of lessened competitor scale to be greatest for the closest pre-scandal competitors of tainted funds. Under the hypothesis of decreasing returns to competitor scale, these funds experience a comparative improvement in their investment opportunities. Therefore, we would expect them to increase capital allocation to active strategies relative to less close competitors of tainted funds, and see relative improvements in performance. I take two different approaches to testing these hypotheses, both of which confirm decreasing returns to competitor scale and the associated fund response.

Since involved funds are directly affected by the scandal, I identify decreasing returns to competitor scale by comparing outcome paths at untainted funds. The first approach compares pre- and post-scandal outcomes as a function of pre-scandal exposure to competition from tainted funds. I measure exposure by the fraction of competitor scale in August 2003 accounted for by prospective tainted families. The competitor size of high exposure funds decreased significantly more during the scandal. Consistent with comparatively improved investment opportunities, high exposure funds increased active share and decreased portfolio liquidity relative to low exposure funds, and experienced comparatively better performance. Statistical tests show no evidence of differential trends by scandal exposure in the pre-period.

The second approach links fund outcomes directly to abnormal outflows at tainted funds. I use a linear model to decompose post-scandal flows at involved funds between time variation common to all funds and abnormal flows attributable to scandal involvement. I show that untainted funds whose tainted competitors experienced greater abnormal outflows saw relative declines in competitor size, improvements in performance, and shifted to more active portfolio management. Variation in competitor size attributable purely to abnormal outflows is negatively related to both fund performance and activeness, providing direct quasi-experimental evidence of decreasing returns to competitor scale.

The picture which emerges from these analyses is one in which portfolio managers optimize investment behavior in real time as they respond to fluctuations in investment opportunities that are not immediately apparent to outside investors. Such a world seems sensible. It is unlikely that retail investors pay the same level of attention to market developments as professional portfolio managers, who make trading decisions based on their perception of investment opportunities on a daily basis. Since they have more flexibility over trading than over expense ratios, portfolio allocation is an important dimension of optimizing behavior. This interpretation is also consistent with recent evidence from the literature on fund optimizing behavior in the face of time-varying investment opportunities. Kacperczyk et al. (2016) argue that mutual funds allocate attention optimally between factor timing and stock picking as the nature of opportunities varies over the business cycle. Pástor et al. (2017a) present evidence that funds exploit improved investment opportunities by increasing turnover.

While the rise of "closet indexing" has received much attention and disapproval, scaling back active management ameliorates the pernicious effects of decreasing returns to scale, as it brings the costs of active trading closer in line with decreased benefits. Absent immediate outflows, deteriorating investment opportunities make a fund "too large," to which optimizing managers react by switching to passive strategies. In this way, "closet indexing" might in some instances be less a symptom of mendacious managers than of imperfect flows causing mismatch between capital and investment opportunities.

The rest of the paper proceeds as follows. Chapter 2 reviews the related literature. Chapter 3 discusses the theoretical framework motivating the empirical analyses. Chapter 4 describes the data and the construction of competitor size. Chapter 5 presents an empirical analysis of capital allocation and competitor scale. Chapter 6 presents evidence from the natural experiment created by the 2003 mutual fund scandal. Chapter 7 concludes.

CHAPTER 2

LITERATURE REVIEW

The investment behavior of mutual funds has previously been studied in the context of decreasing returns to own scale. Pollet and Wilson (2008) investigate funds' response to inflows. They find that funds diversify in response to new flows, especially if they operate in relatively illiquid markets. However, the extent of diversification is small compared to the tendency to mechanically scale up existing holdings. Pástor et al. (2017b) develop and test a model of decreasing returns to scale in which size, turnover, portfolio liquidity, and fund expense ratios are determined jointly in equilibrium. They show that in the cross-section, larger funds tend to trade less, cost less, and hold more liquid portfolios, all of which is consistent with decreasing returns to own scale. However, no paper to date has examined the impact of competition on fund behavior.

The existing literature has provided evidence of a negative relation between fund performance and competition. Wahal and Wang (2011) find that entry by similar funds is associated with decreased flows, performance, and increased exit for incumbents. Pástor et al. (2015) perform a within-fund analysis showing a negative relationship between performance and aggregate industry scale. Hoberg et al. (2018) use holdings-based estimates of fund similarity to measure the number of competing funds, finding that the number of similar funds is negatively related to both the level and the persistence of performance in the cross-section. My primary contribution to this literature is to improve identification by analyzing evidence from a natural experiment provided by the 2003 mutual fund scandal. I also provide additional observational evidence that fund performance is decreasing in competitor scale, especially for funds pursuing less liquid strategies.

My investigation of fund behavior is informed by models of decreasing returns to scale by Berk and Green (2004) and Pástor et al. (2017b). Models of decreasing returns to scale rely on the assumption that trading costs increase in the size of trades due to price impact, which is especially severe when involving illiquid securities. Busse et al. (2017) provide empirical evidence for this assumption using transaction-level data on mutual fund trades. Papers presenting evidence on price pressure due to mutual fund actions include Coval and Stafford (2007), Khan et al. (2012), Lou (2012), Antón and Polk (2014) and Blocher (2016).

The preponderance of existing empirical evidence examining fund performance supports fund level decreasing returns to scale, although the literature is not in full agreement. Chen et al. (2004) document decreasing returns to scale using cross-sectional regressions. Reuter and Zitzewitz (2015) exploit inflows following discrete Morningstar ratings changes to study the size-performance relation in a regression discontinuity framework, finding little evidence of decreasing returns to scale. Pástor et al. (2015) find a negative within-fund association between fund size and performance, but the economic magnitude of the effect is small, and the coefficients statistically insignificant when using bias-free estimation methods. McLemore (2016) studies returns following fund mergers, finding that the increased size of the acquiring fund is accompanied by decreased performance. In contemporaneous work, Harvey and Liu (2017) use a random effects model and estimate economically significant decreasing returns to own scale.

In a broad sense, I contribute to a long line of inquiry into the the nature of skill and constraints among active funds. The typical active fund fails to generate risk-adjusted returns (Jensen, 1968; Malkiel, 1995, 2013; Gruber, 1996; French, 2008; Fama and French, 2010). It would appear at first glance that skill is in short supply among active funds, a puzzle given the vast resources they manage. However, concurrent poor performance and large size is consistent with a combination of skill and decreasing returns to scale (Berk and Green, 2004; Pástor and Stambaugh, 2012). My analysis gives additional credence to the existence of economically important constraints in active management due to decreasing returns to scale. The empirical results I present are consistent with optimizing behavior by portfolio managers in the face of evolving constraints in imperfect capital markets.

CHAPTER 3

THEORETICAL MOTIVATION

In neoclassical models of capital allocation, firms trade off the productivity gains of allocating additional capital to segments in which they possess particular skill with decreasing returns to scale [mp02]. A similar dynamic governs models of active mutual funds such as the canonical Berk and Green (2004) model or Pástor et al. (2017b): funds generate alpha by deploying their skill in active investing, but face decreasing returns to scale. I model competition as a source of time variation in the return to skill in the Berk and Green (2004) or Pástor et al. (2017b) framework, and argue that its impact on fund behavior and performance depends on whether the variation in the investing environment is equally observable to fund managers and outside investors.

Consider fund i managing $q_{i,t}$ assets in Berk and Green (2004). The fund posts a fixed expense ratio f,[1] and splits assets between active and passive management according to $q_{i,t} = A_{i,t} + P_{i,t}$. While active management allows the fund to take advantage of positive NPV investment opportunities in its area of core competence, it also subjects the fund to quadratic trading costs. I parametrize costs as $C(A_{i,t}) = \frac{c_t}{M_t} A_{i,t}^2$, where $A_{i,t}$ is the amount actively managed, M_t the size of the market, and c_t a constant representing period by period trading costs. Normalizing trading costs by M_t implies that price impact per dollar of investment is lower when total market capitalization is higher. With the normalization, the model's predictions are in terms of $FundSize_{i,t} = q_{i,t}/M_t$, instead of the dollar value of assets under management.

Let $\mu_{i,t} = E(R_{t+1} \mid R_1, \ldots, R_t)$ be the fund's expected skill, inferred from publicly available information. In addition, suppose that the returns to skill depend on time-varying external

1. Following Berk and Green, I assume that the fixed expense ratio f satisfies $f < f^*$, where f^* is the expense ratio corresponding to profit maximizing fund size q^*.

factors $x_{i,t}$, such that expected effective skill is $\mu_{i,t}g(x_{i,t})$. The fund's net alpha becomes

$$\frac{A_{i,t}}{q_{i,t}}\mu_{i,t}g(x_{i,t}) - \frac{c_t A_{i,t}^2}{q_{i,t}M_t} - f. \tag{3.1}$$

I focus on competitor size as the external constraint of interest. However, $x_{i,t}$ could in principle be any time-varying external factor affecting the fund's investment opportunities.

Following equation (26) in Berk and Green (2004), the fund's profit maximizing choice for the amount of assets to keep under active management, conditional on overall size and market conditions, is $A_{i,t}^*(\mu_{i,t}g(x_{i,t})) = \frac{\mu_{i,t}g(x_{i,t})M_t}{2c_t}$. This implies that the fraction of assets under active management is governed by the first-order condition:

$$\frac{A_{i,t}^*}{q_{i,t}} = \frac{\mu_{i,t}g(x_{i,t})}{2c_t(q_{i,t}/M_t)}. \tag{3.2}$$

Conditional on its size, the fund optimally responds to deterioration in the NPV of investment opportunities by scaling back active management.

3.1 Symmetric Information

In perfect capital markets investors are symmetrically informed of the fund's time varying investment opportunities, and allocate capital to the fund until its net alpha is zero. The market clearing zero net alpha condition implies that fund size increases with the square of $\mu_{i,t}g(x_{i,t})$:

$$\frac{q_{i,t}^*}{M_t} = \frac{(\mu_{i,t}g(x_{i,t}))^2}{4c_t f}. \tag{3.3}$$

Combining equation (3.2) and (3.3), the equilibrium fraction of assets under active management is:

$$\frac{A_{i,t}^*}{q_{i,t}^*} = \frac{2f}{\mu_{i,t}g(x_{i,t})}. \tag{3.4}$$

In perfect capital markets with symmetrically informed fund managers and outside investors, the share of assets under active management is decreasing in the profitability of investment opportunities $\mu_{i,t} g(x_{i,t})$, conditional on fund expense ratio.

Testing the two above predictions separately would require modeling the evolution of $\mu_{i,t}$. However, we can combine the equilibrium conditions to eliminate fund skill and take logs to obtain

$$2\ln(A^*_{i,t}/q^*_{i,t}) = \ln(f) - \ln(c_t) - \ln(q^*_{i,t}/M_t). \tag{3.5}$$

This leads to the first hypothesis.

Hypothesis 1 (Symmetric Information): *If managers and investors share the same beliefs about the fund's investment opportunities, the share of assets under active management is fully determined by fund size and expense ratio. Business conditions such as competition play no role in determining capital allocation beyond their effect on fund size.*

3.2 Asymmetric Information

Suppose that managers observe $x_{i,t}$, but investors do not. Investors allocate funds as if $g(x_{i,t}) = 1$,

$$\frac{A^*_{i,t}}{q^*_{i,t}} = \frac{2f}{\mu_{i,t}}. \tag{3.6}$$

The equilibrium relation between the share under active management, fund size, and expense ratio now contains an additional term

$$2\ln(A^*_{i,t}/q^*_{i,t}) = \ln(f) - \ln(c_t) - \ln(q^*_{i,t}/M_t) + 2\ln(g(x_{i,t})) \tag{3.7}$$

This gives an alternative hypothesis.

Hypothesis 2 (Asymmetric Information): *If managers have superior information about* *the fund's time-varying investment opportunities relative to outside investors, the share of assets under active management will be positively related to variation in the profitability of opportunities. Business conditions such as competition play a role in determining capital allocation even conditional on fund size and fees.*

Note that under asymmetric information, net alpha is equal to $f_{i,t}(g(x_{i,t})^2 - 1)$. If managers are better informed of investment opportunities than outside investors, we would expect fund to make more when they take more active positions. In the cross-section, conditional on size, we would expect more active funds to perform better, potentially rationalizing findings that variables such as active share or industry concentration predict returns (Cremers and Petajisto, 2009; Kacperczyk et al., 2005).

3.3 Similar Hypotheses Based on an Alternative Approach

I develop hypotheses 1 and 2 based on Berk and Green (2004) and its particular assumptions, including fixed expense ratios and a particular cost structure. A different approach based on Pástor et al. (2017b) yields similar implications without assuming fixed expense ratios, and with the additional feature of multi-dimensional, micro-founded trading costs.

Pástor et al. (2017b) derive from first principles that larger funds that trade more and hold less liquid portfolios incur higher trading costs.[2] Specifically, trading costs are quadratic in $TL^{-1/2}$, the ratio of turnover T to the (square root of) portfolio liquidity L.[3] In their model, funds trade off the costs and benefits of higher turnover and lower portfolio liquidity. The assumption is that funds can exploit a greater number of opportunities by trading more; conversely, they can increase alpha by holding less liquid portfolios, which allows funds to focus on their best ideas in the most mispriced, illiquid segments of the market. The fund's

2. The key assumptions are that funds (expect to) turn over their portfolios proportionately, and incur trading costs for each stock that increase in the size of the trade relative to the stock's market capitalization.

3. More flexible functional forms can also be considered. See the appendix of Pástor et al. (2017b) for more details.

first-order condition, given fund size and trading opportunities, is

$$(T_{i,t}L_{i,t}^{-\frac{1}{2}})^* = \frac{\mu_{i,t}g(x_{i,t})}{2c_t(q_{i,t}/M_t)}.7 \tag{3.8}$$

Under symmetric information, the market clearing zero net alpha condition implies the same fund size $\frac{q^*_{i,t}}{M_t} = \frac{(\mu_{i,t}g(x_{i,t}))^2}{4c_t f_{i,t}}$ as before. In perfect capital markets with symmetrically informed outside investors, equilibrium turnover-liquidity ratio is negatively related to profit opportunities:

$$(T_{i,t}L_{i,t}^{-\frac{1}{2}})^* = \frac{2f_{i,t}}{\mu_{i,t}g(x_{i,t})}. \tag{3.9}$$

With symmetric information, we have the equilibrium relation

$$2\ln(TL^{-1/2})^* = \ln(f_{i,t}) - \ln(c_t) - \ln(q^*_{i,t}/M_t). \tag{3.10}$$

With asymmetric information, the information wedge influences internal capital allocation beyond its effect on fund size

$$2\ln(TL^{-1/2})^* = \ln(f_{i,t}) - \ln(c_t) - \ln(q^*_{i,t}/M_t) + 2\ln(g(x_{i,t})). \tag{3.11}$$

The approach based on the Pástor et al. (2017b) model reproduces hypotheses 1 and 2, with turnover to portfolio liquidity ratio $TL^{-1/2}$ taking the place of share of assets under active management. Ultimately, both the share of actively managed assets and the turnover to portfolio liquidity ratio measure the extent to which fund managers engage in active pursuit of profitable investment opportunities. An advantage of this formulation of the model is that portfolio liquidity is a multidimensional concept. Portfolio liquidity can be decomposed into a product of stock liquidity (market capitalization of holdings) and diversification, the latter of which can be further decomposed as a product of coverage (number of holdings relative to number of tradeable stocks) and balance (a measure of portfolio concentration). This framework allows the researcher to study each dimension, potentially allowing for a richer characterization of fund behavior.

CHAPTER 4

DATA

I build my dataset around two main sources. From the CRSP Survivor-Bias-Free US Mutual Fund database I obtain share class level information on returns, net asset values, expense ratios, TNA, fund turnover, first offer date, name, various fund objective classifications, and flags indicating index fund and ETF/ETN status. The CRSP Mutual Fund database includes data starting from January 1960. From the Thomson Reuters S12 database, I procure fund-level share holdings and additional information on fund investment objectives. Thomson's predecessor first compiled holdings data in March 1980, subsequent to which consistent holdings reports are available.[1] I supplement these two main sources by security-level data on prices and shares outstanding from CRSP, monthly return factors from Ken French's data library,[2] and active share (Cremers and Petajisto, 2009; Petajisto, 2013) from Antti Petajisto's website.[3]

TNA is typically only available at the quarterly or semi-annual frequency in the CRSP files before 1991 (Figure A.1). I interpolate missing TNA by assuming zero net flows. For up to one year following the most recent non-missing TNA value, I replace missing time $t+1$ values of TNA as $TNA_{t+1} = TNA_t(1 + r_{t+1})$, where r corresponds to net returns.

I link CRSP mutual fund data to Thomson holdings data using MFLINKS (Wermers, 2000; Cao and Xue, 2015), currently available until 2016.[4] Since CRSP data are at the share class level, at each date I aggregate variables to the portfolio level by taking the lagged

1. The 1980 March vintage includes a smattering of holding reports dated between 1979 December and 1980 February. For a detailed discussion of vintage dates vs report dates, see Appendix F. In the analysis, I only consider holdings reported during or after 1980 March.

2. http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

3. http://www.petajisto.net/data.html. This dataset also includes the identity of the benchmark against which active share is calculated.

4. Zhu (2017) shows that Thomson's coverage of new share classes deteriorates after 2008. In the Appendix, I replicate analyses using data up to 2008. The results remain similar.

TNA-weighted average of returns, expense ratio, turnover, and summing up TNA. Following Pástor et al. (2017a), I winsorize turnover at the 1% level. The final sample is a fund-month level panel spanning March 1980-December 2016.

4.1 Fund Selection

My aim is to study competition among long-only, general purpose actively managed U.S. domestic equity funds. I purge my sample of fixed income and "balanced" funds, money market funds, international funds, passive index funds, specialist long-short and sector funds, as well as target date funds. I use a variety of filters, based partially on previous research, and developed through a process of case-by-case inspection.[5] The filters primarily rely on a combination of various investment objective classifications, as well as exclusions based on fund names. I describe the fund selection algorithm in Appendix F in precise detail, and outline it below.

Since my analysis relies on within-fund variation, I construct filters at the fund level. I exclude all funds ever classified as International, Municipal Bonds, Bond & Preferred, Balanced, or Metals by Thomson investment objective codes. I exclude a fund if any of its share classes are ever assigned a policy code contrary to a long-only equity strategy,[6] assigned a CRSP objective code indicating sector fund or fixed income fund, flagged as an index fund, or have names indicative of index funds, target date funds, international funds, or tax managed funds. I exclude funds that are identified over 25% of the time as foreign equity by CRSP objective codes. This means that my dataset includes a handful of funds that transition to investing a portion of their assets in foreign markets.

5. The skeleton of my filtering algorithm is the scheme described in Kacperczyk et al. (2008). However, inspection of the fund universe resulting from my implementation of this scheme indicated a significant number of remaining international funds, sector funds, etc. This observation led me to add a number of additional filters.

6. Including codes corresponding to the following classifications: Balanced, Bonds & Preferred Stock, Bonds, Canadian & International, Leverage and/or Short Selling, Leases, Government Securities, Money Market, Preferred Stock, Sector/Highly Speculative, and various Tax Free.

In addition to the exclusion screens, I use objective codes to constructively identify actively managed domestic equity funds. I first use Lipper Class, including funds if any of their share classes are ever assigned a classification consistent with a domestic equity strategy.[7] If Lipper Class is not available, I consider Strategic Insights Objective Codes,[8] and if neither Lipper Class nor Strategic Insights Objective Codes are available, then Weisenberger Objective Codes.[9] I exclude fund-month observations with expense ratio below 0.1% as these are unlikely to correspond to active funds. To lessen the impact of incubation bias (Evans, 2000), I drop fund-month observations with lagged TNA below $15m in 2017 dollars.

4.2 Portfolio Weights

Although Thomson compiles updates on portfolio holdings at regular quarterly intervals, these updates do not exclusively consist of quarter-end reports of fund holdings. As shown in Figure A.2, a significant proportion of reports are dated outside of quarter-end months.

I index each fund i's most recent reporting period at month t as t_r^i, yielding a many-to-one mapping from month t to report date t_r^i for each fund. I consider portfolio holdings as stale beyond six months. Therefore, there are at most six distinct values of t that correspond to each t_r^i. Let Q_{h,i,t_r^i} denote the number of split adjusted shares of security h held by fund i at reporting date t_r^i, $P_{h,t}$ the split adjusted price of security h at month t, and θ_{i,t_r^i} the set of securities classified as U.S. common equity by CRSP in fund i's portfolio reported at t_r^i.

7. Included classes are: Equity Income Funds, Growth Funds, Large-Cap Core Funds, Large-Cap Growth Funds, Large-Cap Value Funds, Mid-Cap Core Funds, Mid-Cap Growth Funds, Mid-Cap Value Funds, Multi-Cap Core Funds, Multi-Cap Growth Funds, Multi-Cap Value Funds, Small-Cap Core Funds, Small-Cap Growth Funds, and Small-Cap Value Funds.

8. Included codes correspond to Equity USA Aggressive Growth, Equity USA Midcaps, Equity USA Growth & Income, Equity USA Growth, Equity USA Income & Growth, or Equity USA Small Companies.

9. Included codes correspond to Growth, Growth-Income, Growth and Current Income, Long-Term Growth, Maximum Capital Gains, or Small Capitalization Growth.

I define the weight of security h in fund i's portfolio at time t as

$$w_{h,i,t} = \frac{Q_{h,i,t_r^i} P_{h,t}}{\sum_{h \in \theta_{i,t_r^i}} Q_{h,i,t_r^i} P_{h,t}}. \tag{4.1}$$

Stacking the portfolio weights for each fund, denote the vector of portfolio weights by $\mathbf{w}_{i,t}$.

4.3 *CompetitorSize*

For each fund, I calculate *CompetitorSize* as the sum of all other funds' size, weighted by the cosine similarity between the funds' stock capitalization adjusted portfolio weights. I cap adjust portfolio weights, as cross-holding a given security is more informative about fund similarity when the market capitalization of the cross-held security is small (Cohen et al., 2005). I define capitalization adjusted weights as portfolio weights scaled by the inverse of the security's weight in the market portfolio:

$$\tilde{w}_{h,i,t} = \frac{w_{h,i,t}}{w_{h,m,t}}, \tag{4.2}$$

where $w_{h,m,t}$ is the weight in the market portfolio. I stack adjusted weights into vectors, denoted $\tilde{\mathbf{w}}_{i,t}$.

Define similarity weights $\psi^k_{i,j,t}$ for fund i with respect to fund j as the cosine similarity between their vectors of capitalization adjusted portfolio weights:[10]

$$\psi_{i,j,t} = \frac{\tilde{\mathbf{w}}_{i,t} \cdot \tilde{\mathbf{w}}_{j,t}}{\|\tilde{\mathbf{w}}_{i,t}\| \|\tilde{\mathbf{w}}_{j,t}\|}. \tag{4.3}$$

CompetitorSize is the similarity-weighted size of all other funds in the industry as of the

10. Cosine similarity represents the cosine of the angle between the funds' adjusted portfolio weight vectors. It is widely used in machine learning, and is used in finance academia with increasing frequency. For example, both Blocher (2016) and Hoberg et al. (2018) use cosine similarity of holdings to measure fund similarity. Cohen et al. (2016) use cosine similarity to measure similarity between company 10-K and 10-Q filings.

fund's most recent reporting date:

$$CompetitorSize_{i,t} = \sum_{j \neq i} \psi_{i,j,t_r^i} FundSize_{j,t_r^i}, \tag{4.4}$$

where

$$FundSize_{j,t_r^i} = \frac{TNA_{j,t_r^i}}{TotalMktCap_{t_r^i}}, \tag{4.5}$$

with $TotalMktCap$ representing the total market capitalization of all U.S. domestic equity in the CRSP universe. $CompetitorSize_{i,t}$ is invariant between each fund's reporting dates.[11] As a reference point, I calculate the total size of the actively managed mutual fund industry following Pástor et al. (2015) as $IndustrySize = \sum_j FundSize_{j,t}$.

4.4 Portfolio Liquidity

I calculate portfolio liquidity and its components (stock liquidity, diversification, coverage, and balance) according to Pástor et al. (2017b), constructing them with respect to the CRSP U.S. domestic equity universe.

4.5 Summary Statistics

Since my analysis relies on within-fund variation, I require each fund to have at least twelve months of non-missing observations of both returns and $CompetitorSize$ to be included in the estimation sample. My sample runs from March 1980 to December 2016, and includes 2,553 distinct funds. Table B.1 reports summary statistics.

The time series of the cross-sectional average competitor size and aggregate industry size are closely related (Figure B.1). Figure 4.1 presents histograms of the distribution of

11. The results remain virtually unchanged if I allow the measure to reflect within report date changes in the implied buy-and-hold portfolio weights and the size of competing funds by calculating it as $\sum_{j \neq i} \psi_{i,j,t} FundSize_{j,t}$.

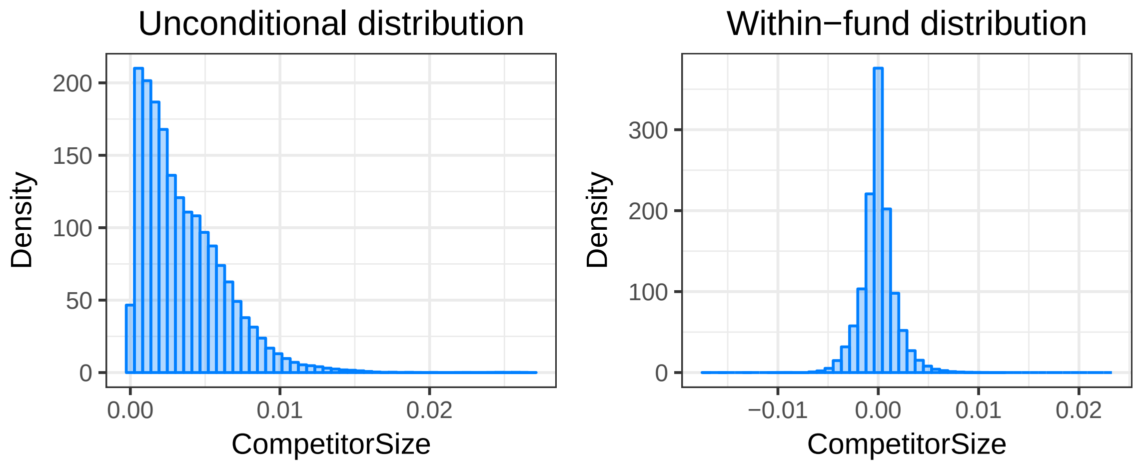

Figure 4.1. Histograms of *CompetitorSize*. The left panel illustrates the variable's unconditional distribution. The right panel shows the distribution after demeaning fund-by-fund.

CompetitorSize. The unconditional distribution is right skewed, as shown in the left panel. This is to be expected, as *CompetitorSize* is a weighted sum of the highly skewed *FundSize*. The distribution is less dispersed once the variable is demeaned fund-by-fund, but exhibits sufficient variation for meaningful within-fund analysis.

4.6 Correlations

Panel A of Table B.2 presents unconditional pairwise correlations between variables, while Panel B presents within-fund pairwise correlations. *CompetitorSize* is positively correlated with *IndustrySize* both unconditionally ($\rho = 0.38$) and at the fund level ($\rho = 0.60$). The residual within-fund variation in *CompetitorSize* with respect to *IndustrySize* reflects heterogenous dynamics in competitor size across funds pursuing different investment strategies.[12]

There is a small but negative correlation between *CompetitorSize* and risk adjusted gross returns, both unconditionally ($\rho =$ -0.02) and within-fund ($\rho =$ -0.03). *CompetitorSize* tends

12. This residual variation is useful for identification, as the time series correlation between *IndustrySize* and a linear time trend is $\rho = 0.92$, making industry level decreasing returns to scale hard to distinguish from simple trends in the data.

to increase over each fund's lifetime. The within-fund correlation between $CompetitorSize$ and $FundAge$ is $\rho = 0.51$. The unconditional correlation is markedly lower ($\rho = 0.10$), indicating that new funds begin their operations exploiting relatively lightly contested investment opportunities. $CompetitorSize$ is highly correlated with portfolio liquidity both unconditionally ($\rho = 0.75$) and within-fund ($\rho = 0.63$), evidence that more liquid market segments are capable of absorbing higher levels of active investment. Consistent with the joint determination of fund size, portfolio liquidity, turnover, and expense ratios in Pástor et al. (2017b), larger funds tend to be more liquid, trade less, and charge lower fees.

CHAPTER 5

CAPITAL ALLOCATION AND COMPETITOR SCALE

5.1 Decreasing Returns to Competitor Scale

I develop empirical tests of hypotheses 1 and 2 using competitor scale as the external constraint on the profitability of investment opportunities. This choice is motivated by both the extant literature and novel evidence from my sample. Pástor et al. (2015) give time series evidence that funds suffer from decreasing returns to aggregate industry scale. In recent work, Hoberg et al. (2018) provide cross-sectional evidence of decreasing profitability due to inter-fund competition. In Appendix C, I relate competitor scale to fund performance, using only within-fund variation. My results are consistent with decreasing returns to competitor scale. The following is a brief summary.

- A one standard deviation increase in competitor size is associated with a 76bp decrease in annual Fama-French three factor adjusted returns.
- Competitor size subsumes the negative effect of aggregate industry size in a head-to-head horse race.
- The negative impact of competitor size is smaller for funds which on average hold more liquid portfolios. This is consistent with liquidity constraints as the channel for decreasing returns to competitor scale.
- The negative impact of competitor size is smaller when funds tilt toward more liquid portfolios. This suggests that increased portfolio liquidity shelters funds from the pernicious effects of decreasing returns to scale.
- Holding skill fixed, funds make more when they hold less liquid portfolios. One interpretation is that funds increase portfolio concentration when they perceive favorable investment opportunities.

5.2 Empirical Strategy

Motivated by decreasing returns to competitor scale, I parametrize the profitability of a fund's time t investment opportunities as

$$g(x_{i,t}) = \mathit{CompetitorSize}_{i,t}^{-\gamma}. \tag{5.1}$$

This is a sensible choice in that alpha before transaction costs is a decreasing function of competitor scale, asymptoting to zero as the market approaches perfect competition.

The fraction of assets under active management in Berk and Green (2004) is similar in spirit to active share (denoted AS) from Cremers and Petajisto (2009), Petajisto (2013).[1] The Pástor et al. (2017b) portfolio choice maps directly into data on turnover and fund holdings. Therefore, letting $y_{i,t} \in \{AS_{i,t}, \left(TL^{-1/2}\right)_{i,t}\}$, the equilibrium relation (with potential information asymmetry) in equation (3.7) implies the regression model

$$\ln(y_{i,t}) = \alpha_t + \eta_1 \ln(\mathit{FundSize}_{i,t}) + \eta_2 \ln(f_{i,t}) + \gamma \ln(\mathit{CompetitorSize}_{i,t}) + \varepsilon_{i,t}. \tag{5.2}$$

Outcome y and *CompetitorSize* are both calculated based on the same portfolio weights. To ensure that my findings are not an artifact of measurement, I consider quarter end holding reports only, and calculate the log change in *CompetitorSize*, holding constant previous quarter-end similarity weights. That is, let t be quarter end dates, and define

$$\Delta CS_{i,t} = \ln\left(\sum_{j\neq i} \psi_{i,j,t-1} \mathit{FundSize}_{j,t}\right) - \ln\left(\sum_{j\neq i} \psi_{i,j,t-1} \mathit{FundSize}_{j,t-1}\right). \tag{5.3}$$

1. Cremers (2017) shows that for funds that do not short or use leverage, active share is equal to one minus the sum of holdings that overlap with the benchmark. Letting J be the set of stocks held by the fund and B the set of stocks in the benchmark b, we have

$$\mathit{ActiveShare} = 1 - \sum_{j \in J \cap B} \min\{w_{i,j}, w_{b,j}\}.$$

Note that the fund's change in capital allocation from $t-1$ to t has no mechanical effect on ΔCS, as it is determined only by changes in other fund size, holding similarity fixed.

I then estimate equation (5.2) in quarterly first differences as

$$\Delta \ln(y_{i,t}) = \alpha_t + \eta_1 \Delta \ln(FundSize_{i,t}) + \eta_2 \Delta \ln(f_{i,t}) + \gamma \Delta CS_{i,t} + \Delta \varepsilon_{i,t}. \tag{5.4}$$

I double cluster standard errors by fund and date × portfolio group.[2] Under the hypothesis of symmetrically informed outside investors, $\gamma = 0$. Under the joint hypothesis of decreasing returns to competitor scale and asymmetric information, $\gamma < 0$. I also perform the analysis separating out portfolio liquidity and its components. In these regressions, the joint hypothesis of decreasing returns to competitor scale and asymmetric information implies $\gamma > 0$.

5.3 Results

Table 5.1 presents results from empirical tests evaluating hypothesis 1 against hypothesis 2. The statistically significant coefficients on ΔCS provide a rejection of the hypothesis that managers and outside investors are symmetrically informed of investment opportunities captured by changes in the scale of competing funds.

The results are consistent with managers reacting optimally to decreasing returns to own and competitor scale by scaling back active management. A one percent increase in competitor size is associated with a 4bp decrease in active share, and a 52bp decrease in the turnover to portfolio liquidity ratio. Since turnover is reported on a fiscal year basis, the estimated negative relation between competitor scale and turnover-liquidity ratio is likely driven by portfolio liquidity: a one percent increase in competitor size is associated with a 79bp increase

2. Fund portfolios are grouped using k-means cluster analysis of raw portfolio weights. Each month, this process constructs $k = 10$ archetypal portfolios (serving as cluster centers). These model portfolios are constructed and then funds are assigned to each such that the sum of squared differences between the weights of fund portfolios and their assigned model portfolio is minimized.

Table 5.1
Capital Allocation and Competitor Scale

Observations are at the fund-quarter level, from 1980-2016. Δ denotes first differences. Dependent variables are noted in the column headers. AS is active share relative to self-declared benchmarks from Petajisto (2013), covering years 1980-2009. T is turnover; L, S, D, C, and B are respectively portfolio liquidity, stock liquidity, diversification, coverage, and balance, as defined in Pástor et al. (2017b). $\Delta CS_{i,t} = \ln\left(\sum_{j\neq i} \psi_{i,j,t-1} FundSize_{j,t}\right) - \ln\left(\sum_{j\neq i} \psi_{i,j,t-1} FundSize_{j,t-1}\right)$ is the change in log competitor size, holding previous quarter end similarity weights fixed. Standard errors are double clustered by fund and portfolio group × quarter, and reported in parentheses. Asterisks denote statistical significance: *** p <0.01, ** p <0.05, * p <0.1.

Dep. Var.:	$\Delta\ln(AS)$	$\Delta\ln(TL^{-1/2})$	$\Delta\ln(L)$	$\Delta\ln(S)$	$\Delta\ln(D)$	$\Delta\ln(C)$	$\Delta\ln(B)$
ΔCS	-0.045***	-0.521***	0.794***	0.624***	0.599***	0.179***	0.515***
	(0.016)	(0.056)	(0.079)	(0.077)	(0.064)	(0.029)	(0.058)
$\Delta\ln(FundSize)$	-0.016***	-0.111***	0.184***	0.101***	0.163***	0.098***	0.101***
	(0.002)	(0.012)	(0.013)	(0.010)	(0.011)	(0.007)	(0.009)
$\Delta\ln(f)$	-0.021	0.037	0.039*	-0.011	0.053**	0.029**	0.034*
	(0.017)	(0.027)	(0.021)	(0.019)	(0.022)	(0.012)	(0.019)
$\Delta\ln(T)$			-0.005	-0.004	-0.003	0.003	-0.006*
			(0.004)	(0.003)	(0.004)	(0.002)	(0.003)
$\Delta\ln(D)$				-0.355***			
				(0.019)			
$\Delta\ln(S)$					-0.632***	-0.194***	-0.539***
					(0.017)	(0.014)	(0.022)
$\Delta\ln(B)$						-0.126***	
						(0.011)	
$\Delta\ln(C)$							-0.293***
							(0.023)
Fixed Effects							
• Quarter	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	35,285	57,240	57,240	57,240	57,240	57,240	57,240
R^2	0.025	0.020	0.047	0.252	0.245	0.098	0.219
R^2 (proj. model)	0.004	0.007	0.031	0.233	0.231	0.077	0.209

in portfolio liquidity. Increases in competitor scale are also associated with statistically significant increases in each component of portfolio liquidity. As the scale of their competitors increase, funds tend to increase the stock liquidity and diversification of their portfolios, including coverage and balance.

Pástor et al. (2017b) argue that unit trading costs might vary by fund segment, implying a model with segment × time fixed effects. To accommodate segment level variation in trading costs, I repeat the analysis with benchmark × quarter fixed effects in Table D.2. The results remain similar. The results are also similar if, following Zhu (2017), I restrict the estimation data to pre-2008 observations (Tables D.1, D.3).

CHAPTER 6

EVIDENCE FROM THE 2003 MUTUAL FUND SCANDAL

In September 2003, the New York Attorney General's office launched investigations into several high-profile mutual fund families for illegal trading practices. Families were charged with allowing favored clients to trade fund shares at stale prices at the expense of ordinary shareholders (Houge and Wellman, 2005; Zitzewitz, 2006). By the end of October 2004, official investigations had been announced against a total of twenty-five mutual fund families.

Houge and Wellman (2005) and McCabe (2008) argue that investors penalized tainted funds with large outflows. This is borne out in my data. Figure 6.1 plots mean net flows by scandal involvement.[1]

The two series track each other closely in the two years prior to the scandal, and diverge abruptly in September 2003. The wedge between the two groups persists until the end of 2006, coincident with the final settlements negotiated with the Securities and Exchange Commission (Zitzewitz, 2009).[2]

I conclude that the scandal caused a significant reallocation of resources away from tainted

---

1. I follow Table 1 of Houge and Wellman (2005) for classifying fund families embroiled in the scandal. The following is the list of fund families tainted by the scandal by month of the news date of investigation. September 2003: Alliance Bernstein, Franklin Templeton, Gabelli, Janus, Nations, One Group, Putnam, Strong. October 2003: Alger, Federated. November 2003: Excelsior/US Trust, Fremont, Loomis Sayles, PBHG. December 2003: AIM/Invesco, MFS, Heartland. January 2004: Columbia, Scudder, Seligman. February 2004: PIMCO. March 2004: ING, RS. August 2004: Evergreen. October 2004: Sentinel.
I identify funds belonging to these families as of August 2003 in my sample based on the share class names in the CRSP mutual fund dataset. I classify 289 of the 1,461 funds in my sample in August 2003 with existing holdings and gross returns as members of future tainted families. Table E.1 presents a snapshot of summary statistics as of August 2003 by future scandal involvement. Tainted funds are slightly older, larger, and have higher turnover to portfolio liquidity and expense ratios.

2. The difference is statistically significant. Using a two year pre- and post-scandal window of observations, I estimate a regression of the form

$$flow_{i,t} = \alpha_i + \alpha_t + \gamma PostNews_{i,t} + \sum_{\tau=1}^{12} R_{i,t-\tau} + \varepsilon_{i,t},$$

where $PostNews_{i,t}$ is an indicator for tainted funds after news of their involvement in the scandal break. I find γ =-11.1% per year, with t-statistic of -5.2 (clustered by fund and portfolio group $\times$ month).

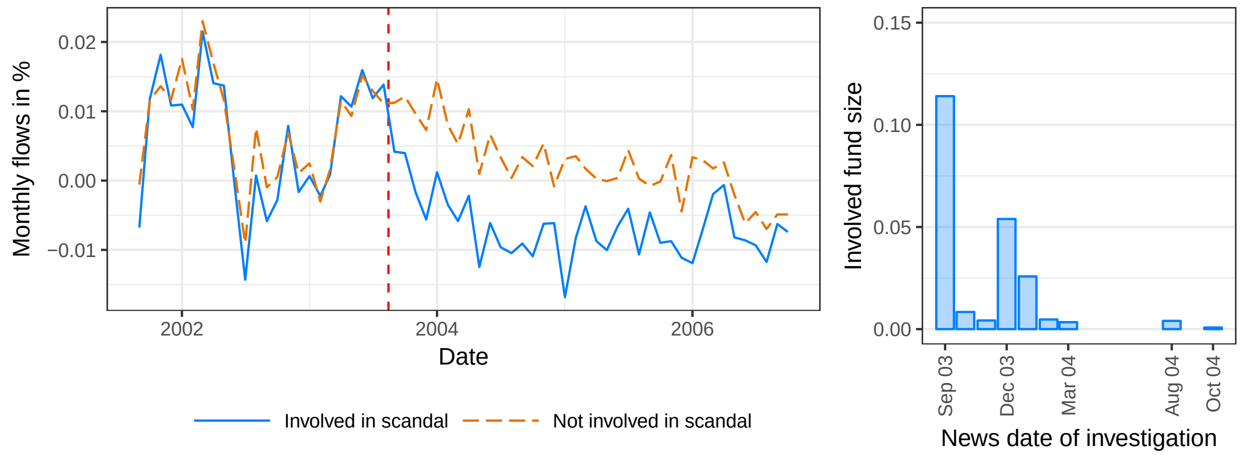

Figure 6.1. Flows and assets by scandal involvement. The left panel plots mean monthly net flows. The vertical line corresponds to August 2003, the month before the announcement of the first investigations. The right panel shows the total net assets of funds coming under investigation in a given month, as a fraction of the total net assets managed by all funds in my sample.

funds. Unless flows are perfectly offsetting, this shift causes a relative reduction in the competitor size of the most similar funds. Under decreasing returns to competitor scale, we would expect the investment opportunities of these funds to improve in relative terms, leading them to differentially expand active management and earn higher risk-adjusted returns.

I test these hypotheses by comparing untainted funds with differential pre-scandal similarity to prospective scandal funds. I discard tainted funds as the internal upheaval following the scandal likely had a direct impact on their performance and investment behavior.[3] I take two approaches. The first is a straightforward difference-in-differences-style comparison of fund outcomes before and after the scandal as a function of their pre-scandal exposure to tainted funds. The second approach links fund outcomes directly to variation in competitor size attributable to abnormal flows among tainted funds. I first present an analysis on fund capital allocation, followed by an analysis of fund performance.

3. In the aftermath of the investigations, several executives stepped down, and a number of portfolio managers were fired. Perhaps the highest profile casualty of the scandal was Richard S. Strong, founder of Strong Capital Management, who resigned in December 2003. Strong would go on to pay $60 million in settlements and be barred from the industry. Strong Capital itself was acquired by Wells Fargo in 2004.

6.1 Before and After Analysis

I relate fund-by-fund differences in pre-scandal $[2003m8 - W,\ 2003m8]$ and post-scandal $[2004m11,\ 2004m11 + W]$ outcomes to pre-scandal exposure to competition from tainted funds. I consider $W \in \{1, 2\}$ year windows. For a fund to be included in the estimation sample, it must have available holdings information for August 2003, and I must observe it both in the pre- and the post-scandal period.

I measure pre-scandal exposure as the proportion of competitor size attributable to prospective tainted funds as of August 2003. Let Φ denote the set of funds that belong to families later investigated, and define

$$ScandalExposure_i = \frac{\sum_{j \in \Phi} \psi_{i,j,2003m8} FundSize_{j,2003m8}}{\sum_{j \neq i} \psi_{i,j,2003m8} FundSize_{j,2003m8}}. \tag{6.1}$$

On average, 22% of untainted funds' competitor size is due to tainted fund families. Exposure ranges from 7% to 42%, with lower quartile 20% and upper quartile 26%.

To present interpretable summary statistics, I sort funds into high and low exposure groups depending on whether their $ScandalExposure$ is above or below the cross-sectional median. Table E.2 gives a snapshot taken in August 2003. High exposure funds are slightly smaller, have higher turnover to portfolio liquidity ratios, expense ratios, $CompetitorSize$, and worse performance. Fund age is almost identical across the two groups, limiting the plausibility of life cycle effects as an explanation for differences in outcome paths.

Figure 6.2 summarizes the identifying variation in the data. I plot the groupwise cross-sectional mean of within-fund deviations for log competitor size, log active share, log portfolio liquidity, and log turnover. The differential impact of the scandal across groups is identified by the difference in the pre- and post-scandal period wedges between the series. The $CompetitorSize$ of the low exposure group overall trends upward, despite a small dip in the middle of the scandal period. The $CompetitorSize$ of high exposure funds drops more

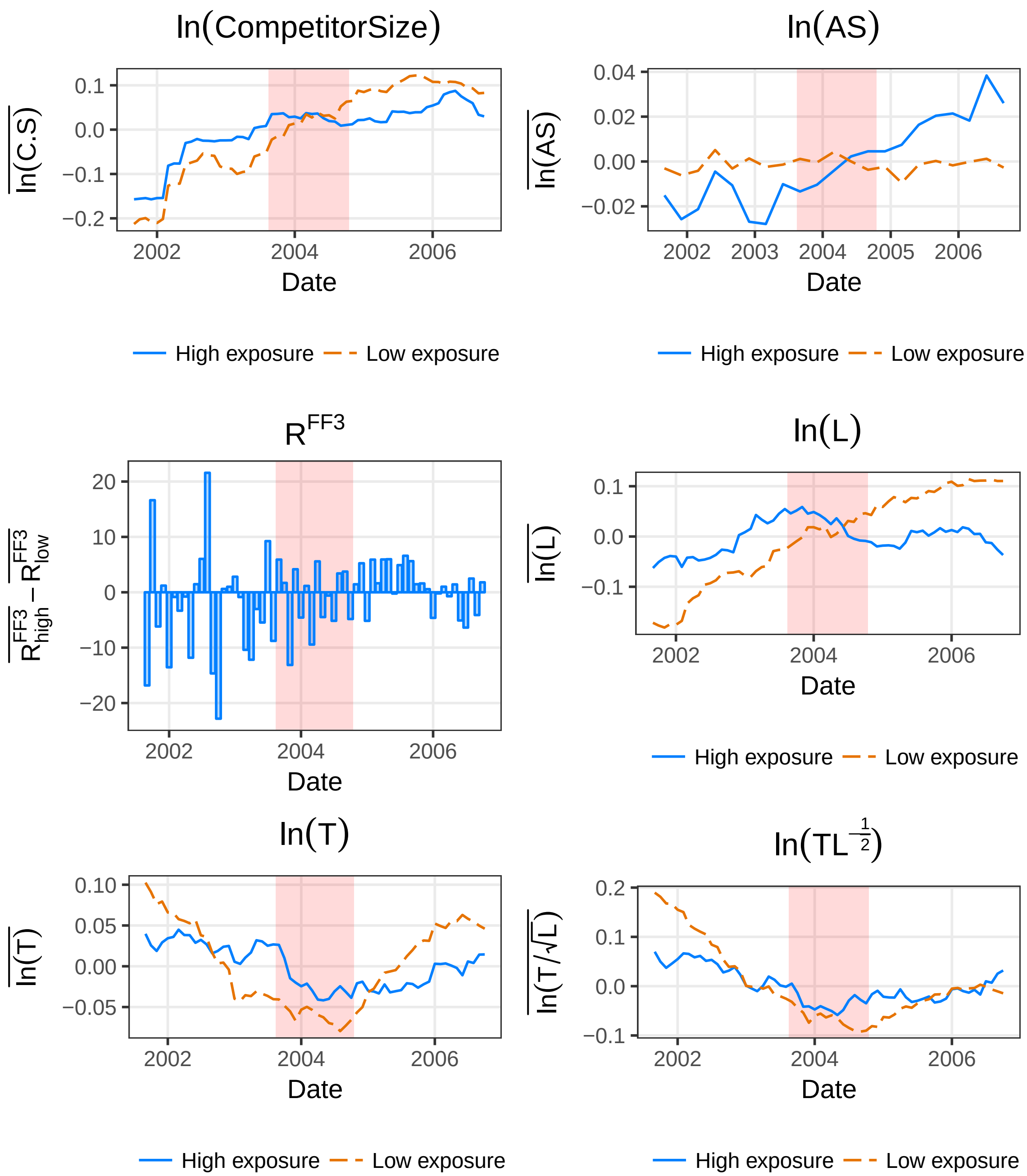

Figure 6.2. Untainted fund outcomes by *ScandalExposure*. Funds are sorted into high and low exposure groups depending on whether their *ScandalExposure* is above or below the cross-sectional median. The R^{FF3} panel plots the difference between the cross-sectional means of the within-fund deviations of three factor adjusted gross returns across high and low exposure groups. Other panels plot the cross-sectional groupwise means of respective variables' deviations from within-fund means. The $ln(AS)$ panel plots only quarter-end months as the variable is seldom reported within quarter. The shaded area corresponds to the scandal period Sep 2003-Oct 2004.

substantively during the scandal, and remains flat for almost a year after the end of the scandal period. The historical accident of scandal-related outflows at involved funds appear to have insulated their closest competitors from contemporaneous increases in the aggregate size of the industry.

The active share of low exposure funds is essentially flat during this period, whereas the active share of high exposure funds is flat in the pre-period, and then increases steadily during and after the scandal. The portfolio liquidity of high and low exposure funds exhibit broadly parallel increases in the pre-period. Following the scandal, the portfolio liquidity of low exposure funds continues to increase, whereas the portfolio liquidity of high exposure funds decreases during the scandal period and then levels off. These phenomena are consistent with high exposure funds responding to improved prospects by shifting resources away from the benchmark, tilting toward less liquid, more concentrated portfolios.

The patterns in turnover do not lend themselves to easy interpretation. The turnover of low exposure funds trends downward in the first half of the sample, and then swings upward during the second half, whereas the turnover of high exposure funds remains relatively flat, with an upward blip during the year ending in Sep 2003. Note that funds only report turnover once a year, as a cumulative measure that applies for the most recently concluded fiscal year. This is in contrast to holdings-based measures such as active share or portfolio liquidity, which can typically be calculated quarterly, based on unambiguously timed snapshots of portfolio holdings. The poor measurement of turnover's timing makes it a less suitable outcome variable for this analysis, which is designed to exploit tightly timed differences in fund outcomes as a function of exposure to competition by tainted funds.

Returns are highly volatile, which presents a challenge for providing a visual comparison of trends across groups. To compare relative fund performance before and after the scandal, for each month I plot the difference between high and low exposure groups' mean within-fund three factor adjusted returns. High exposure funds relatively underperform low exposure funds in the pre-scandal period, are essentially even during the scandal, and enjoy a string of relative outperformance in the year after the end of the scandal period. The differential relative before and after performance of the two groups is consistent with decreasing returns to competitor scale.

To formally test for differential differences in before and after outcomes as a function of ex ante exposure to competition from prospective scandal funds, I perform regressions of the form

$$y_{i,t} = \alpha_i + \alpha_t + \gamma \left(\mathbb{I}_t \times ScandalExposure_i \right) + \mathbf{X}_{i,t}\Gamma + \varepsilon_{i,t}, \tag{6.2}$$

where $\mathbf{X}_{i,t}$ includes log fund size and expense ratio, as dictated by theory. In the regression, exposure is a continuous variable. I double cluster standard errors by fund and portfolio group × time. I normalize *ScandalExposure* by its interquartile range ($\approx 6\%$).

Table 6.1 presents results. The one (two) year window estimate implies a statistically significant 6.4% (3.4%) post-scandal reduction in *CompetitorSize* for untainted funds at the 75^{th} percentile of *ScandalExposure* relative to untainted funds at the 25^{th} percentile of *ScandalExposure*. The same difference in *ScandalExposure* is associated with a statistically significant 2.6% (3.4%) relative increase in active share. The increase in turnover-liquidity ratio is positive but not statistically significant (1.9% at the one year horizon and 1.0% at the two year horizon). The weak response is due to turnover: increasing *ScandalExposure* from its 25^{th} to its 75^{th} percentile is associated with a highly significant 10.1% (10.1%) decrease in portfolio liquidity at the one (two) year horizon. Examining each component of portfolio liquidity separately reveals a shift toward lower portfolio liquidity among high exposure funds on all dimensions, as evidenced by statistically significant, negative coefficients associated with $\mathbb{I} \times ScanEx$ for all outcomes except balance.

Table 6.1
Capital Allocation and the Scandal: Before and After Analysis

Dependent variables are identified in the column headers. $\ln(C.S.)$ is an abbreviation for $\ln(CompetitorSize)$. Observations are at the fund-report date level, including only funds not directly involved in the scandal over the period $\{(2003m8 - W, 2003m8], [2004m11, 2004m11 + W)\}$, where W corresponds to the number of years specified in the panel headers. $\mathbb{I} \times ScanEx$ is the interaction of $ScandalExposure$ (normalized by its interquartile range) and an indicator for the post-scandal period. Standard errors are double clustered by fund and portfolio group × date, and reported in parentheses. Asterisks denote statistical significance: *** p <0.01, ** p <0.05, * p <0.1.

Dep. Var.:	$\ln(C.S.)$	$\ln(AS)$	$\ln(TL^{-1/2})$	$\ln(L)$	$\ln(S)$	$\ln(D)$	$\ln(C)$	$\ln(B)$
Panel A: 1 year window								
$\mathbb{I} \times ScanEx$	-0.064***	0.026***	0.019	-0.101***	-0.095***	-0.049**	-0.026*	-0.026
	(0.019)	(0.006)	(0.033)	(0.024)	(0.017)	(0.022)	(0.014)	(0.017)
$\ln(FundSize)$	0.130***	-0.013**	-0.163***	0.171***	0.083***	0.139***	0.072***	0.077***
	(0.021)	(0.007)	(0.032)	(0.026)	(0.014)	(0.025)	(0.020)	(0.018)
$\ln(f)$	-0.004	-0.015	0.073	0.015	-0.067	0.065	0.096	-0.024
	(0.098)	(0.029)	(0.141)	(0.115)	(0.076)	(0.106)	(0.076)	(0.077)
$\ln(T)$				-0.079***	-0.065***	-0.045*	0.017	-0.064***
				(0.027)	(0.016)	(0.025)	(0.015)	(0.019)
$\ln(D)$					-0.228***			
					(0.029)			
$\ln(S)$						-0.450***	-0.243***	-0.238***
						(0.050)	(0.041)	(0.050)
$\ln(B)$							-0.056*	
							(0.034)	
$\ln(C)$								-0.083*
								(0.048)
Fixed Effects								
• Fund	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
• Time	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	7,073	6,067	6,893	6,893	6,893	6,893	6,893	6,893
R^2	0.927	0.914	0.896	0.962	0.987	0.941	0.942	0.892
R^2 (proj. model)	0.052	0.022	0.023	0.076	0.156	0.121	0.083	0.062
Panel B: 2 year window								
$\mathbb{I} \times ScanEx$	-0.034*	0.029***	0.010	-0.101***	-0.105***	-0.040*	-0.032**	-0.011
	(0.020)	(0.006)	(0.033)	(0.025)	(0.018)	(0.022)	(0.015)	(0.016)
$\ln(FundSize)$	0.148***	-0.018***	-0.200***	0.183***	0.088***	0.147***	0.073***	0.084***
	(0.017)	(0.006)	(0.026)	(0.022)	(0.014)	(0.021)	(0.016)	(0.015)
$\ln(f)$	-0.098	0.005	0.171	-0.074	-0.169**	0.039	0.104	-0.061
	(0.073)	(0.027)	(0.126)	(0.109)	(0.077)	(0.091)	(0.072)	(0.072)
$\ln(T)$				-0.063***	-0.065***	-0.025	0.023	-0.049***
				(0.024)	(0.014)	(0.022)	(0.014)	(0.015)
$\ln(D)$					-0.216***			
					(0.022)			
$\ln(S)$						-0.425***	-0.217***	-0.236***
						(0.036)	(0.033)	(0.040)
$\ln(B)$							-0.055**	
							(0.024)	
$\ln(C)$								-0.077**
								(0.033)
Fixed Effects								
• Fund	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
• Time	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	13,656	11,764	13,291	13,291	13,291	13,291	13,291	13,291
R^2	0.895	0.878	0.858	0.943	0.980	0.913	0.914	0.848
R^2 (proj. model)	0.065	0.027	0.044	0.083	0.154	0.114	0.070	0.063

The main concern with identification based on comparing pre- and post-event periods across groups is that the measured effect might be the manifestation of favorable trends across the groups in the pre-period. I test for differential trends in the pre-period as a function of $ScandalExposure$ by estimating the regression

$$y_{i,t} = \alpha_i + \alpha_t + \gamma \left(t \times ScandalExposure_i \right) + \mathbf{X}_{i,t}\Gamma + \varepsilon_{i,t}, \tag{6.3}$$

where t is a linear time trend and $\mathbf{X}_{i,t}$ includes the usual controls. I estimate this regression on pre-period observations. Differential pre-trends by $ScandalExposure$ would be a concern if the coefficient on the trend interaction was statistically significant and of the same sign as the corresponding interaction coefficient in Table 6.1. Results from these specifications fail to reject the null hypothesis of no differential trends in the pre-period (Table E.3), with the exception of a slight favorable trend in the portfolio-liquidity ratio at the two year horizon due to the patterns in turnover seen in Figure 6.2.

6.2 Linking *CompetitorSize* Directly to Abnormal Flows

The analysis above does not explicitly model untainted fund outcomes as a function of the relevant shock to competitor scale, namely, the abnormal outflows from competing tainted funds. I aim to fill this gap in the following. I first estimate outflows at tainted funds attributable to the scandal. In turn, I relate untainted fund outcomes to variation in competitor size explained by abnormal tainted competitor outflows.

I use a linear model to decompose variation in fund flows between the effects of the scandal and baseline variation. I pool tainted and untainted funds in the two year window surrounding the scandal period, consisting of observations from September 2001 to October 2006. Consider scandal funds as being from the same cohort d if news of investigation into their trading practices broke in month d. Denote the cohort of fund j as $j^{(d)}$. Let $\mathbb{I}_{t \geq j^{(d)}}$ be an indicator for post investigation months for fund j, and define $\mathbb{I}_{d,t}$ as cohort $\times$ time dummy

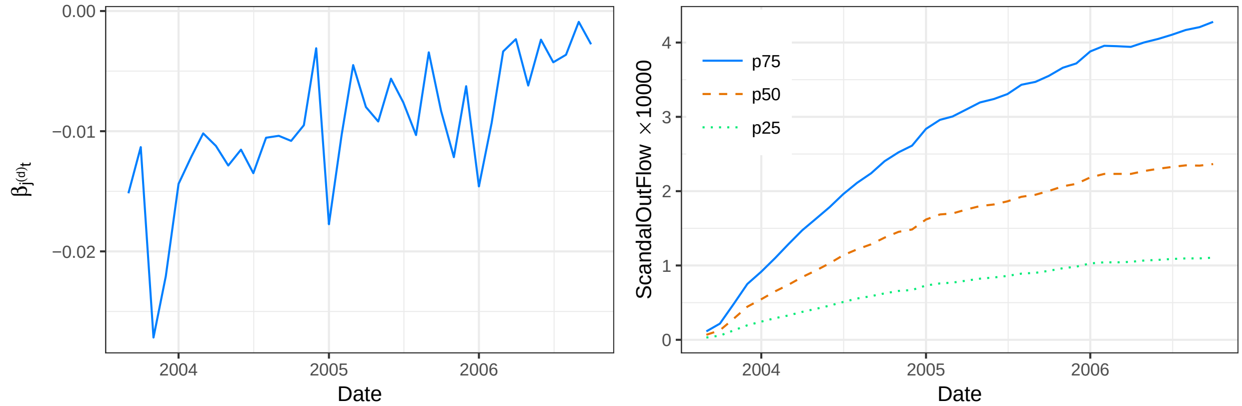

Figure 6.3. Estimated abnormal outflows from scandal funds. The left panel shows the cross-sectional mean coefficient on post-scandal cohort × time fixed effects from equation (6.4). The right panel shows the time series of cross-sectional percentiles of *ScandalOutFlow* across untainted funds.

variables. I regress flows on the full set of post-investigation cohort × time indicators, controlling for fund and time fixed effects:

$$flow_{j,t} = \alpha_j + \alpha_t + \beta_{j^{(d)},t} \left(\mathbb{I}_{t \geq j^{(d)}} \mathbb{I}_{j^{(d)},t} \right) + \varepsilon_{j,t}. \tag{6.4}$$

I interpret the betas as the path of abnormal flows attributable to the scandal for each cohort of tainted funds. I cumulate abnormal flows for each fund at each post-scandal date as

$$\hat{f}_{j,t} = \prod_{\tau \geq j^{(d)}}^{t} \left(1 + \hat{\beta}_{j^{(d)},t} \right) - 1. \tag{6.5}$$

I construct *ScandalOutFlow* for untainted fund i as the similarity- and size-weighted cumulative abnormal negative net flow among tainted funds $j \in \Phi$:

$$ScandalOutFlow_{i,t} = -\sum_{j \in \Phi} \psi_{i,j,2003m8} \left(\hat{f}_{j,t} FundSize_{j,2003m8} \right). \tag{6.6}$$

One can interpret *ScandalOutFlow* as the expected decrease in *CompetitorSize* for untainted funds due to scandal-related outflows among tainted funds, given the pattern of fund similarities immediately preceding the scandal.

Figure 6.3 plots time series characteristics of abnormal flows and $ScandalOutFlow$. Abnormal flows are most negative in the immediate aftermath of the announcement of the first investigations, and gradually converge to zero near the end of 2006. This pattern maps into almost linearly increasing cumulative outflows in the first two years after the scandal, reflected in the observed pattern in $ScandaOutFlow$. Importantly for identifying differential spillover effects of the scandal, total predicted outflows at competing tainted funds vary substantially in the cross-section.

This line of analysis at its core relies on differences in pre- and post-scandal outcomes among untainted funds as a function of post-scandal outflows among competing tainted funds. To illustrate the identifying variation, I sort funds into high and low outflow groups based whether their fund-level mean $ScandalOutFlow$ is above or below the median. I then plot cross-sectional mean within-fund demeaned outcomes for each group in Figure 6.4. The patterns are similar to Figure 6.2: the high outflow group exhibits a relative post-scandal decline in competitor size and portfolio liquidity, and an increase in active share. The two groups exhibit differential trends in turnover before 2003, but there is convergence before the scandal, and a relative increase in the turnover of the high group around the second half of 2004.

I formally test the link between tainted fund flows and untainted fund outcomes through the regression specification

$$y_{i,t} = \alpha_i + \alpha_t + \gamma ScandalOutFlow_{i,t} + \mathbf{X}_{i,t}\Gamma + \varepsilon_{i,t}, \tag{6.7}$$

where $\mathbf{X}_{i,t}$ includes log size and expense ratio. To make γ readily interpretable, I normalize $ScandalOutFlow$ by its interquartile range.

Table 6.2 presents results. Moving from the 25$^{\text{th}}$ to the 75$^{\text{th}}$ percentile of $ScandalOutFlow$ is associated with a 18.2% relative decline in competitor size using a one year event window, and 16.9% using a two year event window. These coefficients are three to five

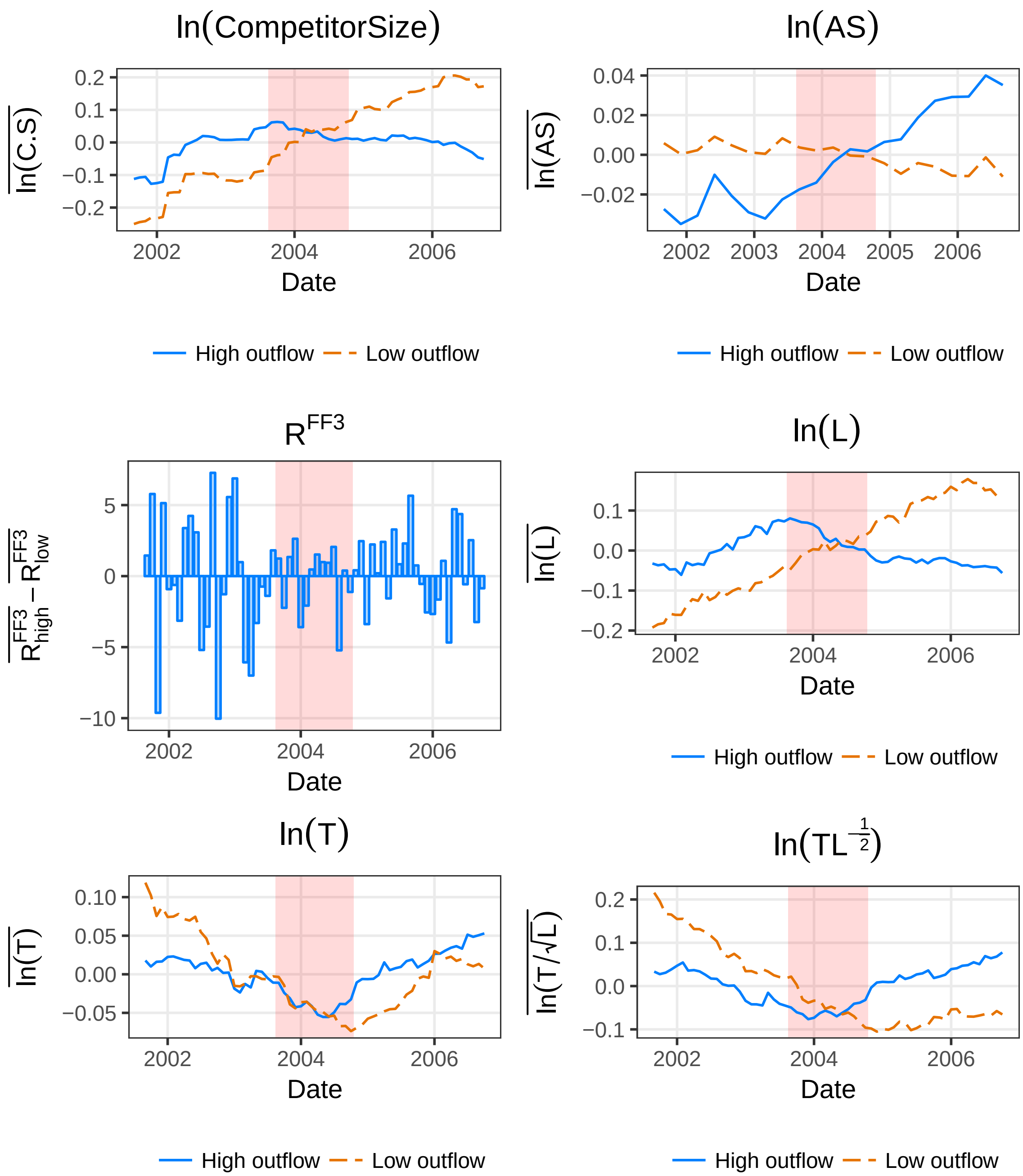

Figure 6.4. Untainted fund outcomes by mean *ScandalOutFlow*. Funds are sorted into high and low groups depending on whether their mean *ScandalOutflow* is above or below the cross-sectional median. The R^{FF3} panel plots the difference between the cross-sectional means of the within-fund deviations of three factor adjusted gross returns across high and low exposure groups. Other panels plot the cross-sectional groupwise means of respective variables' deviations from within-fund means. The $ln(AS)$ panel plots only quarter-end months as the variable is seldom reported within quarter. The shaded area corresponds to the scandal period Sep 2003-Oct 2004.

times the magnitude of the corresponding coefficients in Table 6.1, likely reflecting that $ScandalOutFlow$ is a closer proxy of the underlying quasi-exogenous shock of interest. All coefficients of interest are highly significant in this specification. An interquartile range increase in $ScandalOutFlow$ is associated at the one (two) year horizon with a 5.9% (6.1%) increase in active share, a 12.2% (11.6%) increase in the turnover-liquidity ratio, and a 22.2% (20.9%) decrease in portfolio liquidity. A closer look reveals a negative relation between $ScandalOutFlow$ and each dimension of portfolio liquidity. These results are consistent with outflows opening up investment opportunities for competing funds, to which they respond by increasing costly active management.

As in the previous section, I test for differential trends in the pre-period as a function of $ScandalOutFlow$ by estimating the regression

$$y_{i,t} = \alpha_i + \alpha_t + \gamma \left(t \times \overline{ScandalOutFlow_i}\right) + \mathbf{X}_{i,t}\Gamma + \varepsilon_{i,t}, \tag{6.8}$$

where $\overline{ScandalOutFlow_i}$ is fund-level mean $ScandalOutFlow$, t is a linear time trend, and $\mathbf{X}_{i,t}$ includes the usual controls. Results are presented in Table E.6. Two coefficients reach significance at the 10% level,[4] and all others are insignificant, meaning that the null hypothesis of no differential pre-trends is rarely (and even then only weakly) rejected.

4. The one-year coefficient for $CompetitorSize$ is positive, which leans against its negative association with $ScandalOutFlow$. On the other hand, the two-year coefficient for $\ln(TL^{-1/2})$ is also positive, which is a potential cause for concern. This slight pre-trend is unsurprising in light of the turnover pattern in Figure 6.4.

Table 6.2
Capital Allocation and the Scandal: Using Abnormal Flows

Dependent variables are identified in the column headers. ln($C.S.$) is an abbreviation for ln($CompetitorSize$). Observations are at the fund-report date level, including only funds not directly involved in the scandal over the period $\{[2003m9 - W, 2004m10+W]\}$, where W corresponds to the number of years specified. $ScandalOutFlow$ is normalized by its interquartile range. Standard errors are double clustered by fund and portfolio group × time, and reported in parentheses. Asterisks denote statistical significance: *** p <0.01, ** p <0.05, * p <0.1.

Dep. Var.:	ln($C.S.$)	ln(AS)	ln($TL^{-1/2}$)	ln(L)	ln(S)	ln(D)	ln(C)	ln(B)
Panel A: 1 year window								
$ScandalOutFlow$	-0.182***	0.059***	0.122***	-0.222***	-0.120***	-0.182***	-0.076***	-0.127***
	(0.017)	(0.007)	(0.026)	(0.023)	(0.015)	(0.021)	(0.015)	(0.015)
ln($FundSize$)	0.092***	-0.009*	-0.130***	0.142***	0.080***	0.115***	0.074***	0.056***
	(0.017)	(0.005)	(0.027)	(0.021)	(0.013)	(0.021)	(0.017)	(0.015)
ln(f)	-0.066	-0.008	0.068	-0.028	-0.061	0.008	0.077	-0.064
	(0.070)	(0.024)	(0.110)	(0.089)	(0.059)	(0.083)	(0.061)	(0.059)
ln(T)				-0.057***	-0.045***	-0.038**	0.011	-0.052***
				(0.019)	(0.011)	(0.018)	(0.011)	(0.014)
ln(D)					-0.267***			
					(0.025)			
ln(S)						-0.522***	-0.269***	-0.318***
						(0.044)	(0.035)	(0.046)
ln(B)							-0.097***	
							(0.029)	
ln(C)								-0.155***
								(0.043)
Fixed Effects								
• Fund	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
• Time	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	12,084	10,324	11,689	11,689	11,689	11,689	11,689	11,689
R^2	0.936	0.933	0.903	0.967	0.988	0.949	0.950	0.901
R^2 (proj. model)	0.088	0.073	0.025	0.107	0.172	0.174	0.107	0.103
Panel B: 2 year window								
$ScandalOutFlow$	-0.169***	0.061***	0.116***	-0.209***	-0.125***	-0.162***	-0.075***	-0.104***
	(0.015)	(0.007)	(0.024)	(0.021)	(0.014)	(0.019)	(0.015)	(0.013)
ln($FundSize$)	0.113***	-0.011**	-0.174***	0.155***	0.083***	0.127***	0.071***	0.070***
	(0.014)	(0.005)	(0.023)	(0.020)	(0.012)	(0.019)	(0.015)	(0.013)
ln(f)	-0.134**	0.012	0.183*	-0.110	-0.168**	-0.017	0.075	-0.090
	(0.059)	(0.023)	(0.102)	(0.096)	(0.071)	(0.080)	(0.061)	(0.064)
ln(T)				-0.049**	-0.050***	-0.024	0.017	-0.043***
				(0.019)	(0.011)	(0.018)	(0.012)	(0.013)
ln(D)					-0.259***			
					(0.020)			
ln(S)						-0.506***	-0.254***	-0.305***
						(0.035)	(0.031)	(0.039)
ln(B)							-0.088***	
							(0.023)	
ln(C)								-0.127***
								(0.032)
Fixed Effects								
• Fund	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
• Time	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	18,889	16,197	18,309	18,309	18,309	18,309	18,309	18,309
R^2	0.908	0.898	0.867	0.951	0.982	0.924	0.922	0.860
R^2 (proj. model)	0.103	0.086	0.046	0.117	0.181	0.164	0.095	0.096

In additional analyses I isolate the variation in $CompetitorSize$ attributable to abnormal flows at tainted competitors, and measure its impact on capital allocation. I perform two-stage least squares (2SLS) regressions, instrumenting for $\ln(CompetitorSize)$ by $ScandalOutFlow$ in the specification

$$y_{i,t} = \alpha_i + \alpha_t + \gamma \ln(CompetitorSize_{i,t}) + \mathbf{X}_{i,t}\Gamma + \varepsilon_{i,t}, \tag{6.9}$$

where $y_{i,t}$ is log active share, log turnover-liquidity ratio, or log portfolio liquidity and its components, and $\mathbf{X}$ the usual controls. Table E.9 presents results. As expected based on the first column of Table 6.2, the first stage F-statistics are high, and $ScandalOutFlow$ passes the relevance criterion. Consistent with the reduced form results, variation in competitor size attributable to $ScandalOutFlow$ is associated with decreased active management and increased portfolio liquidity.

6.3 Controlling for Sector Level Shocks

As an additional robustness check to ensure my results are not an artifact of common sector level shocks, I re-estimate the analysis using benchmark × time fixed effects. The results remain similar (Tables E.4, E.5, E.7, E.10).

6.4 Fund Performance

The analysis presented so far is consistent with competitors of funds tainted by scandal reacting to improved investment opportunities by increasing capital allocated to active strategies. According to this line of reasoning we would expect the same funds to experience relatively improved performance. To investigate, in Table 6.3 I perform analyses similar to those presented above, but with risk adjusted gross returns as the outcome variable of interest. The results demonstrate that close competitors of tainted funds indeed saw an increase in relative performance following the scandal, even after controlling for benchmark × month fixed effects. The inter-quartile difference in $ScandalExposure$ is associated with an increase in annualized three-factor benchmarked returns at the one (two) year horizon of 3.8% to 5.7% (1.8% to 2.5%), depending on specification. The corresponding difference in $ScandalOutFlow$ is associated with increases of 1.7% to 3.2% (0.4% to 1.1%). 2SLS specifications that instrument for $\ln(CompetitorSize)$ with $ScandalOutFlow$ indicate that the variation in competitor scale due purely to abnormal outflows at scandal-afflicted funds is negatively related to fund performance.

6.5 Investor Flows

I have argued that observing a relation between investment opportunities and funds' internal capital allocation after controlling for fund size is indicative of information asymmetry between managers and outside investors. If outside investors are less informed, we would expect their reaction to improvements in investment opportunities to lag fund managers' actions. To investigate, I regress net flows on either $\mathbb{I} \times ScandalExposure$ or $ScandalOutFlow$, along with a host of fund and month (or benchmark × month) fixed effects, using both one and two year estimation windows. In one set of regressions, I do not control for past returns. In the other, I control for one year of lagged three-factor adjusted excess returns.

Table E.11 presents results. In Panel A, I do not control for past performance. Consistent with sluggish investor response, high $ScandalExposure$ and $ScandalOutFlow$ do not have a positive association with investor flows at the one year horizon. At the two year horizon, a positive relation between $ScandalOutFlow$ and net flow transpires. However, this positive relation can be fully explained by backward-looking, return-chasing investor behavior: when I control for past performance in Panel B, the association vanishes. The two-year horizon appears to be long enough for the improved prospects of high $ScandalOutFlow$ funds to manifest in improved actual performance, to which investors respond by increasing the capital allocated to these funds.

Table 6.3
Fund Performance and the Scandal

The dependent variable is Fama-French 3 factor adjusted gross returns, in annual percent units. Observations are at the fund-month level. The estimation sample includes only funds not tainted by the scandal. In columns (1)-(4) regressions are estimated by ordinary least squares. In columns (5)-(6), regressions are estimated by two stage least squares, instrumenting $\ln(CompetitorSize)$ with $ScandalOutFlow$. In columns (1)-(2), the sample includes $\{(2003m8 - W, 2003m8], [2004m11, 2004m11 + W)\}$, where W corresponds to the number of years specified. In columns (3)-(6), the sample is taken over the period $\{[2003m9 - W, 2004m10 + W]\}$. $ScandalExposure$ (abbreviated to $ScanEx$) is the fraction of untainted funds' $CompetitorSize$ due to portfolio similarity with future scandal funds in August 2003. I normalize $ScandalExposure$ by its interquartile range. $\mathbb{I}$ is an indicator for the post scandal period. $ScandalOutFlow$ is the similarity-weighted cumulative abnormal outflows attributable to the scandal among involved funds. $ScandalOutFlow$ is normalized by its interquartile range. Benchmarks are the indexes which yield the lowest active share, taken from Petajisto (2013). I use the most recently available benchmark when one is missing. Standard errors are double clustered by fund and portfolio group × month in odd columns, and by fund and benchmark × month in even columns (5)-(6). Standard errors are reported in parentheses. Asterisks denote statistical significance: *** p <0.01, ** p <0.05, * p <0.1.

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: 1 year window						
$\mathbb{I} \times ScanEx$	5.723***	3.802***				
	(1.119)	(0.692)				
$ScandalOutFlow$			1.718**	3.213***		
			(0.856)	(0.733)		
$\ln(CompetitorSize)$					-9.703**	-16.121***
					(4.804)	(3.882)
$\ln(FundSize)$	-3.125***	-3.277***	-4.693***	-4.338***	-3.785***	-2.830***
	(0.683)	(0.511)	(0.652)	(0.499)	(0.754)	(0.615)
Fixed Effects						
• Fund	Yes	Yes	Yes	Yes	Yes	Yes
• Month	Yes	No	Yes	No	Yes	No
• Benchmark × Month	No	Yes	No	Yes	No	Yes
Observations	24,885	23,018	41,295	38,157	41,295	38,157
R^2	0.111	0.213	0.101	0.215		
R^2 (proj. model)	0.019	0.010	0.008	0.008		
F (first stage)					135.7	104.9
Panel B: 2 year window						
$\mathbb{I} \times ScanEx$	2.513***	1.827***				
	(0.834)	(0.478)				
$ScandalOutFlow$			0.415	1.126***		
			(0.541)	(0.409)		
$\ln(CompetitorSize)$					-2.539	-7.138***
					(3.303)	(2.665)
$\ln(FundSize)$	-3.710***	-3.697***	-4.161***	-3.965***	-3.865***	-3.121***
	(0.479)	(0.344)	(0.445)	(0.324)	(0.581)	(0.468)
Fixed Effects						
• Fund	Yes	Yes	Yes	Yes	Yes	Yes
• Month	Yes	No	Yes	No	Yes	No
• Benchmark × Month	No	Yes	No	Yes	No	Yes
Observations	48,194	44,675	65,413	60,555	65,413	60,555
R^2	0.083	0.217	0.087	0.220		
R^2 (proj. model)	0.011	0.009	0.009	0.009		
F (first stage)					130.9	77.9

CHAPTER 7

CONCLUSION

I studied the impact of competition on the investment decisions of active mutual funds. While there is a growing literature concerned with the deleterious effect of competition on fund performance, the link between competition and fund behavior has remained unexamined until now. I argued that the equilibrium response of active funds to competition depends on whether managers and outside investors share the same information set regarding the fund's prospective profitability. Therefore, studying the link between competition and fund behavior provides a window into whether managers and outside investors make disparate assessments of fund prospects.

I provided empirical evidence by studying the response of fund behavior to changes in the scale of competing funds. I found that funds decrease capital allocation to costly active strategies in response to increased competitor scale, which is consistent with fund managers being better informed about fund prospects than outside investors. I bolstered the causal interpretation of these findings by exploited the 2003 mutual fund scandal as a quasi-exogenous shock to the size of funds tainted by the scandal.

My study highlighted that fund performance is not only a function of skill, but also opportunity. Whether a fund can profitably deploy skill is shaped by several factors. The size and actions of other funds determines the amount of competition the fund faces for lucrative investments. The decisions of outside investors determines the size of the fund, and therefore its trading costs. Fund managers must take all of these into consideration when making their investment decisions. I have taken a reduced-form approach in linking competition to fund behavior, performance, and outside investor flows. Future research might find it fruitful to examine these elements in a unified structural framework.

APPENDIX A

RAW DATA CHARACTERISTICS

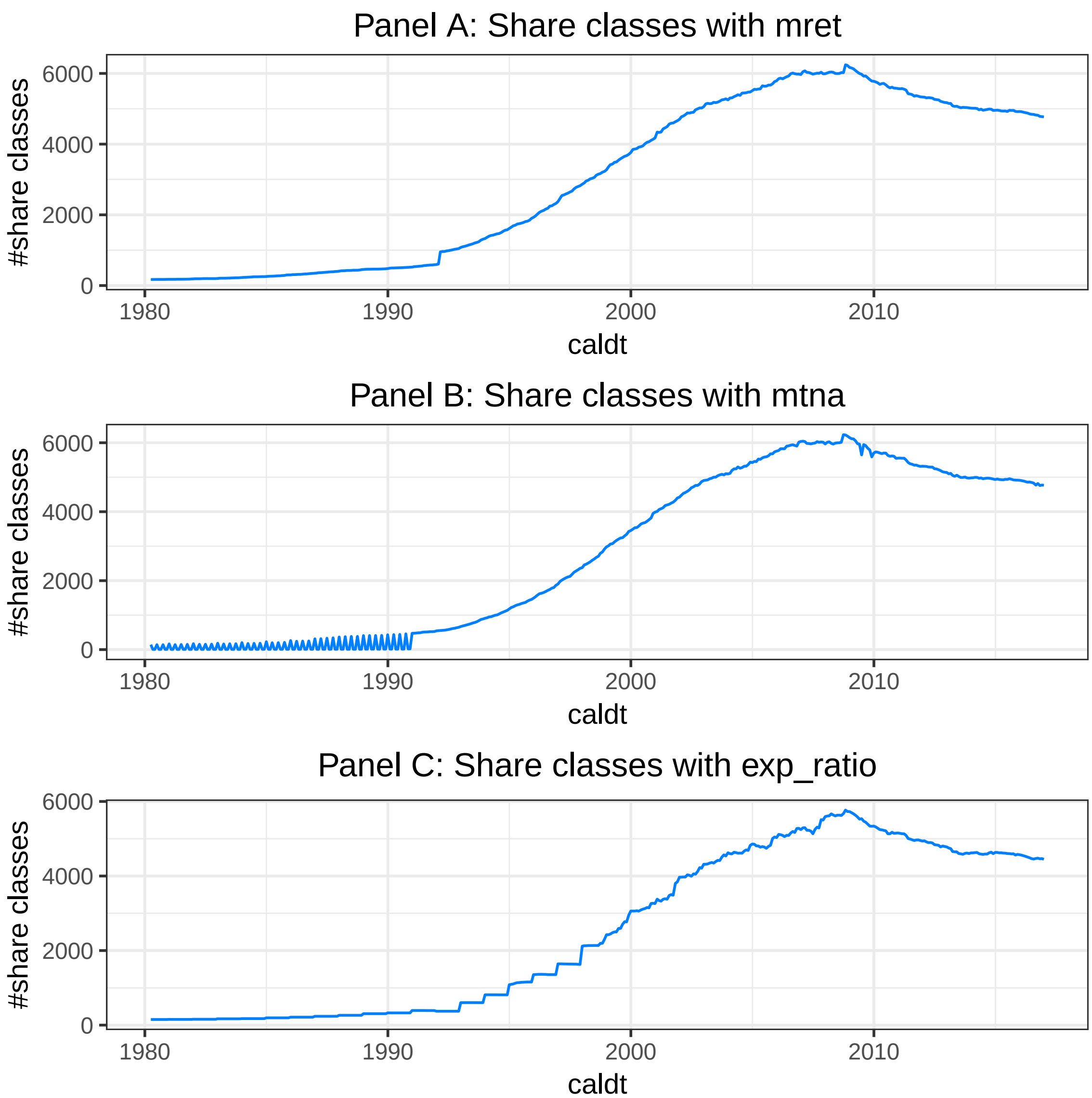

Figure A.1. Data availability in the CRSP Mutual Fund dataset. Number of actively managed domestic equity fund share classes with non-missing `mret`, `mtna`, and `exp_ratio` in the CRSP Mutual Fund Dataset by date.

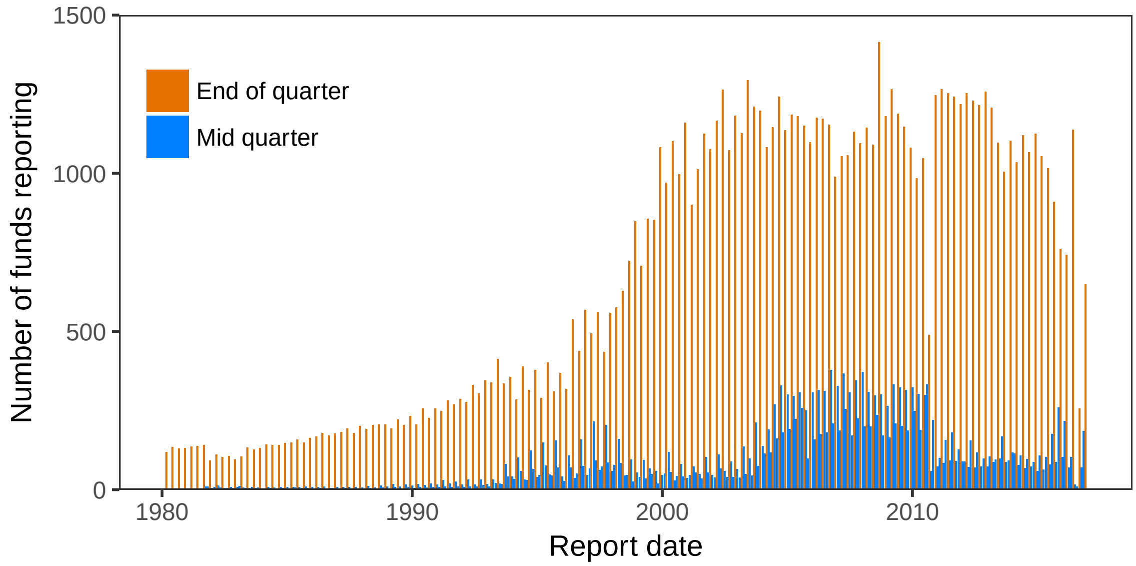

Figure A.2. Fund report dates in Thomson. Time series plot of the number of funds reporting portfolio holdings during a given month. End-of-quarter months (March, June, September, December) are differentiated from mid-quarter months by a distinct color.

APPENDIX B
SUMMARY STATISTICS

Table B.1
Summary Statistics

Alphas, returns, and expense ratios are expressed in annualized percentages. L, S, D, C, and B are portfolio liquidity, stock liquidity, diversification, coverage, and balance, respectively, as defined in Pástor et al. (2017b), calculated with respect to the market portfolio of U.S. common equity. $CompetitorSize$ is the portfolio similarity weighted size of each fund's competitors, as defined in Section 4.3. $IndustrySize$ is the total net assets of the funds in the sample, divided by the total market capitalization of all U.S. common equity in CRSP. $FundAge$ is the number of years since the fund's inception. $FundSize$ is the TNA of each fund as a fraction of the total market capitalization of all U.S. common equity in CRSP. AS is active share relative to self declared benchmarks (Cremers and Petajisto, 2009; Petajisto, 2013). T is turnover ratio as defined by CRSP, winsorized at 1%.

	N	mean	sd	p1	p10	p25	p50	p75	p90	p99
Panel A: Fund level means										
$\bar{R}^{FF3}$	2,393	-0.16	4.1	-12.96	-3.67	-1.2	0.26	1.58	2.81	6.82
$\bar{R}^{FF3}$ (net)	2,553	-1.35	4.05	-13.51	-4.8	-2.42	-0.92	0.36	1.62	5.48
$\overline{\text{Expense ratio}}$	2,413	1.29	0.42	0.39	0.83	1.01	1.25	1.51	1.84	2.48
$\bar{L} \times 10$	2,553	0.49	0.65	0.01	0.04	0.08	0.26	0.68	1.2	2.93
$\bar{S}$	2,553	10.26	9.57	0.15	0.43	1.3	8.99	16.51	22.46	38.41
$\bar{D} \times 10$	2,553	0.1	0.23	0	0.02	0.03	0.05	0.1	0.18	0.72
$\bar{C} \times 10$	2,553	0.23	0.44	0.04	0.07	0.09	0.14	0.22	0.4	1.47
$\bar{B}$	2,553	0.38	0.16	0.06	0.18	0.26	0.37	0.5	0.6	0.73
Panel B: Fund-month level statistics										
R^{FF3}	363,872	0.44	22.85	-61.52	-23.34	-10.54	0.3	11.15	24.05	65.82
R^{FF3} (net)	384,643	-0.78	22.71	-62.86	-24.5	-11.68	-0.87	9.89	22.68	64.42
Expense ratio	362,972	1.23	0.42	0.33	0.76	0.96	1.18	1.45	1.77	2.41
$CompetitorSize$	384,643	0.35	0.28	0.02	0.06	0.14	0.29	0.51	0.72	1.22
$IndustrySize$	384,643	0.14	0.03	0.03	0.09	0.13	0.14	0.16	0.17	0.17
$FundSize \times 10^4$	384,643	1.12	3.91	0.01	0.03	0.07	0.22	0.77	2.25	16.32
TNA (2017$100m)	384,643	16.91	64.2	0.18	0.41	0.98	3.2	11.08	31.83	248.46
$FundAge$	384,502	14.76	13.81	0.75	2.92	5.75	10.75	18.42	31.08	69.83
AS	63,036	0.81	0.15	0.38	0.6	0.72	0.85	0.93	0.97	1
$\ln(TL^{-1/2})$	342,879	1.38	1.18	-1.71	-0.1	0.63	1.4	2.19	2.86	3.96
T	342,879	0.8	0.67	0.03	0.18	0.34	0.62	1.05	1.63	3.73
$L \times 10$	384,643	0.48	0.67	0.01	0.03	0.07	0.23	0.64	1.22	3.17
S	384,643	9.92	9.65	0.13	0.41	1.23	8.29	15.75	22.44	40.03
$D \times 10$	384,643	0.09	0.23	0	0.01	0.02	0.05	0.1	0.18	0.7
$C \times 10$	384,643	0.22	0.41	0.03	0.06	0.09	0.14	0.22	0.37	1.58
B	384,643	0.37	0.18	0.04	0.14	0.23	0.36	0.51	0.63	0.78

Table B.2
Correlations

		Comp. Size	R^{FF3}	Exp. ratio	*Fund Size*	*Fund Age*	*AS*	*T*	*L*	$\ln(TL^{-1/2})$	*Ind. Size*
	Panel A: Unconditional										
1	*CompetitorSize*	1.00									
2	R^{FF3}	-0.02	1.00								
3	Expense ratio	-0.25	0.01	1.00							
4	*FundSize*	0.20	-0.00	-0.20	1.00						
5	*FundAge*	0.10	-0.01	-0.24	0.29	1.00					
6	*AS*	-0.69	-0.02	0.24	-0.14	-0.13	1.00				
7	*T*	-0.04	-0.00	0.20	-0.10	-0.11	0.06	1.00			
8	*L*	0.75	-0.01	-0.27	0.19	0.11	-0.83	-0.09	1.00		
9	$\ln(TL^{-1/2})$	-0.47	0.01	0.35	-0.21	-0.21	0.54	0.70	-0.53	1.00	
10	*IndustrySize*	0.38	-0.01	0.09	-0.05	-0.09	-0.17	0.02	0.11	-0.02	1.00
	Panel B: Within-fund correlations										
11	*CompetitorSize*	1.00									
12	R^{FF3}	-0.03	1.00								
13	Expense ratio	-0.04	0.03	1.00							
14	*FundSize*	0.26	-0.02	-0.06	1.00						
15	*FundAge*	0.51	-0.03	-0.16	0.14	1.00					
16	*AS*	-0.53	-0.00	-0.01	-0.16	-0.37	1.00				
17	*T*	-0.07	0.01	0.11	-0.07	-0.11	0.02	1.00			
18	*L*	0.63	-0.01	-0.04	0.21	0.24	-0.70	-0.08	1.00		
19	$\ln(TL^{-1/2})$	-0.34	0.02	0.16	-0.15	-0.29	0.33	0.73	-0.38	1.00	
20	*IndustrySize*	0.60	-0.01	0.01	0.17	0.76	-0.39	-0.01	0.26	-0.18	1.00

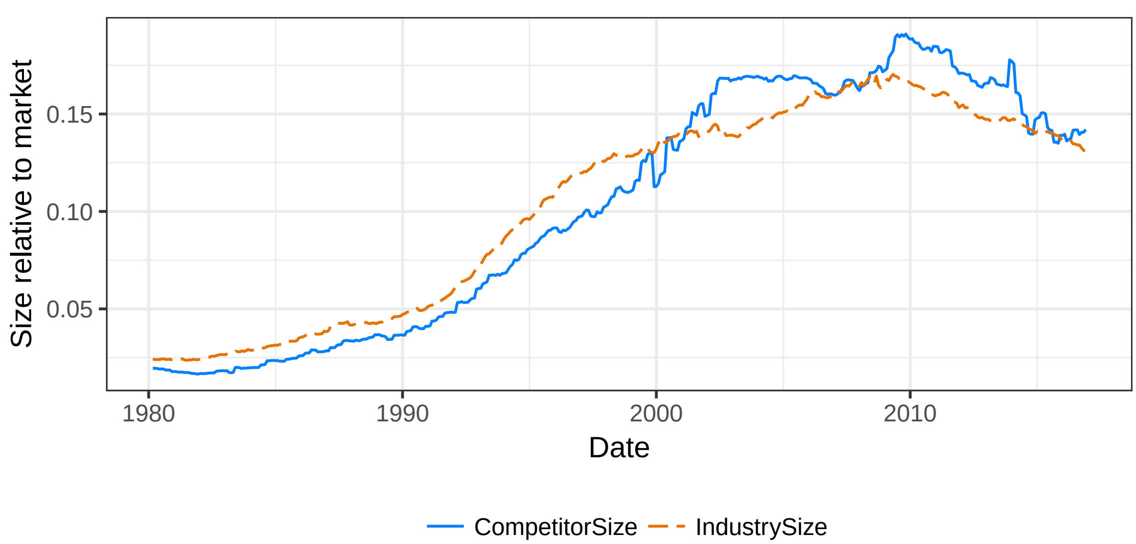

Figure B.1. Time series of *CompetitorSize*. Cross-sectional mean of *CompetitorSize* (scaled by 40 for exposition) against the time series of *IndustrySize*.

APPENDIX C
FUND PERFORMANCE AND COMPETITOR SCALE

C.1 Benchmarking Returns

Ideally, I would like to benchmark mutual fund returns by factors that are both near costlessly tradeable for funds, and span dimensions of risk that are of concern to investors. In the absence of such an ideal benchmark, I employ the conventional option of benchmarking returns with Fama-French factors. Define three factor benchmark adjusted gross returns as

$$R_{i,t}^{FF3} = R_{i,t} - \left[\hat{\beta}_i^{RMRF} RMRF_t + \hat{\beta}_i^{SMB} SMB_t + \hat{\beta}_i^{HML} HML_t\right], \tag{C.1}$$

where $R_{i,t}$ is the gross return of fund i at month t in excess of the risk free rate, expressed in percentages. $RMRF$, SMB, and HML are the usual market, size, and value factors. The beta hats are the sample estimates of each fund's exposure to the respective factors, estimated by fund level regressions of the form

$$R_{i,t} = \alpha_i + \beta_i^{RMRF} RMRF_t + \beta_i^{SMB} SMB_t + \beta_i^{HML} HML_t + \varepsilon_{i,t}. \tag{C.2}$$

Therefore, $R_{i,t}^{FF3} = \hat{\alpha}_i + \hat{\varepsilon}_{i,t}$, i.e. each period's benchmark adjusted returns are equal to the sum of the fund's estimated gross alpha and the given month's residual from the Fama-French time series regressions.

Benchmarking with a factor model has some shortcomings and advantages. The long-short SMB and HML portfolios are not tradeable for mutual funds. Berk and van Binsbergen (2015) argue that at each point in time the performance of active funds ought to be measured against the returns of the lowest cost passive funds readily available to retail investors. This is an eminently sensible suggestion for studying funds' value added for retail investors, but not an obviously superior method for testing whether fund alpha is decreasing in competitor scale. Cremers et al. (2012) note that Fama-French benchmarks imply nonzero alphas for a number of mainstream passive benchmarks. My results are robust to following their suggestion of benchmarking with index-based factors. Unlike self-designated benchmarks, Fama-French factors are not gameable by funds. They are also widely available.[1] Unlike characteristic based benchmarks, factor based benchmarks are not subject to errors in holdings data, are available monthly, and account for the unobserved actions of funds. Lastly, regardless of whether size and value correspond to risk, Fama-French factors capture a large fraction of variance in cross-sectional returns.

1. As argued by Pástor et al. (2015), Morningstar benchmarks share the feature of non-gameability, but are proprietary.

C.2 Regression Setup

I implement within-fund regression specifications of the following form to test for decreasing returns to competitor scale:

$$R^{FF3}_{i,t+1} = \alpha_i + \gamma CompetitorSize_{i,t} + \mathbf{X}_{i,t}\Gamma + \varepsilon_{i,t+1}, \tag{C.3}$$

where α_i are firm fixed effects, and $\mathbf{X}_{i,t}$ is a vector of controls including $IndustrySize$ and year-month fixed effects.[2] The coefficient of interest is γ. To make the economic magnitude of the coefficient easier to interpret, I annualize returns and divide $CompetitorSize$ and $IndustrySize$ by their respective standard deviations before performing the regressions. I re-scale $FundSize$ by the difference between the 50^{th} and 10^{th} percentiles of its distribution.

I include fund fixed effects throughout to take into account the possibility that baseline fund skill and average competitor size are related in the cross-section, i.e. $\text{Cov}(\alpha_i, \overline{CompetitorSize_i}) \neq 0$. We would expect talented managers to be endogenously allocated where they are most capable of taking advantage of investment opportunities. Bolstering this view, Berk et al. (2017) show that fund families funnel capital toward skilled managers. A mechanical concern is that in the cross-section $CompetitorSize$ tends to be higher for funds in large cap sectors. Since more liquid market segments can absorb a larger amount of active investment, not all cross-sectional variation of $CompetitorSize$ reflects variation in the effective inter-fund competition for investment opportunities. Controlling for fund fixed effects is a parsimonious way of controlling for such fixed differences in funds' operating environment.

For the specifications including fund fixed effects only, deviations of $CompetitorSize$ from its within-fund mean provide the variation identifying the coefficient of interest. For regressions including both fund and year-month fixed effects, the coefficient of interest is identified based on deviations of $CompetitorSize$ from its within-fund mean, relative to the average within-fund deviation at each date. Year-month fixed effects control nonparametrically for common time series variation in returns and industry competition, ruling out the possibility that the identified effect of competitor size is an artifact of other aggregate developments, such as shared time-varying exposure to competition from hedge funds. However, including overly fine cross-sectional dummy variables would risk soaking up the variation of interest.

I include $IndustrySize$ to demonstrate that $CompetitorSize$ captures distinct variation in decreasing returns faced by funds. Controlling for $FundSize$ is relevant for separating industry-level decreasing returns to scale from fund-level decreasing returns to scale.[3]

2. Since $IndustrySize$ only varies in the time series, it is omitted in regressions featuring year-month fixed effects. Similarly, $FundAge$ is fully absorbed by the combination of fund and year-month fixed effects.

3. Interpreting the coefficient on $FundSize$ is problematic in within-fund regressions of returns. To be unbiased, within-fund regressions require strict exogeneity of the regressors [Chamberlain (1982); stambaugh99], meaning $\text{Cov}(x_{i,t}, \varepsilon_{i,s}) = 0 \ \forall s \in \{1, 2, \dots, T_i\}$. Since past idiosyncratic high (low) returns mechanically increase (decrease) total net assets, we will typically have $\text{Cov}(FundSize_{i,t}, \varepsilon_{i,s}) > 0$ for $s < t$, and a downward bias in the estimated coefficient. In simulations, Harvey and Liu (2017) estimate the bias around 14%. Pástor et al. (2015) propose a recursive demeaning (RD) procedure for eliminating this bias.

For constructing standard errors, each month I sort funds into ten mutually exclusive (but not necessarily equal sized) portfolio groups based on their most recently reported holdings.[4] I double cluster standard errors by fund and year-month × portfolio group, to account for both within-fund and cross-sectional correlation in errors. In practice, I find that clustering by fund in regressions of returns is essentially irrelevant, as within-fund correlation in the error term is negligible. On the other hand, the cross-sectional correlation structure of regression errors is substantive. The number of portfolio groups is similar to the number of Morningstar sectors. Each month's largest portfolio group cluster on average accounts for over a third of observations. Therefore, portfolio group × month clusters allow for extensive within month correlation of errors, without reducing the number of clusters unreasonably.

C.3 Results

Table C.1 presents results from the equation (C.3) regression specifications. There is a consistently negative, statistically significant within fund relation between $CompetitorSize$ and fund performance. Coefficients range from -1.03 in the univariate within-fund regression to -0.76 in the specification featuring the full set of fund and year-month fixed effects and own size. While $IndustrySize$ is associated with a statistically significant -0.40 coefficient in the specification with no other controls (column (2)), adding $CompetitorSize$ to the specification (column (3)) subsumes its negative effect, with the coefficient on $IndustrySize$ dropping to an insignificant 0.08. Although coefficients associated with own fund size are consistently negative and statistically significant, they are known to be biased and should be interpreted with caution.[5]

The point estimates they report with the RD procedure are similar to those from the fixed effects OLS regressions, but the standard errors increase almost twenty-fold. Given that estimating the magnitude of decreasing returns to own size is not the focus of this study and the unfavorable tradeoff between bias and variance, I choose to not implement the RD procedure.

4. Funds are grouped using k-means cluster analysis of raw portfolio weights. Each month, this process constructs $k = 10$ archetypal portfolios (serving as cluster centers). These model portfolios are constructed and then funds are assigned to them such that the sum of squared differences between the weights of fund portfolios and their assigned model portfolio is minimized.

5. Consistent with Harvey and Liu (2017), I find much larger estimates of decreasing returns to own size when using a log transform of $FundSize$.

Table C.1
Fund Performance and Competitor Scale

Observations are at the fund-month level, over the period 1980-2016. The dependent variable is three-factor adjusted gross returns, in annualized percentages. *CompetitorSize* and *IndustrySize* are normalized by their respective sample standard deviations. *FundSize* is normalized by the difference between the 50th and 10th percentile of its distribution. Standard errors are double clustered by fund and year-month × portfolio group, and reported in parentheses. Asterisks denote statistical significance: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

	(1)	(2)	(3)	(4)
CompetitorSize	-0.973***		-1.033***	-0.763***
	(0.176)		(0.220)	(0.142)
IndustrySize		-0.400**	0.076	
		(0.183)	(0.229)	
FundSize				-0.024***
				(0.009)
Fixed Effects				
• Fund	Yes	Yes	Yes	Yes
• Month	No	No	No	Yes
Observations	363,872	363,872	363,872	363,872
R^2	0.012	0.012	0.012	0.103
R^2 (proj. model)	0.001	0.000	0.001	0.001

Expense ratios provide an informative comparison for the magnitude of the *CompetitorSize* coefficients. The mean expense ratio in my sample is 1.23% per year, with an interquartile range of 0.49%. A one standard deviation increase in *CompetitorSize* is associated with a drop in performance on the order of two thirds the typical fund expense ratio. Decreasing returns to competitor scale are a meaningful impediment to sustainable profitable operations for funds.

C.4 The Role of Portfolio Liquidity

Economic reasoning dictates that decreasing returns to competitor scale operate through the price impact of competing funds. Therefore, we would expect decreasing returns to be more severe for funds relying on less liquid strategies. I test this by comparing the magnitude of decreasing returns across funds with different levels of average portfolio liquidity. Specifically, I run regressions of the form

$$\begin{aligned} R^{FF3}_{i,t+1} = \alpha_i + \alpha_t + \gamma_1 CompetitorSize_{i,t} + \gamma_2 \left(CompetitorSize_{i,t} \times \overline{PortLiq_i} \right) \\ + \eta_1 FundSize_{i,t} + \eta_2 \left(FundSize_{i,t} \times \overline{PortLiq_i} \right) + \varepsilon_{i,t+1}, \end{aligned} \quad \text{(C.4)}$$

where $\overline{Port.Liq_i}$ is either the fund-level average portfolio liquidity measure proposed by Pástor et al. (2017b), or any of its sub-components of stock liquidity (average relative market capitalization of holdings), coverage (number of stocks held relative to total stocks in the market), and balance (a measure of how closely portfolio weights track market weights of stocks in the portfolio). Diversification is the product of coverage and balance. I re-scale each variable so that a unit increase corresponds to the interquartile range of within-fund means. If decreasing returns to scale are rooted in liquidity constraints, we expect $\gamma_2 > 0$.

Table C.2
The Role of Portfolio Liquidity

Observations are at the fund-month level, over the period 1980-2016. The dependent variable is three-factor adjusted gross returns, in annualized percentages. L, S, D, C, B are portfolio liquidity, stock liquidity, diversification, coverage, and balance, as defined in Pástor et al. (2017b). $\bar{L}$, $\bar{S}$, $\bar{D}$, $\bar{C}$, $\bar{B}$ denote fund-level means. Each X variable is normalized by its interquartile range. Standard errors are double clustered by fund and year-month $\times$ portfolio group, and reported in parentheses. Asterisks denote statistical significance: *** p <0.01, ** p <0.05, * p <0.1.

	(1)	(2)	(3)	(4)	(5)
Panel A: Fund Level Average Portfolio Liquidity					
$\bar{X}$ =	$\bar{L}$	$\bar{S}$	$\bar{D}$	$\bar{C}$	$\bar{B}$
$Comp.Size \times \bar{X}$	0.231***	0.320	0.048***	0.069***	0.300
	(0.075)	(0.282)	(0.017)	(0.020)	(0.192)
$FundSize \times \bar{X}$	0.028***	0.011	0.006**	0.004*	0.005
	(0.010)	(0.022)	(0.003)	(0.002)	(0.015)
$CompetitorSize$	-1.177***	-1.073***	-0.858***	-0.925***	-1.262***
	(0.208)	(0.314)	(0.152)	(0.162)	(0.359)
$FundSize$	-0.075***	-0.030	-0.034***	-0.036***	-0.032
	(0.021)	(0.022)	(0.011)	(0.013)	(0.027)
Fixed Effects					
• Fund	Yes	Yes	Yes	Yes	Yes
• Month	Yes	Yes	Yes	Yes	Yes
Observations	363,872	363,872	363,872	363,872	363,872
R^2	0.103	0.103	0.103	0.103	0.103
R^2 (proj. model)	0.001	0.001	0.001	0.001	0.001
Panel B: Real Time Portfolio Liquidity					
X =	L	S	D	C	B
$Comp.Size \times X$	0.161***	-0.238	0.040**	0.058***	0.415***
	(0.041)	(0.250)	(0.019)	(0.020)	(0.130)
$FundSize \times X$	0.008**	0.002	0.011***	0.006**	0.015**
	(0.004)	(0.008)	(0.002)	(0.003)	(0.007)
X	-0.753**	-1.053*	-0.053	0.027	-1.673***
	(0.294)	(0.554)	(0.091)	(0.066)	(0.279)
$CompetitorSize$	-0.869***	-0.462*	-0.893***	-0.957***	-1.082***
	(0.207)	(0.271)	(0.157)	(0.164)	(0.290)
$FundSize$	-0.050***	-0.025**	-0.052***	-0.043***	-0.048***
	(0.014)	(0.013)	(0.010)	(0.011)	(0.015)
Fixed Effects					
• Fund	Yes	Yes	Yes	Yes	Yes
• Month	Yes	Yes	Yes	Yes	Yes
Observations	363,872	363,872	363,872	363,872	363,872
R^2	0.103	0.103	0.103	0.103	0.103
R^2 (proj. model)	0.001	0.001	0.001	0.001	0.001

Panel A of Table C.2 presents results from the regressions. All γ_2 coefficients are positive and three out of five are statistically significant. The economic magnitudes are large as well. Increasing average portfolio liquidity from the 25$^{\text{th}}$ to the 75$^{\text{th}}$ percentile of its distribution changes the impact of a one standard deviation increase in $CompetitorSize$ by 23bp in annualized returns. Decomposing portfolio liquidity into its components demonstrates that the majority of the effect is attributable to stock liquidity, with a lesser amount attributable to diversification, including coverage and balance.[6]

6. In unreported results, I find a similar pattern of more severe decreasing returns for funds employing less liquid strategies using ad hoc measures of portfolio liquidity such as the portfolio-weighted average market

A related question is whether funds can actively ameliorate the pernicious effects of decreasing returns to (competitor) scale by choosing more liquid portfolios. I test this by replacing fund-level average measures of portfolio liquidity in equation (C.4) with real time values:

$$\begin{aligned} R^{FF3}_{i,t+1} = \alpha_i + \alpha_t + \gamma_1 CompetitorSize_{i,t} + \gamma_2 \left(CompetitorSize_{i,t} \times PortLiq_{i,t}\right) \\ + \eta_1 FundSize_{i,t} + \eta_2 \left(FundSize_{i,t} \times PortLiq_{i,t}\right) \\ + \gamma_3 PortLiq_{i,t} + \varepsilon_{i,t+1}. \end{aligned} \tag{C.5}$$

Panel B of Table C.2 presents results from the regressions. With the exception of stock liquidity, the interaction terms are all positive and statistically significant, suggesting that increased portfolio liquidity shelters the fund from decreasing returns to scale. Note that the main effect of portfolio liquidity is negative, suggesting that funds make more when they hold more concentrated portfolios. This is consistent with funds responding to time-varying investment opportunities by decreasing portfolio liquidity when expected returns are high.

C.5 Results Using Pre-2008 Data

Zhu (2017) finds that Thomson's coverage of new funds deteriorates after 2008, leading to potential bias. To make sure this data quality issue is not responsible for my findings, I re-do the analysis using only data up to 2008. The results, reported in Table C.3, remain substantively similar.

Table C.3
Fund Performance and Competitor Scale — Pre-2008 Data

Observations are at the fund-month level, over the period 1980-2007. The dependent variable is three-factor adjusted gross returns, in annualized percentages. *CompetitorSize* and *IndustrySize* are normalized by their respective sample standard deviations. *FundSize* is normalized by the difference between the 50th and 10th percentile of its distribution. Standard errors are double clustered by fund and year-month × portfolio group, and reported in parentheses. Asterisks denote statistical significance: *** p <0.01, ** p <0.05, * p <0.1.

	(1)	(2)	(3)	(4)
CompetitorSize	-1.065***		-1.203***	-0.839***
	(0.222)		(0.310)	(0.209)
IndustrySize		-0.401**	0.152	
		(0.186)	(0.256)	
FundSize				-0.035***
				(0.012)
Fixed Effects				
• Fund	Yes	Yes	Yes	Yes
• Month	No	No	No	Yes
Observations	230,471	230,471	230,471	230,471
R^2	0.018	0.017	0.018	0.106
R^2 (proj. model)	0.001	0.000	0.001	0.001

weight of holdings, number of stocks held, share of largest five holdings, the Herfindahl-Hischman Index of portfolio weights, as well as own size and turnover.

Table C.4
The Role of Portfolio Liquidity — Pre-2008 Data

Observations are at the fund-month level, over the period 1980-2007. The dependent variable is three-factor adjusted gross returns, in annualized percentages. L, S, D, C, B are portfolio liquidity, stock liquidity, diversification, coverage, and balance, as defined in Pástor et al. (2017b). $\bar{L}$, $\bar{S}$, $\bar{D}$, $\bar{C}$, $\bar{B}$ denote fund-level means. Each X variable is normalized by its interquartile range. Standard errors are double clustered by fund and year-month × portfolio group, and reported in parentheses. Asterisks denote statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

	(1)	(2)	(3)	(4)	(5)
Panel A: Fund Level Average Portfolio Liquidity					
$\bar{X}$ =	$\bar{L}$	$\bar{S}$	$\bar{D}$	$\bar{C}$	$\bar{B}$
$Comp.Size \times \bar{X}$	0.255**	0.181	0.037*	0.066**	0.482*
	(0.123)	(0.379)	(0.021)	(0.031)	(0.265)
$FundSize \times \bar{X}$	0.031**	0.026	0.012	0.003	0.006
	(0.014)	(0.031)	(0.010)	(0.005)	(0.025)
$CompetitorSize$	-1.305***	-1.043**	-0.898***	-0.973***	-1.619***
	(0.320)	(0.475)	(0.220)	(0.235)	(0.495)
$FundSize$	-0.093***	-0.053*	-0.055**	-0.046**	-0.046
	(0.029)	(0.031)	(0.021)	(0.021)	(0.043)
Fixed Effects					
• Fund	Yes	Yes	Yes	Yes	Yes
• Month	Yes	Yes	Yes	Yes	Yes
Observations	230,471	230,471	230,471	230,471	230,471
R^2	0.106	0.106	0.106	0.106	0.106
R^2 (proj. model)	0.001	0.001	0.001	0.001	0.001
Panel B: Real Time Portfolio Liquidity					
X =	L	S	D	C	B
$Comp.Size \times X$	0.170***	-0.263	0.075**	0.077	0.480***
	(0.060)	(0.327)	(0.035)	(0.047)	(0.177)
$FundSize \times X$	0.006	0.008	0.014**	0.004	0.011
	(0.005)	(0.009)	(0.006)	(0.004)	(0.007)
X	-0.812**	-0.700	-0.430***	-0.029	-2.129***
	(0.385)	(0.644)	(0.162)	(0.137)	(0.375)
$CompetitorSize$	-0.920***	-0.463	-0.879***	-1.024***	-1.129***
	(0.304)	(0.429)	(0.230)	(0.245)	(0.396)
$FundSize$	-0.054***	-0.043***	-0.065***	-0.046***	-0.052***
	(0.016)	(0.016)	(0.015)	(0.014)	(0.017)
Fixed Effects					
• Fund	Yes	Yes	Yes	Yes	Yes
• Month	Yes	Yes	Yes	Yes	Yes
Observations	230,471	230,471	230,471	230,471	230,471
R^2	0.106	0.106	0.106	0.106	0.106
R^2 (proj. model)	0.001	0.001	0.001	0.001	0.001

APPENDIX D

ADDITIONAL RESULTS: CAPITAL ALLOCATION

Table D.1
Capital Allocation and Competitor Scale — Pre-2008 Data

Observations are at the fund-quarter level, from 1980-2007. Δ denotes first differences. Dependent variables are noted in the column headers. AS is active share relative to self-declared benchmarks from Petajisto (2013). T is turnover; L, S, D, C, and B are respectively portfolio liquidity, stock liquidity, diversification, coverage, and balance, as defined in Pástor et al. (2017b). $\Delta CS_{i,t} = \ln\left(\sum_{j\neq i}\psi_{i,j,t-1}FundSize_{j,t}\right) - \ln\left(\sum_{j\neq i}\psi_{i,j,t-1}FundSize_{j,t-1}\right)$ is the change in log competitor size, holding previous quarter end similarity weights fixed. Standard errors are double clustered by fund and portfolio group $\times$ quarter, and reported in parentheses. Asterisks denote statistical significance: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Dep. Var.:	$\Delta\ln(AS)$	$\Delta\ln(TL^{-1/2})$	$\Delta\ln(L)$	$\Delta\ln(S)$	$\Delta\ln(D)$	$\Delta\ln(C)$	$\Delta\ln(B)$
ΔCS	-0.040**	-0.535***	0.862***	0.668***	0.666***	0.169***	0.594***
	(0.016)	(0.065)	(0.073)	(0.078)	(0.059)	(0.028)	(0.058)
$\Delta\ln(FundSize)$	-0.017***	-0.149***	0.224***	0.131***	0.199***	0.118***	0.122***
	(0.003)	(0.017)	(0.018)	(0.015)	(0.015)	(0.009)	(0.013)
$\Delta\ln(f)$	-0.024	0.002	0.051**	-0.004	0.065**	0.037***	0.041*
	(0.018)	(0.029)	(0.024)	(0.022)	(0.026)	(0.014)	(0.021)
$\Delta\ln(T)$			-0.004	-0.003	-0.003	0.003	-0.006
			(0.005)	(0.005)	(0.005)	(0.003)	(0.005)
$\Delta\ln(D)$				-0.394***			
				(0.024)			
$\Delta\ln(S)$					-0.639***	-0.166***	-0.567***
					(0.019)	(0.017)	(0.026)
$\Delta\ln(B)$						-0.122***	
						(0.015)	
$\Delta\ln(C)$							-0.278***
							(0.029)
Fixed Effects							
• Quarter	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	31,063	32,901	32,901	32,901	32,901	32,901	32,901
R^2	0.023	0.026	0.056	0.278	0.274	0.090	0.260
R^2 (proj. model)	0.004	0.012	0.044	0.262	0.261	0.067	0.252

Table D.2
Capital Allocation and Competitor Scale — Benchmark × Quarter FE

Observations are at the fund-quarter level, from 1980-2007. Δ denotes first differences. Dependent variables are noted in the column headers. AS is active share relative to self-declared benchmarks from Petajisto (2013). T is turnover; L, S, D, C, and B are respectively portfolio liquidity, stock liquidity, diversification, coverage, and balance, as defined in Pástor et al. (2017b). Benchmarks are the indexes which yield the lowest active share, taken from Petajisto (2013). I use the most recently available benchmark when one is missing. Standard errors are double clustered by fund and benchmark × quarter, and reported in parentheses. Asterisks denote statistical significance: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Dep. Var.:	$\Delta\ln(AS)$	$\Delta\ln(TL^{-1/2})$	$\Delta\ln(L)$	$\Delta\ln(S)$	$\Delta\ln(D)$	$\Delta\ln(C)$	$\Delta\ln(B)$
ΔCS	-0.022*	-0.410***	0.525***	0.357***	0.433***	0.094***	0.402***
	(0.012)	(0.070)	(0.064)	(0.066)	(0.056)	(0.033)	(0.053)
$\Delta\ln(FundSize)$	-0.014***	-0.106***	0.169***	0.086***	0.155***	0.096***	0.094***
	(0.002)	(0.012)	(0.011)	(0.008)	(0.010)	(0.007)	(0.009)
$\Delta\ln(f)$	-0.023	0.027	0.047**	-0.004	0.058**	0.030**	0.040*
	(0.019)	(0.031)	(0.024)	(0.018)	(0.025)	(0.013)	(0.021)
$\Delta\ln(T)$			-0.006	-0.004	-0.005	0.004	-0.008**
			(0.004)	(0.004)	(0.004)	(0.002)	(0.004)
$\Delta\ln(D)$				-0.361***			
				(0.019)			
$\Delta\ln(S)$					-0.649***	-0.209***	-0.548***
					(0.017)	(0.016)	(0.022)
$\Delta\ln(B)$						-0.134***	
						(0.012)	
$\Delta\ln(C)$							-0.289***
							(0.024)
Fixed Effects							
• Benchmark × Quarter	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	35,285	51,972	51,972	51,972	51,972	51,972	51,972
R^2	0.113	0.054	0.102	0.314	0.291	0.139	0.264
R^2 (proj. model)	0.002	0.004	0.017	0.236	0.240	0.080	0.216

Table D.3
Capital Allocation and Competitor Scale — Benchmark × Quarter FE, Pre-2008 Data

Observations are at the fund-quarter level, from 1980-2007. Δ denotes first differences. Dependent variables are noted in the column headers. AS is active share relative to self-declared benchmarks from Petajisto (2013). T is turnover; L, S, D, C, and B are respectively portfolio liquidity, stock liquidity, diversification, coverage, and balance, as defined in Pástor et al. (2017b). Benchmarks are the indexes which yield the lowest active share, taken from Petajisto (2013). I use the most recently available benchmark when one is missing. Standard errors are double clustered by fund and benchmark × quarter, and reported in parentheses. Asterisks denote statistical significance: *** p <0.01, ** p <0.05, * p <0.1.

Dep. Var.:	$\Delta \ln(AS)$	$\Delta \ln(TL^{-1/2})$	$\Delta \ln(L)$	$\Delta \ln(S)$	$\Delta \ln(D)$	$\Delta \ln(C)$	$\Delta \ln(B)$
ΔCS	-0.020*	-0.431***	0.588***	0.378***	0.502***	0.083**	0.487***
	(0.012)	(0.081)	(0.075)	(0.080)	(0.065)	(0.039)	(0.061)
$\Delta \ln(FundSize)$	-0.015***	-0.140***	0.207***	0.113***	0.188***	0.117***	0.112***
	(0.003)	(0.016)	(0.015)	(0.012)	(0.014)	(0.009)	(0.012)
$\Delta \ln(f)$	-0.025	-0.002	0.056**	-0.002	0.071**	0.039**	0.046*
	(0.020)	(0.033)	(0.027)	(0.021)	(0.030)	(0.015)	(0.024)
$\Delta \ln(T)$			-0.008	-0.004	-0.007	0.003	-0.010**
			(0.006)	(0.005)	(0.005)	(0.003)	(0.005)
$\Delta \ln(D)$				-0.392***			
				(0.024)			
$\Delta \ln(S)$					-0.645***	-0.178***	-0.568***
					(0.020)	(0.018)	(0.026)
$\Delta \ln(B)$						-0.131***	
						(0.016)	
$\Delta \ln(C)$							-0.284***
							(0.031)
Fixed Effects							
• Benchmark × Quarter	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	31,063	30,968	30,968	30,968	30,968	30,968	30,968
R^2	0.114	0.062	0.105	0.330	0.310	0.132	0.295
R^2 (proj. model)	0.002	0.007	0.023	0.255	0.261	0.070	0.252

APPENDIX E
ADDITIONAL RESULTS: MUTUAL FUND SCANDAL

Table E.1
Fund Characteristics as of August 2003 by Scandal Involvement

Means of various characteristics as of August 2003, depending on whether the family the fund belongs to was later implicated in the late trading scandal. Returns are annualized, in percentages.

Scandal involvement	No	Yes
N	1,172	289
$CompetitorSize \times 10^2$	0.41	0.49
TNA (100m \$)	10.73	12.52
R^{FF3}	-6.75	-10.31
Fund age	11.62	14.73
Expense ratio	1.35	1.46
AS	0.72	0.79
T	0.84	1
L	0.06	0.06
$ln(TL^{-1/2})$	1.35	1.46

Table E.2
Untainted Fund Characteristics by $ScandalExposure$ as of August 2003

$ScandalExposure$:	Below median	Above median
N	586	586
$CompetitorSize \times 10^2$	0.38	0.44
TNA (100m \$)	10.98	10.48
R^{FF3}	-2.34	-11.11
Fund age	11.58	11.66
Expense ratio	1.31	1.38
AS	0.71	0.74
T	0.72	0.96
L	0.06	0.05
$ln(TL^{-1/2})$	1.18	1.51

Table E.3
Capital Allocation and $ScandalExposure$: Testing for Pre-Trends

Dependent variables are identified in the column headers. ln($C.S.$) is an abbreviation for ln($CompetitorSize$). Observations are at the fund-report date level, including only funds not directly involved in the scandal over the period $[2003m8 - W, 2003m8]$, where W corresponds to the number of years specified in the panel headers. $t \times ScanEx$ is the interaction of a linear time trend with $ScandalExposure$ (normalized by its interquartile range). Standard errors are double clustered by fund and portfolio group × time, and reported in parentheses. Asterisks denote statistical significance: *** p <0.01, ** p <0.05, * p <0.1.

Dep. Var.:	ln($C.S.$)	ln(AS)	ln($TL^{-1/2}$)	ln(L)	ln(S)	ln(D)	ln(C)	ln(B)
				Panel A: 1 year window				
$t \times ScanEx$	0.114	0.015	0.930	0.301	0.395	0.090	0.236	-0.107
	(0.429)	(0.094)	(0.584)	(0.453)	(0.388)	(0.397)	(0.223)	(0.332)
ln($FundSize$)	0.083	0.026	-0.068	0.090	0.052	0.075	0.023	0.063
	(0.060)	(0.025)	(0.069)	(0.066)	(0.051)	(0.064)	(0.037)	(0.051)
ln(f)	0.078	0.065	0.307	0.042	0.012	0.045	-0.058	0.101
	(0.182)	(0.063)	(0.187)	(0.180)	(0.097)	(0.176)	(0.109)	(0.130)
ln(T)				-0.008	-0.000	-0.009	-0.011	-0.001
				(0.020)	(0.015)	(0.019)	(0.015)	(0.015)
ln(D)					-0.350***			
					(0.075)			
ln(S)						-0.527***	-0.273***	-0.346***
						(0.129)	(0.083)	(0.117)
ln(B)							-0.126**	
							(0.056)	
ln(C)								-0.235**
								(0.103)
Fixed Effects								
• Fund	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
• Time	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,137	2,706	3,093	3,093	3,093	3,093	3,093	3,093
R^2	0.973	0.973	0.969	0.989	0.995	0.982	0.982	0.965
R^2 (proj. model)	0.003	0.005	0.007	0.004	0.187	0.186	0.127	0.111
				Panel B: 2 year window				
$t \times ScanEx$	0.270	-0.010	0.563*	0.146	-0.005	0.179	-0.018	0.213
	(0.208)	(0.055)	(0.325)	(0.239)	(0.145)	(0.241)	(0.134)	(0.208)
ln($FundSize$)	0.196***	-0.012	-0.197***	0.262***	0.142***	0.221***	0.088***	0.162***
	(0.032)	(0.009)	(0.043)	(0.041)	(0.031)	(0.038)	(0.021)	(0.034)
ln(f)	-0.023	-0.021	0.278***	0.014	-0.078	0.068	0.018	0.058
	(0.080)	(0.034)	(0.103)	(0.107)	(0.081)	(0.091)	(0.062)	(0.082)
ln(T)				-0.016	-0.017	-0.008	-0.009	-0.001
				(0.026)	(0.017)	(0.023)	(0.015)	(0.017)
ln(D)					-0.304***			
					(0.037)			
ln(S)						-0.538***	-0.241***	-0.369***
						(0.078)	(0.054)	(0.076)
ln(B)							-0.106***	
							(0.036)	
ln(C)								-0.179***
								(0.063)
Fixed Effects								
• Fund	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
• Time	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5,996	5,190	5,898	5,898	5,898	5,898	5,898	5,898
R^2	0.949	0.955	0.938	0.977	0.990	0.960	0.958	0.926
R^2 (proj. model)	0.036	0.002	0.027	0.053	0.177	0.181	0.086	0.124

Table E.4
Capital Allocation and the Scandal: Before and After Analysis — Benchmark × Time FE

Dependent variables are identified in the column headers. ln($C.S.$) is an abbreviation for ln($CompetitorSize$). Observations are at the fund-report date level, including only funds not directly involved in the scandal over the period $\{(2003m8 - W, 2003m8], [2004m11, 2004m11 + W)\}$, where W corresponds to the number of years specified in the panel headers. $\mathbb{I} \times ScanEx$ is the interaction of $ScandalExposure$ (normalized by its interquartile range) and an indicator for the post-scandal period. Benchmarks are the indexes which yield the lowest active share, taken from Petajisto (2013). I use the most recently available benchmark when one is missing. Standard errors are double clustered by fund and benchmark × date, and reported in parentheses. Asterisks denote statistical significance: *** p <0.01, ** p <0.05, * p <0.1.

Dep. Var.:	ln($C.S.$)	ln(AS)	ln($TL^{-1/2}$)	ln(L)	ln(S)	ln(D)	ln(C)	ln(B)
Panel A: 1 year window								
$\mathbb{I} \times ScanEx$	-0.065***	0.018***	-0.015	-0.097***	-0.068***	-0.062**	-0.035**	-0.030
	(0.022)	(0.006)	(0.038)	(0.029)	(0.023)	(0.025)	(0.016)	(0.020)
ln($FundSize$)	0.125***	-0.006	-0.120***	0.145***	0.054***	0.126***	0.056***	0.077***
	(0.021)	(0.006)	(0.032)	(0.026)	(0.014)	(0.026)	(0.019)	(0.017)
ln(f)	0.024	-0.015	0.082	0.036	-0.060	0.083	0.068	0.020
	(0.106)	(0.027)	(0.149)	(0.119)	(0.074)	(0.114)	(0.072)	(0.085)
ln(T)				-0.060**	-0.055***	-0.029	0.024	-0.054***
				(0.028)	(0.016)	(0.027)	(0.015)	(0.020)
ln(D)					-0.189***			
					(0.024)			
ln(S)						-0.429***	-0.255***	-0.198***
						(0.055)	(0.045)	(0.051)
ln(B)							-0.045	
							(0.033)	
ln(C)								-0.067
								(0.048)
Fixed Effects								
• Fund	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
• Benchmark × Time	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6,549	6,067	6,464	6,464	6,464	6,464	6,464	6,464
R^2	0.931	0.932	0.905	0.966	0.990	0.944	0.948	0.899
R^2 (proj. model)	0.043	0.008	0.012	0.052	0.108	0.101	0.081	0.046
Panel B: 2 year window								
$\mathbb{I} \times ScanEx$	-0.032	0.020***	-0.014	-0.085***	-0.080***	-0.038	-0.036**	-0.004
	(0.023)	(0.006)	(0.038)	(0.030)	(0.024)	(0.026)	(0.017)	(0.020)
ln($FundSize$)	0.140***	-0.010*	-0.169***	0.166***	0.059***	0.144***	0.063***	0.088***
	(0.018)	(0.006)	(0.026)	(0.024)	(0.013)	(0.023)	(0.016)	(0.017)
ln(f)	-0.075	0.007	0.181	-0.076	-0.144**	0.018	0.050	-0.030
	(0.071)	(0.025)	(0.129)	(0.109)	(0.063)	(0.094)	(0.063)	(0.074)
ln(T)				-0.045*	-0.047***	-0.017	0.024*	-0.041***
				(0.023)	(0.012)	(0.022)	(0.013)	(0.015)
ln(D)					-0.171***			
					(0.020)			
ln(S)						-0.404***	-0.238***	-0.187***
						(0.044)	(0.036)	(0.044)
ln(B)							-0.042*	
							(0.025)	
ln(C)								-0.060*
								(0.036)
Fixed Effects								
• Fund	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
• Benchmark × Time	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	12,685	11,764	12,495	12,495	12,495	12,495	12,495	12,495
R^2	0.905	0.898	0.869	0.950	0.986	0.919	0.924	0.858
R^2 (proj. model)	0.055	0.010	0.030	0.062	0.104	0.096	0.071	0.046

Table E.5
Capital Allocation and $ScandalExposure$: Testing for Pre-Trends — Benchmark $\times$ Time FE

Dependent variables are identified in the column headers. $\ln(C.S.)$ is an abbreviation for $\ln(CompetitorSize)$. Observations are at the fund-report date level, including only funds not directly involved in the scandal over the period $[2003m8 - W, 2003m8]$, where W corresponds to the number of years specified in the panel headers. $t \times ScanEx$ is the interaction a linear time trend with $ScandalExposure$ (normalized by its interquartile range). Benchmarks are the indexes which yield the lowest active share, taken from Petajisto (2013). I use the most recently available benchmark when one is missing. Standard errors are double clustered by fund and benchmark $\times$ time, and reported in parentheses. Asterisks denote statistical significance: *** p <0.01, ** p <0.05, * p <0.1.

Dep. Var.:	$\ln(C.S.)$	$\ln(AS)$	$\ln(TL^{-1/2})$	$\ln(L)$	$\ln(S)$	$\ln(D)$	$\ln(C)$	$\ln(B)$
Panel A: 1 year window								
$t \times ScanEx$	0.178	0.067	1.072	0.093	-0.270	0.262	-0.069	0.353
	(0.466)	(0.112)	(0.881)	(0.469)	(0.353)	(0.436)	(0.245)	(0.405)
$\ln(FundSize)$	0.076	0.015	-0.059	0.079	0.042	0.068	0.004	0.072*
	(0.056)	(0.016)	(0.076)	(0.055)	(0.030)	(0.055)	(0.036)	(0.041)
$\ln(f)$	0.157	0.037	0.297	0.152	0.082	0.131	-0.010	0.154
	(0.204)	(0.047)	(0.264)	(0.161)	(0.090)	(0.154)	(0.078)	(0.136)
$\ln(T)$				-0.013	-0.003	-0.013	-0.013	-0.003
				(0.023)	(0.016)	(0.022)	(0.016)	(0.017)
$\ln(D)$					-0.294***			
					(0.056)			
$\ln(S)$						-0.598***	-0.321***	-0.379***
						(0.077)	(0.085)	(0.103)
$\ln(B)$							-0.124*	
							(0.072)	
$\ln(C)$								-0.220*
								(0.132)
Fixed Effects								
• Fund	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
• Benchmark $\times$ Time	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,917	2,706	2,908	2,908	2,908	2,908	2,908	2,908
R^2	0.976	0.981	0.971	0.991	0.997	0.984	0.984	0.968
R^2 (proj. model)	0.003	0.003	0.006	0.004	0.178	0.178	0.125	0.102
Panel B: 2 year window								
$t \times ScanEx$	0.386	-0.012	0.850*	0.311	-0.050	0.400	-0.041	0.474*
	(0.294)	(0.052)	(0.439)	(0.308)	(0.183)	(0.292)	(0.139)	(0.267)
$\ln(FundSize)$	0.199***	-0.015	-0.199***	0.283***	0.139***	0.241***	0.095***	0.176***
	(0.037)	(0.010)	(0.048)	(0.045)	(0.026)	(0.044)	(0.024)	(0.041)
$\ln(f)$	-0.015	-0.025	0.322***	0.019	-0.072	0.071	0.031	0.049
	(0.083)	(0.031)	(0.106)	(0.107)	(0.072)	(0.094)	(0.059)	(0.082)
$\ln(T)$				-0.008	-0.013	-0.001	-0.003	0.002
				(0.025)	(0.016)	(0.023)	(0.015)	(0.018)
$\ln(D)$					-0.267***			
					(0.037)			
$\ln(S)$						-0.513***	-0.248***	-0.332***
						(0.084)	(0.058)	(0.077)
$\ln(B)$							-0.101**	
							(0.040)	
$\ln(C)$								-0.173**
								(0.072)
Fixed Effects								
• Fund	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
• Benchmark $\times$ Time	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5,604	5,190	5,574	5,574	5,574	5,574	5,574	5,574
R^2	0.954	0.962	0.943	0.980	0.993	0.964	0.963	0.933
R^2 (proj. model)	0.035	0.003	0.028	0.058	0.151	0.161	0.086	0.104

Table E.6
Capital Allocation and $ScandalOutFlow$: Testing for Pre-Trends

Dependent variables are identified in the column headers. ln($C.S.$) is an abbreviation for ln($CompetitorSize$). Observations are at the fund-report date level, including only funds not directly involved in the scandal over the period $[2003m8 - W, 2003m8]$, where W denotes the number of years specified. $\overline{ScandalOutFlow}$ is the fund-level mean $ScandalOutFlow$ over the specified post-scandal window, normalized by interquartile range. Benchmarks are the indexes which yield the lowest active share, taken from Petajisto (2013). I use the most recently available benchmark when one is missing. Standard errors are double clustered by fund and portfolio group × time, and reported in parentheses. Asterisks denote statistical significance: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Dep. Var.:	ln($C.S.$)	ln(AS)	ln($TL^{-1/2}$)	ln(L)	ln(S)	ln(D)	ln(C)	ln(B)
Panel A: 1 year window								
$t \times \overline{ScandalOutFlow}$	0.049*	-0.002	0.046	0.030	0.030	0.017	-0.008	0.026
	(0.027)	(0.007)	(0.040)	(0.027)	(0.023)	(0.024)	(0.015)	(0.019)
ln($FundSize$)	0.075	0.024	-0.067	0.067	0.052	0.048	0.005	0.049
	(0.056)	(0.023)	(0.064)	(0.061)	(0.054)	(0.058)	(0.035)	(0.046)
ln(f)	0.067	0.057	0.295	0.041	0.026	0.033	-0.057	0.087
	(0.176)	(0.060)	(0.182)	(0.174)	(0.096)	(0.172)	(0.103)	(0.129)
ln(T)				-0.003	0.002	-0.005	-0.006	-0.001
				(0.019)	(0.013)	(0.019)	(0.014)	(0.014)
ln(D)					-0.366***			
					(0.074)			
ln(S)						-0.561***	-0.274***	-0.376***
						(0.134)	(0.081)	(0.123)
ln(B)							-0.115**	
							(0.053)	
ln(C)								-0.221**
								(0.103)
Fixed Effects								
• Fund	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
• Time	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,357	2,881	3,291	3,291	3,291	3,291	3,291	3,291
R^2	0.974	0.973	0.968	0.990	0.995	0.983	0.983	0.966
R^2 (proj. model)	0.006	0.005	0.005	0.003	0.207	0.205	0.128	0.124
Panel B: 2 year window								
$t \times \overline{ScandalOutFlow}$	0.024	-0.006	0.036*	0.020	0.009	0.019	0.007	0.014
	(0.016)	(0.006)	(0.020)	(0.017)	(0.011)	(0.016)	(0.010)	(0.012)
ln($FundSize$)	0.184***	-0.009	-0.198***	0.246***	0.141***	0.212***	0.077***	0.162***
	(0.033)	(0.008)	(0.043)	(0.041)	(0.031)	(0.038)	(0.020)	(0.034)
ln(f)	-0.027	-0.024	0.273***	0.026	-0.047	0.060	0.012	0.055
	(0.081)	(0.032)	(0.103)	(0.107)	(0.084)	(0.089)	(0.058)	(0.083)
ln(T)				-0.009	-0.018	-0.001	-0.009	0.008
				(0.025)	(0.017)	(0.023)	(0.014)	(0.017)
ln(D)					-0.339***			
					(0.041)			
ln(S)						-0.604***	-0.248***	-0.436***
						(0.080)	(0.050)	(0.082)
ln(B)							-0.106***	
							(0.033)	
ln(C)								-0.184***
								(0.061)
Fixed Effects								
• Fund	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
• Time	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6,438	5,541	6,318	6,318	6,318	6,318	6,318	6,318
R^2	0.947	0.954	0.937	0.977	0.990	0.960	0.957	0.926
R^2 (proj. model)	0.032	0.003	0.026	0.047	0.215	0.219	0.088	0.161

Table E.7
Capital Allocation and $ScandalOutFlow$ — Benchmark × Time FE

Dependent variables are identified in the column headers. ln($C.S.$) is an abbreviation for ln($CompetitorSize$). Observations are at the fund-report date level, including only funds not directly involved in the scandal over the period $\{[2003m9 - W, 2004m10 + W]\}$, where W corresponds to the number of years specified. $ScandalOutFlow$ is normalized by its interquartile range. Benchmarks are the indexes which yield the lowest active share, taken from Petajisto (2013). I use the most recently available benchmark when one is missing. Standard errors are double clustered by fund and benchmark × time, and reported in parentheses. Asterisks denote statistical significance: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Dep. Var.:	ln($C.S.$)	ln(AS)	ln($TL^{-1/2}$)	ln(L)	ln(S)	ln(D)	ln(C)	ln(B)
Panel A: 1 year window								
$ScandalOutFlow$	-0.203***	0.058***	0.092**	-0.250***	-0.111***	-0.215***	-0.074***	-0.163***
	(0.021)	(0.008)	(0.038)	(0.030)	(0.019)	(0.028)	(0.018)	(0.020)
ln($FundSize$)	0.096***	-0.005	-0.103***	0.127***	0.054***	0.111***	0.065***	0.059***
	(0.017)	(0.005)	(0.028)	(0.022)	(0.012)	(0.021)	(0.016)	(0.015)
ln(f)	-0.050	-0.010	0.093	-0.011	-0.055	0.023	0.065	-0.036
	(0.076)	(0.023)	(0.116)	(0.090)	(0.055)	(0.088)	(0.058)	(0.065)
ln(T)				-0.057***	-0.040***	-0.040**	0.010	-0.052***
				(0.019)	(0.011)	(0.019)	(0.011)	(0.015)
ln(D)					-0.227***			
					(0.022)			
ln(S)						-0.501***	-0.265***	-0.295***
						(0.049)	(0.038)	(0.049)
ln(B)							-0.091***	
							(0.029)	
ln(C)								-0.144***
								(0.044)
Fixed Effects								
• Fund	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
• Benchmark × Time	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	11,181	10,324	10,949	10,949	10,949	10,949	10,949	10,949
R^2	0.939	0.945	0.908	0.971	0.991	0.951	0.953	0.907
R^2 (proj. model)	0.076	0.053	0.012	0.089	0.132	0.155	0.091	0.097
Panel B: 2 year window								
$ScandalOutFlow$	-0.161***	0.059***	0.097***	-0.226***	-0.106***	-0.188***	-0.070***	-0.134***
	(0.019)	(0.007)	(0.032)	(0.027)	(0.018)	(0.026)	(0.018)	(0.018)
ln($FundSize$)	0.118***	-0.008	-0.153***	0.147***	0.058***	0.129***	0.066***	0.075***
	(0.015)	(0.005)	(0.023)	(0.021)	(0.012)	(0.020)	(0.015)	(0.015)
ln(f)	-0.113*	0.011	0.200*	-0.110	-0.145**	-0.030	0.044	-0.074
	(0.059)	(0.022)	(0.105)	(0.097)	(0.058)	(0.084)	(0.055)	(0.068)
ln(T)				-0.044**	-0.039***	-0.025	0.014	-0.040***
				(0.019)	(0.010)	(0.018)	(0.011)	(0.013)
ln(D)					-0.209***			
					(0.019)			
ln(S)						-0.481***	-0.258***	-0.269***
						(0.040)	(0.033)	(0.042)
ln(B)							-0.076***	
							(0.024)	
ln(C)								-0.113***
								(0.034)
Fixed Effects								
• Fund	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
• Benchmark × Time	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	17,525	16,197	17,188	17,188	17,188	17,188	17,188	17,188
R^2	0.913	0.913	0.875	0.956	0.987	0.928	0.929	0.869
R^2 (proj. model)	0.078	0.058	0.030	0.097	0.128	0.144	0.084	0.084

Table E.8
Capital Allocation and $ScandalOutFlow$: Testing for Pre-Trends — Benchmark × Time FE

Dependent variables are identified in the column headers. $\ln(C.S.)$ is an abbreviation for $\ln(CompetitorSize)$. Observations are at the fund-report date level, including only funds not directly involved in the scandal over the period $[2003m8 - W, 2003m8]$, where W denotes the number of years specified. $\overline{ScandalOutFlow}$ is the fund-level mean $ScandalOutFlow$ over the specified post-scandal window, normalized by interquartile range. Benchmarks are the Standard errors are double clustered by fund and benchmark × time, and reported in parentheses. Asterisks denote statistical significance: *** p <0.01, ** p <0.05, * p <0.1.

Dep. Var.:	$\ln(C.S.)$	$\ln(AS)$	$\ln(TL^{-1/2})$	$\ln(L)$	$\ln(S)$	$\ln(D)$	$\ln(C)$	$\ln(B)$
Panel A: 1 year window								
$t \times \overline{ScandalOutFlow}$	0.076***	-0.000	0.092*	0.020	0.005	0.020	-0.005	0.027
	(0.027)	(0.008)	(0.053)	(0.030)	(0.018)	(0.030)	(0.019)	(0.026)
$\ln(FundSize)$	0.074	0.013	-0.061	0.064	0.051	0.047	-0.006	0.057
	(0.054)	(0.015)	(0.075)	(0.051)	(0.034)	(0.051)	(0.033)	(0.039)
$\ln(f)$	0.153	0.026	0.269	0.160	0.096	0.134	-0.007	0.153
	(0.200)	(0.046)	(0.238)	(0.163)	(0.092)	(0.156)	(0.077)	(0.135)
$\ln(T)$				-0.010	-0.003	-0.010	-0.010	-0.002
				(0.021)	(0.015)	(0.021)	(0.015)	(0.016)
$\ln(D)$					-0.302***			
					(0.055)			
$\ln(S)$						-0.623***	-0.317***	-0.403***
						(0.077)	(0.082)	(0.097)
$\ln(B)$							-0.113*	
							(0.068)	
$\ln(C)$								-0.208
								(0.130)
Fixed Effects								
• Fund	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
• Benchmark × Time	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,121	2,881	3,093	3,093	3,093	3,093	3,093	3,093
R^2	0.977	0.980	0.970	0.992	0.997	0.984	0.984	0.969
R^2 (proj. model)	0.010	0.001	0.007	0.003	0.190	0.189	0.123	0.110
Panel B: 2 year window								
$t \times \overline{ScandalOutFlow}$	0.056***	-0.008	0.037	0.037*	0.021*	0.031	0.008	0.027
	(0.018)	(0.007)	(0.023)	(0.021)	(0.011)	(0.021)	(0.012)	(0.017)
$\ln(FundSize)$	0.186***	-0.011	-0.202***	0.260***	0.132***	0.227***	0.083***	0.173***
	(0.037)	(0.009)	(0.050)	(0.043)	(0.026)	(0.041)	(0.022)	(0.039)
$\ln(f)$	-0.030	-0.027	0.326***	0.015	-0.054	0.050	0.022	0.034
	(0.081)	(0.029)	(0.108)	(0.105)	(0.073)	(0.091)	(0.056)	(0.080)
$\ln(T)$				-0.002	-0.016	0.007	-0.004	0.011
				(0.024)	(0.016)	(0.023)	(0.014)	(0.019)
$\ln(D)$					-0.298***			
					(0.040)			
$\ln(S)$						-0.584***	-0.258***	-0.402***
						(0.086)	(0.056)	(0.085)
$\ln(B)$							-0.101***	
							(0.039)	
$\ln(C)$								-0.176**
								(0.070)
Fixed Effects								
• Fund	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
• Benchmark × Time	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6,016	5,541	5,967	5,967	5,967	5,967	5,967	5,967
R^2	0.952	0.961	0.942	0.980	0.992	0.963	0.962	0.932
R^2 (proj. model)	0.037	0.004	0.024	0.053	0.185	0.193	0.088	0.133

Table E.9
Capital Allocation and the Scandal: Instrumenting Competitor Size with Abnormal Flows

Dependent variables are identified in the column headers. ln($C.S.$) is an abbreviation for ln($CompetitorSize$). Observations are at the fund-report date level, including only funds not directly involved in the scandal over the period $\{[2003m9 - W, 2004m10 + W]\}$, where W corresponds to the number of years specified. $ScandalOutFlow$ is the similarity-weighted cumulative abnormal outflows attributable to the scandal among involved funds. $ScandalOutFlow$ is normalized by its interquartile range. Regressions are estimated via two stage least squares, instrumenting for ln($CompetitorSize_{i,t}$) with $ScandalOutFlow_{i,t}$. The F-statistic of the first stage relation is reported at the bottom of each panel. Standard errors are double clustered by fund and portfolio group × time, and are reported in parentheses. Asterisks denote statistical significance: *** p <0.01, ** p <0.05, * p <0.1.

Dep. Var.:	ln(AS)	ln($TL^{-1/2}$)	ln(L)	ln(S)	ln(D)	ln(C)	ln(B)
Panel A: 1 year window							
ln($CompetitorSize$)	-0.330***	-0.668***	1.213***	1.015***	1.078***	0.826***	0.856***
	(0.046)	(0.141)	(0.090)	(0.115)	(0.089)	(0.141)	(0.074)
ln($FundSize$)	0.021**	-0.068**	0.030*	0.026*	0.025	0.030*	0.004
	(0.009)	(0.031)	(0.018)	(0.015)	(0.017)	(0.017)	(0.012)
ln(f)	-0.025	0.034	0.034	0.012	0.049	0.067	-0.009
	(0.032)	(0.106)	(0.080)	(0.069)	(0.074)	(0.066)	(0.051)
ln(T)			-0.043***	-0.041***	-0.032***	-0.014	-0.043***
			(0.012)	(0.011)	(0.011)	(0.010)	(0.009)
ln(D)				-0.739***			
				(0.058)			
ln(S)					-0.705***	-0.560***	-0.528***
					(0.035)	(0.060)	(0.040)
ln(B)						-0.637***	
						(0.097)	
ln(C)							-0.422***
							(0.041)
Fixed Effects							
• Fund	Yes	Yes	Yes	Yes	Yes	Yes	Yes
• Time	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	10,324	11,689	11,689	11,689	11,689	11,689	11,689
F (first stage)	107.7	111.7	111.4	70.9	100.2	44.6	90.9
Panel B: 2 year window							
ln($CompetitorSize$)	-0.358***	-0.687***	1.241***	1.090***	1.046***	0.803***	0.792***
	(0.046)	(0.143)	(0.097)	(0.124)	(0.092)	(0.134)	(0.081)
ln($FundSize$)	0.028***	-0.097***	0.018	0.013	0.020	0.020	0.007
	(0.009)	(0.027)	(0.019)	(0.016)	(0.017)	(0.017)	(0.012)
ln(f)	-0.034	0.102	0.034	-0.001	0.076	0.094*	0.004
	(0.026)	(0.097)	(0.067)	(0.064)	(0.058)	(0.052)	(0.048)
ln(T)			-0.035***	-0.037***	-0.020*	-0.003	-0.033***
			(0.013)	(0.011)	(0.012)	(0.011)	(0.009)
ln(D)				-0.775***			
				(0.061)			
ln(S)					-0.657***	-0.513***	-0.484***
					(0.029)	(0.049)	(0.037)
ln(B)						-0.607***	
						(0.088)	
ln(C)							-0.408***
							(0.038)
Fixed Effects							
• Fund	Yes	Yes	Yes	Yes	Yes	Yes	Yes
• Time	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	16,197	18,309	18,309	18,309	18,309	18,309	18,309
F (first stage)	125.7	133.7	133.3	88.9	116.7	55.6	102.4

Table E.10
Capital Allocation and the Scandal: Instrumenting Competitor Size with Abnormal Flows — Benchmark × Time FE

Dependent variables are identified in the column headers. ln($C.S.$) is an abbreviation for ln($CompetitorSize$). Observations are at the fund-report date level, including only funds not directly involved in the scandal over the period $\{[2003m9 - W, 2004m10 + W]\}$, where W corresponds to the number of years specified. $ScandalOutFlow$ is the similarity-weighted cumulative abnormal outflows attributable to the scandal among involved funds. $ScandalOutFlow$ is normalized by its interquartile range. I estimate regressions via two stage least squares, instrumenting for ln($CompetitorSize_{i,t}$) with $ScandalOutFlow_{i,t}$. The F-statistic of the first stage relation is reported at the bottom of each panel. Standard errors are double clustered by fund and benchmark × time, and are reported in parentheses. Asterisks denote statistical significance: *** p <0.01, ** p <0.05, * p <0.1.

Dep. Var.:	ln(AS)	ln($TL^{-1/2}$)	ln(L)	ln(S)	ln(D)	ln(C)	ln(B)
Panel A: 1 year window							
ln($CompetitorSize$)	-0.297***	-0.453**	1.225***	0.961***	1.125***	0.839***	0.955***
	(0.045)	(0.177)	(0.107)	(0.154)	(0.108)	(0.200)	(0.094)
ln($FundSize$)	0.021**	-0.059*	0.008	0.005	0.008	0.016	-0.009
	(0.008)	(0.033)	(0.020)	(0.015)	(0.019)	(0.019)	(0.014)
ln(f)	-0.022	0.077	0.030	0.004	0.047	0.057	0.006
	(0.030)	(0.110)	(0.079)	(0.065)	(0.075)	(0.066)	(0.055)
ln(T)			-0.041***	-0.037***	-0.032***	-0.014	-0.041***
			(0.012)	(0.010)	(0.011)	(0.011)	(0.009)
ln(D)				-0.700***			
				(0.081)			
ln(S)					-0.706***	-0.562***	-0.545***
					(0.040)	(0.080)	(0.044)
ln(B)						-0.646***	
						(0.139)	
ln(C)							-0.463***
							(0.047)
Fixed Effects							
• Fund	Yes	Yes	Yes	Yes	Yes	Yes	Yes
• Time	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	10,324	10,949	10,949	10,949	10,949	10,949	10,949
F (first stage)	81.1	90.3	90.5	42.3	84.2	27.4	77.8
Panel B: 2 year window							
ln($CompetitorSize$)	-0.367***	-0.594***	1.392***	1.237***	1.252***	1.078***	1.050***
	(0.054)	(0.188)	(0.125)	(0.218)	(0.124)	(0.266)	(0.114)
ln($FundSize$)	0.033***	-0.084***	-0.016	-0.016	-0.010	-0.005	-0.018
	(0.010)	(0.031)	(0.023)	(0.020)	(0.021)	(0.024)	(0.016)
ln(f)	-0.031	0.143	0.024	0.002	0.059	0.063	0.019
	(0.025)	(0.100)	(0.068)	(0.067)	(0.059)	(0.054)	(0.049)
ln(T)			-0.035***	-0.035***	-0.024**	-0.016	-0.033***
			(0.013)	(0.012)	(0.012)	(0.013)	(0.009)
ln(D)				-0.834***			
				(0.116)			
ln(S)					-0.689***	-0.611***	-0.536***
					(0.037)	(0.094)	(0.044)
ln(B)						-0.795***	
						(0.184)	
ln(C)							-0.499***
							(0.050)
Fixed Effects							
• Fund	Yes	Yes	Yes	Yes	Yes	Yes	Yes
• Time	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	16,197	17,188	17,188	17,188	17,188	17,188	17,188
F (first stage)	69.7	73.5	73.3	31.5	63.3	18.9	55.0

Table E.11
Investor Flows and the Scandal

The dependent variable is net flows in monthly percent units. Observations are at the fund-month level, including only funds untainted by the scandal, over 1-2 year windows around the scandal. $\mathbb{I}$ is an indicator for the post scandal period. $ScandalExposure$ ($ScanEx$) and $ScandalOutFlow$ are normalized by interquartile range. Benchmarks are the indexes which yield the lowest active share, taken from Petajisto (2013). Standard errors are double clustered by fund and portfolio group × month in the month FE specifications, and by fund and benchmark × month in the benchmark × month specifications, reported in parentheses. Asterisks denote statistical significance: *** p <0.01, ** p <0.05, * p <0.1.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	1 year window				**2 year window**			
Panel A: Ignoring Past Performance								
$\mathbb{I} \times ScanEx$	-0.102	-0.052			0.286*	0.218		
	(0.161)	(0.194)			(0.152)	(0.167)		
$ScandalOutFlow$			0.082	0.068			0.369***	0.278**
			(0.136)	(0.166)			(0.111)	(0.124)
Fixed Effects								
• Fund	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
• Month	Yes	No	Yes	No	Yes	No	Yes	No
• Benchmark × Month	No	Yes	No	Yes	No	Yes	No	Yes
Observations	25,179	23,122	41,715	38,293	48,883	44,916	66,252	60,849
R^2	0.296	0.307	0.297	0.307	0.234	0.257	0.243	0.261
R^2 (proj. model)	0.000	0.000	0.000	0.000	0.001	0.000	0.001	0.001
Panel B: Controlling for Past Performance								
$\mathbb{I} \times ScanEx$	-0.343**	-0.140			-0.248*	-0.080		
	(0.148)	(0.179)			(0.128)	(0.149)		
$ScandalOutFlow$			-0.103	-0.242			0.102	-0.007
			(0.125)	(0.155)			(0.093)	(0.109)
R^{FF3}_{t-1}	0.181***	0.188***	0.176***	0.192***	0.168***	0.165***	0.165***	0.169***
	(0.025)	(0.027)	(0.022)	(0.022)	(0.021)	(0.019)	(0.018)	(0.017)
R^{FF3}_{t-2}	0.158***	0.176***	0.159***	0.172***	0.161***	0.166***	0.159***	0.164***
	(0.021)	(0.022)	(0.019)	(0.018)	(0.020)	(0.017)	(0.017)	(0.015)
R^{FF3}_{t-3}	0.151***	0.158***	0.137***	0.148***	0.140***	0.145***	0.133***	0.141***
	(0.024)	(0.021)	(0.020)	(0.017)	(0.020)	(0.015)	(0.017)	(0.014)
R^{FF3}_{t-4}	0.080***	0.087***	0.079***	0.085***	0.099***	0.103***	0.094***	0.099***
	(0.018)	(0.018)	(0.015)	(0.015)	(0.017)	(0.017)	(0.015)	(0.014)
R^{FF3}_{t-5}	0.085***	0.098***	0.082***	0.091***	0.099***	0.109***	0.093***	0.102***
	(0.024)	(0.022)	(0.018)	(0.016)	(0.019)	(0.016)	(0.016)	(0.014)
R^{FF3}_{t-6}	0.102***	0.118***	0.086***	0.101***	0.116***	0.126***	0.102***	0.114***
	(0.019)	(0.021)	(0.015)	(0.017)	(0.016)	(0.016)	(0.013)	(0.014)
R^{FF3}_{t-7}	0.069***	0.080***	0.058***	0.063***	0.104***	0.110***	0.088***	0.092***
	(0.026)	(0.021)	(0.019)	(0.017)	(0.019)	(0.018)	(0.016)	(0.016)
R^{FF3}_{t-8}	0.093***	0.104***	0.066***	0.075***	0.094***	0.094***	0.076***	0.078***
	(0.025)	(0.018)	(0.019)	(0.015)	(0.017)	(0.014)	(0.014)	(0.012)
R^{FF3}_{t-9}	0.087***	0.098***	0.065***	0.081***	0.105***	0.114***	0.086***	0.101***
	(0.023)	(0.017)	(0.017)	(0.014)	(0.017)	(0.014)	(0.014)	(0.011)
R^{FF3}_{t-10}	0.120***	0.135***	0.077***	0.085***	0.103***	0.115***	0.079***	0.089***
	(0.023)	(0.021)	(0.016)	(0.015)	(0.015)	(0.016)	(0.013)	(0.013)
R^{FF3}_{t-11}	0.120***	0.130***	0.065***	0.080***	0.106***	0.108***	0.074***	0.083***
	(0.019)	(0.019)	(0.015)	(0.014)	(0.013)	(0.013)	(0.012)	(0.011)
R^{FF3}_{t-12}	0.067***	0.090***	0.052***	0.074***	0.066***	0.075***	0.056***	0.070***
	(0.019)	(0.018)	(0.015)	(0.014)	(0.013)	(0.011)	(0.012)	(0.010)
Fixed Effects								
• Fund	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
• Month	Yes	No	Yes	No	Yes	No	Yes	No
• Benchmark × Month	No	Yes	No	Yes	No	Yes	No	Yes
Observations	24,529	22,672	40,718	37,595	47,247	43,770	64,241	59,440
R^2	0.314	0.329	0.311	0.326	0.257	0.280	0.260	0.280
R^2 (proj. model)	0.025	0.028	0.019	0.022	0.035	0.035	0.028	0.028

APPENDIX F
DATA CONSTRUCTION

Code for reproducing the paper in full is available at https://github.com/laszlo-jakab/competitor-scale.[1] For a standalone toolbox to calculate the *CompetitorSize* variable using portfolio weights and fund sizes, see the `R` package `compsizer`, available at https://github.com/laszlo-jakab/compsizer.

F.1 CRSP Survivor-Bias-Free US Mutual Fund Database

F.1.1 Source Datasets

I download the following datasets from WRDS: `MONTHLY_TNA`, `MONTHLY_RETURNS`, `MONTHLY_NAV`, `FUND_HDR`, `FUND_HDR_HIST`, `FUND_FEES`, `FUND_STYLE`. The following is a summary of the steps I took for cleaning the data.

F.1.2 Data Cleaning

- `MONTHLY_TNA`
  - Set `mtna = -99` to missing.
  - Keep observations with positive, non-missing `mtna`.
  - Align `caldt` across share classes each month by taking the min (in a few cases, `caldt` is beyond the last trading day).
- `MONTHLY_RETURNS`
  - keep observations with non-missing `mret`.
  - align `caldt` across share classes each month as above.
  - keep returns only if they apply for a single month, as evidenced by non-missing `mnav` or `mret` in the previous month.
- `FUND_FEES`
  - Set `-99` to missing for all variables.
  - Set `exp_ratio <= 0` to missing.
  - Make sure `begdt` and `enddt` do not overlap (in a handful of cases, the next `begdt` is the same as the current `enddt`; if so, I decrement `enndt` by one day).
  - Break out `turn_ratio`, as `begdt` and `enddt` are not appropriate for this variable.
  - `turn_ratio` is valid for 12 months ending on fiscal_yearend when it is present, otherwise, for the 12 months ending on `begdt`; construct custom beginning and end dates for `turn_ratio` based on this fact.

1. Programs built under `R 3.4.3`, using packages `data.table 1.10.4-3`, and `zoo 1.8-1`. Fixed effects and instrumental variable regressions estimated using the `lfe 2.6-2291` package (Gaure, 2013).

- FUND_HDR, FUND_HDR_HIST
  - Append datasets.
  - Clean beginning and end dates.

F.1.3 Generating Fund-Level Dataset

1. Combine returns, TNA, expense ratio, turnover
   - Match returns and TNA on crsp_fundno and caldt.
   - Overlap match combined returns and TNA to exp_ratio on crsp_fundno, and begdt <= caldt <= enddt.
   - Perform similar overlap match for turnover, on user-defined beginning and end dates.
2. Fill in gaps in mtna by carrying forward the last non-missing observation, compounded by net returns.
3. Generate one-month lag TNA to be used as weights in the fund-level aggregation; fill in first observation for each fund by current TNA to retain as many observations as possible (this is used only for the fund-level aggregation). Keep only if lagged TNA is positive and non-missing.
4. Merge in the fund identifier wficn from MFLINKS.
5. Generate gross returns as mret.gross = mret + exp_ratio / 12.
6. Aggregate to fund level.
   - For each of net returns, gross returns, expense ratio, and turnover, calculate lagged-TNA-weighted means at the fund-month level, using only non-missing observations for each.
   - Sum TNA by fund-month.
   - Merge fund-month level variables into single dataset.
7. Clean fund-month level dataset.
   - Keep only if lagged TNA in 2017 dollars is in excess of 15 million.
   - If lagged TNA missing, drop if current TNA is below 15 million in 2017 dollars.
   - Drop if current TNA is zero.
   - Drop if expense ratio is below 0.1%.
   - Keep only if a fund has twelve months of available observations.

F.2 CRSP US Stock Database

I download the following datasets from WRDS: MSF, STOCKNAMES. The following is a summary of the steps I took for cleaning the data. - Set prc = -99 to missing, and replace the variable with its absolute value. - Keep if prc and shrout are non-missing. - Overlap match to STOCKNAMES on crsp_fundno and namedt <= caldt <= nameendt|. - Keep only common stock (shrcd 10 or 11).

F.3 Thomson Reuters Holdings Database

I download from WRDS data on fund share holdings from the Thomson Reuters Mutual Fund and Investment Company Common Stock Holdings Database (henceforth referred to as Thomson). The first holdings are reported as of 1979m12, but there is a jump in the number of reporting funds in 1980m3, which I consider the first complete report date.

While Thomson collects holdings quarterly (on `fdate` "file dates"; these represent "vintages" of Thomson's data feed), funds do not consistently report holdings at that frequency. The Thomson dataset notes the date on which funds reported their holdings, and hence the date for which the holdings are valid, as `rdate` at each data vintage. Share holdings are adjusted for splits at the vintage date. A recent WRDS investigation suggests that Thomson occasionally applies the split adjustment inappropriately, causing inconsistencies in holdings around stock splits. Part of this problem occurs while Thomson carries forward stale holdings data across different vintages. To the extent that the problematic split adjustments are linked to stale data, I avoid the issue by excluding stale reports as follows.

At each vintage date, Thomson provides the most recently reported holdings of each fund, even if the fund has not furnished an updated holdings report since the last vintage date. I handle this "stale data problem" by keeping only the first Thomson vintage for each holdings report. In the absence of an intervening fresh holdings report, I consider each report to represent the fund's buy-and-hold portfolio for the six months following the report. Reports can be filed on any day of the month. Since I am constructing a fund-month-level dataset, I assign each report to the month during which it was filed, regardless of the exact report day. This amounts to treating each report as if it was made at the end of month.

WRDS delivers the Thomson holdings data in two parts. The S12 Type 1 table includes fund report dates, data vintage dates, as well as investment objective codes. The S12 Type 3 table includes share holdings for each data vintage, including stale data. The Type 3 table identifies shares by `cusip`. After selecting the first vintage for each report date in the Type 1 table and discarding all subsequent vintages, I link the Type 1 and Type 3 tables on fund identifiers and vintage date (`fundno-fdate`).

F.4 MFLINKS

MFLINKS, originally developed by Wermers (2000), links CRSP share classes with fund level portfolio holdings in Thomson. The database was updated by Cao and Xue (2015), and currently it runs until the end of 2016.

The database links share classes in the CRSP mutual fund database, identified uniquely by `crsp_fundno`, to unique fund identifiers `wficn`. (There are six instances where a `crsp_fundno` is not uniquely matched to a `wficn`; I drop these observations.) On the Thomson side, it maps each `fundno-fdate` pair to the appropriate `wficn`.

F.5 Calculating Fund Portfolio Weights

F.5.1 Linking CUSIP and `permno`

I use the `STOCKNAMES` file to produce a linking table between CUSIP and `permno` in order to combine share Thomson share holding data with CRSP stock data. I take `ncusip` as the relevant identifier in this file; when it is missing, I fill it in with the appropriate `cusip`. I then generate all unique pairs between the (filled-in) `ncusip` and `permno`. This algorithm produces a one to many mapping between (filled-in) `ncusip` and `permno`.

F.5.2 Applying share adjustment

Holdings in Thomson are adjusted for splits as of the vintage date. I therefore match holdings to CRSP cumulative share adjustment factors (`cfacshr`) by vintage date as `shares=shares*cfacshr`. I retain only share holdings which can be matched to records in CRSP classified as US common equity.

F.5.3 Prices

After adjusting shares, I forward fill portfolio holdings for up to 6 months between report dates, simulating a buy-and-hold strategy. I then match holdings with prices and cumulative price adjustment factors (`cfacpr`) on the (forward-filled) report dates, and calculate holding values as `shares*prc/cfacpr` at each date, and use these to calculate portfolio weights.

F.6 Identifying actively managed domestic equity funds

The following steps describe the algorithm for identifying actively managed US equity funds.

1. Exclude funds without a `wficn` fund identifier from MFLINKS.
2. Exclude funds that do not meet the data availability, size, or expense ratio requirements in the CRSP mutual fund dataset.
3. Exclude funds with investment objective codes (`ioc`) in the Thomson Reuters S12 Type 1 database of the following type:
    - `ioc = 1` International
    - `ioc = 5` Municipal Bond
    - `ioc = 6` Bond & Preferred
    - `ioc = 7` Balanced
    - `ioc = 8` Metals Remaining funds either have a missing investment objective code, or are categorized as `ioc = 2` Aggressive Growth, `ioc==3` Growth, `ioc = 4` Growth & Income, `ioc = 9` Unclassified.

4. Drop if any share class of the fund is ever classified as an index fund by CRSP (`index_fund_flag != ""`), or if the name of any of the fund's share classes ever contains "index" in the CRSP database.
5. Exclude if any of the fund's share classes is ever assigned any of the following policy code (`policy`) in CRSP:
    - `Bal` Balanced fund
    - `B & P` Bonds and preferred stock (mainly convertible funds)
    - `Bonds` Bonds
    - `C & I` Canadian and international
    - `Hedge` Leverage and/or short selling
    - `Leases` Holds equity in lease contracts
    - `GS` Government securities
    - `MM` Money market
    - `Pfd` Preferred stock
    - `Spec` Sector or highly speculative fund
    - `TF` Tax free fund
    - `TFE` Tax-free exchange fund
    - `TFM` Tax free money market fund
6. Use CRSP objective codes (`crsp_obj_cd`) to exclude target date funds and sector funds, as follows.
    - Target date funds: exclude fund if any share class ever has `crsp_obj_cd` equal to `MT`
    - Sector funds: exclude fund if any share class is ever has `crsp_obj_cd` beginning with `EDS`
7. Exclude funds with any share class whose lower case name ever includes any of the following geographical and allocation-related strings
    - `internat`
    - `euro`
    - `japan`
    - `emerging market`
    - `balanced`
    - `bond fund`
8. To further ensure the exclusion of target date funds, drop funds with any share classes whose lower case name ever includes any of the following strings
    - `20[0-9][0-9]`
    - `retire`
    - `target`
9. Exclude Tax-managed funds with any share classes whose lower case name ever includes `tax-` or `tax`, followed by any of `manage`, `efficien`, `exempt`, `smart`, `advantage`, `aware`, or `sensitive`.

10. After the above filters are applied, I define domestic equity funds constructively through the following procedure. Consider a share class domestic equity if

    i. Lipper class indicates domestic equity. I consider this to be the case if `lipper_class` is any of

        - `EIEI` Equity income
        - `G` Growth
        - `LCCE` Large cap core
        - `LCGE` Large cap growth
        - `LCVE` Large cap value
        - `MCCE` Mid cap core
        - `MCGE` Mid cap growth
        - `MCVE` Mid cap value
        - `MLCE` Multi cap core
        - `MLGE` Multi cap growth
        - `MLVE` Multi cap value
        - `SCCE` Small cap core
        - `SCGE` Small cap growth
        - `SCVE` Small cap value

    ii. If Lipper class is unavailable, rely on Strategic Insights Objective Codes to identify equity funds. I flag funds as domestic equity if `si_obj_cd` is any of

        - `AGG` Equity US aggressive growth
        - `GMC` Equity US midcaps
        - `GRI` Equity US growth & income
        - `GRO` Equity US growth
        - `ING` Equity US income & growth

        - `SCG` Equity US small companies

    iii. If neither Lipper class nor Strategic Insights Objective Code is available, I use Wiesenberger Fund Type codes. I flag funds as domestic equity if `wbrger_obj_cd` is any of - `G` Growth - `G-I` Growth-income - `GCI` Growth and current income - `LTG` Long term growth - `MCG` Maximum capital gains - `SCG` Small cap growth

I include a fund if any of its share classes is flagged as domestic equity by the above procedure at some point.

11. Drop if any share classes ever classified as fixed income, i.e. with `crsp_obj_cd` beginning with `I`
12. Drop funds if it is classified as a foreign fund (`crsp_obj_cd` beginning with `EF`) over 25% of the time.

REFERENCES

Antón, M. and Polk, C. (2014). Connected stocks. *Journal of Finance*, 69(3):1099–1127.

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., van Binsbergen, J. H., and Liu, B. (2017). Matching capital and labor. *Journal of Finance*, 72(6):2467–2504.

Blocher, J. (2016). Network externalities in mutual funds. *Journal of Financial Markets*, 30:1–26.

Busse, J. A., Chordia, T., Jiang, L., and Tang, Y. (2017). Mutual fund trading costs. Working paper. Available at SSRN: https://ssrn.com/abstract=2350583.

Cao, B. and Xue, J. (2015). Mutual fund links update procedures. WRDS Research Note. Available at: https://wrds-web.wharton.upenn.edu/wrds/support/Data/_001Manuals%20and%20Overviews/_090MFLINKS/Mutual%20Fund%20Links%20Update%20Procedure.pdf.cfm.

Chamberlain, G. (1982). Multivariate regression models for panel data. *Journal of Econometrics*, 18(1):5–46.

Chen, J., Hong, H., Huang, M., and Kubik, J. D. (2004). Does fund size erode mutual fund performance? The role of liquidity and organization. *American Economic Review*, 94(5):1276–1302.

Cohen, L. H., Malloy, C. J., and Nguyen, Q. H. (2016). Lazy prices. Working paper. Available at SSRN: https://ssrn.com/abstract=1658471.

Cohen, R. B., Coval, J. D., and Pástor, Ľuboš. (2005). Judging fund managers by the company they keep. *Journal of Finance*, 60(3):1057–1096.

Coval, J. D. and Stafford, E. (2007). Asset fire sales (and purchases) in equity markets. *Journal of Financial Economics*, 86(2):479–512.

Cremers, M. (2017). Active share and the three pillars of active management: Skill, conviction and opportunity. *Financial Analysts Journal*, 73(2):61–79.

Cremers, M. and Petajisto, A. (2009). How active is your fund manager? A new measure that predicts performance. *Review of Financial Studies*, 22(9):3329–3365.

Cremers, M., Petajisto, A., and Zitzewitz, E. W. (2012). Should benchmark indices have alpha? Revisiting performance evaluation. *Critical Finance Review*, 2:1–48.

Evans, R. (2000). Mutual fund incubation. *Journal of Finance*, 65(4):1581–1611.

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

French, K. R. (2008). Presidential address: The cost of active investing. *Journal of Finance*, 63(4):1537–1573.

Gaure, S. (2013). lfe: Linear group fixed effects. *The R Journal*, 5(2):104–117. User documentation of the 'lfe' package.

Gruber, M. J. (1996). Another puzzle: The growth in actively managed mutual funds. *Journal of Finance*, 51(3):783–810.

Harvey, C. R. and Liu, Y. (2017). Decreasing returns to scale, fund flows, and performance. Duke I&E Research Paper No. 2017-13. Available at SSRN: https://ssrn.com/abstract=2990737.

Hoberg, G., Kumar, N., and Prabhala, N. (2018). Mutual fund competition, managerial skill, and alpha persistence. *Review of Financial Studies*, 31(5):1896–1929.

Houge, T. and Wellman, J. (2005). Fallout from the mutual fund trading scandal. *Journal of Business Ethics*, 62:129–139.

Jensen, M. C. (1968). The performance of mutual funds in the period 1945–1964. *Journal of Finance*, 23(2):389–416.

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

Kacperczyk, M., Sialm, C., and Zheng, L. (2005). On the industry concentration of actively managed equity mutual funds. *Journal of Finance*, 60(4):1983–2011.

Kacperczyk, M., Sialm, C., and Zheng, L. (2008). Unobserved actions of mutual funds. *Review of Financial Studies*, 21(6):2379–2416.

Khan, M., Kogan, L., and Serafeim, G. (2012). Mutual fund trading pressure: Firm-level stock price impact and timing of SEOs. *Journal of Finance*, 68(4):1371–1395.

Lou, D. (2012). A flow-based explanation for return predictability. *Review of Financial Studies*, 25(12):3457–3489.

Malkiel, B. G. (1995). Returns from investing in equity mutual funds, 1971–1991. *Journal of Finance*, 50(2):549–572.

Malkiel, B. G. (2013). Asset management fees and the growth of finance. *Journal of Economic Perspectives*, 27(2):97–108.

McCabe, P. (2008). The economics of the mutual fund trading scandal. Finance and economics discussion series, Federal Reserve Board, Washington D.C.

McLemore, P. (2016). Do mutual funds have decreasing returns to scale? evidence from fund mergers. Working paper. Available at SSRN: https://ssrn.com/abstract=2490824.

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

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

Pástor, v., Stambaugh, R. F., and Taylor, L. A. (2017a). Do funds make more when they trade more? *Journal of Finance*, 72(4):1483–1528.

Pástor, v., Stambaugh, R. F., and Taylor, L. A. (2017b). Fund tradeoffs. Working paper. Available at SSRN: https://ssrn.com/abstract=3011700.

Petajisto, A. (2013). Active share and mutual fund performance. *Financial Analysts Journal*, 69(4):73–93.

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

Reuter, J. and Zitzewitz, E. W. (2015). How much does size erode mutual fund performance? a regression discontinuity approach. Working paper. Available at SSRN: https://ssrn.com/abstract=1661447.

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

Wermers, R. (2000). Mutual fund performance: An empirical decomposition into stock-picking talent, style, transaction costs, and expenses. *Journal of Finance*, 55(4):1655–1695.

Zhu, Q. (2017). The missing new funds. Working paper. Available at SSRN: https://ssrn.com/abstract=3004268.

Zitzewitz, E. W. (2006). How widespread was late trading in mutual funds? *American Economic Review Papers and Proceedings*, 96(2):284–289.

Zitzewitz, E. W. (2009). Prosecutorial discretion in mutual fund settlement negotiations, 2003-7. *The B.E. Journal of Economic Analysis & Policy*, 9(1):Article 24.