NBER WORKING PAPER SERIES

FUND TRADEOFFS

Lubos Pastor
Robert F. Stambaugh
Lucian A. Taylor

Working Paper 23670
http://www.nber.org/papers/w23670

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2017, Revised October 2019

This paper previously circulated under the title "Portfolio Liquidity and Diversification: Theory and Evidence." The views in this paper are the responsibility of the authors, not the institutions they are affiliated with, nor of the National Bureau of Economic Research. We are grateful for comments from our conference discussants Daniel Andrei, James Choi, Wayne Ferson, Mark Grinblatt, Paul Irvine, Marcin Kacperczyk, Clemens Sialm, Ashish Tiwari, and Mungo Wilson. We are also grateful for comments from Jonathan Berk, Gene Fama, Laszlo Jakab, Don Keim, Rob Vishny, Jeff Wurgler, and Yao Zeng, conference participants at the AIM Investment Conference, AQR Insight Award Finalists Competition, Conference on Professional Asset Management (Erasmus), Duke/UNC Asset Pricing Conference, ESSFM Gerzensee, European Finance Association, FARFE (MIT), Five-Star Conference on Research in Finance (NYU Stern), GSU-CEAR Finance Conference, NBER Conference on New Developments in Long-Term Asset Management, Q Group, Western Finance Association, and seminar participants at Boston University, Tulane University, University of Chicago, University of Illinois at Urbana-Champaign, University of Pennsylvania, University of Utah, and WU Vienna. We are also grateful to Will Cassidy, Yeguang Chi, and Pierre Jaffard for superb research assistance. This research was funded in part by the Fama-Miller Center for Research in Finance and the Center for Research in Security Prices at Chicago Booth.

At least one co-author has disclosed a financial relationship of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w23670.ack

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2017 by Lubos Pastor, Robert F. Stambaugh, and Lucian A. Taylor. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Fund Tradeoffs

Lubos Pastor, Robert F. Stambaugh, and Lucian A. Taylor

NBER Working Paper No. 23670
August 2017, Revised October 2019
JEL No. G11,G23

ABSTRACT

We study tradeoffs among active mutual funds' characteristics. In both our equilibrium model and the data, funds with larger size, lower expense ratio, and higher turnover hold more-liquid portfolios. Portfolio liquidity, a concept introduced here, depends not only on the liquidity of the portfolio's holdings but also on the portfolio's diversification. We also confirm other model-predicted tradeoffs: Larger funds are cheaper. Larger and cheaper funds are less active, based on our new measure of activeness. Better-diversified funds hold less-liquid stocks; they are also larger, cheaper, and trade more. These tradeoffs provide novel evidence of diseconomies of scale in active management.

Lubos Pastor
University of Chicago
Booth School of Business
5807 South Woodlawn Ave
Chicago, IL 60637
and NBER
lubos.pastor@chicagobooth.edu

Robert F. Stambaugh
Finance Department
The Wharton School
University of Pennsylvania
Philadelphia, PA 19104-6367
and NBER
stambaugh@wharton.upenn.edu

Lucian A. Taylor
Finance Department
The Wharton School
University of Pennsylvania
2300 Steinberg Hall - Dietrich Hall
3620 Locust Walk
Philadelphia, PA 19104-6367
luket@wharton.upenn.edu

1. Introduction

Mutual funds manage tens of trillions of dollars. Through their investment decisions, these funds play a major role in allocating capital in the economy. Myriad studies investigate fund performance, but few address fund characteristics. We show that fund characteristics, particularly the tradeoffs among them, provide new insights into the economics of mutual funds. Chief among these insights are novel perspectives on diseconomies of scale in active asset management. Assuming such diseconomies, Berk and Green (2004) argue that investors allocate money across funds such that, in equilibrium, each fund has zero expected performance relative to a passive benchmark. If expected performance is always zero, then investigating scale diseconomies by empirically linking fund size with performance faces an inherent challenge. While multiple studies report a negative size-performance relation, the evidence is fairly sensitive to the methodological approach.[1] In contrast, we argue that relations among fund characteristics offer strong evidence of scale diseconomies in active management. For example, we show that larger funds tend to trade less and hold more-liquid portfolios, a clear indication of decreasing returns to scale.

We derive equilibrium relations among four key fund characteristics: fund size, expense ratio, turnover, and portfolio liquidity. This last characteristic is novel. While the literature presents a variety of liquidity measures for individual securities, it offers little guidance for assessing liquidity at the portfolio level. We introduce the concept of portfolio liquidity, and we show that diseconomies of scale lead funds to trade off this characteristic against others in important ways. Our measure of portfolio liquidity is derived theoretically based on the simple idea that a portfolio is more liquid if it has lower trading costs. Specifically, if one trades equal dollar amounts of two portfolios, the portfolio with lower trading costs has greater liquidity.

Our equilibrium model relates portfolio liquidity to fund size, expense ratio, and turnover. When choosing its characteristics, a fund recognizes that lower liquidity and higher turnover raise expected gross profits but also raise transaction costs. Those costs increase in the fund's size as well. This role of fund size is recognized by investors when they decide how much capital to allocate to the fund, as in Berk and Green (2004).

The model implies a novel link between the four key mutual fund characteristics. Funds with larger size, lower expense ratios, and higher turnover should have more-liquid portfolios. We investigate these implications in a sample of 2,789 active U.S. equity mutual funds from

[1] See, for example, Pástor, Stambaugh, and Taylor (2015), Reuter and Zitzewitz (2015), and Zhu (2018). For additional evidence on returns to scale in mutual funds, see Chen et al. (2004), Bris et al. (2007), Pollet and Wilson (2008), Yan (2008), and Harvey and Liu (2017), among others.

1979 through 2014. When we estimate the cross-sectional regression of portfolio liquidity on fund size, expense ratio, and turnover in our panel dataset, we find strong support for the model. All three slopes have their predicted signs and are highly significant, both economically and statistically. Funds that are larger, less expensive, and trade more tend to hold more-liquid portfolios, as the model predicts. According to our model, these tradeoffs are induced by diseconomies of scale. Intuitively, a large fund optimally reduces its trading costs by trading less and holding a more-liquid portfolio.

A fund's size trades off negatively with its expense ratio in our model. As in Berk and Green (2004), a fund's fee revenue is determined by its skill, so charging a higher expense ratio simply dictates a smaller fund. We find strong evidence of this tradeoff—a large negative correlation between fund size and expense ratio—in the data.

With decreasing returns to scale, a fund also faces a tradeoff between its size and how actively it pursues profit opportunities. Our model delivers a novel measure of a fund's activeness that combines the fund's turnover with its portfolio liquidity. The latter characteristic depends on the portfolio's weights versus the benchmark, with a less-liquid portfolio being more active. In that respect our measure of activeness resembles the popular active share measure of Cremers and Petajisto (2009). Portfolio holdings are only part of the story, though. In our model, a fund is also more active if it trades more. The model implies that smaller funds as well as more expensive funds should be more active, and we find evidence of both tradeoffs in the data. Similar properties for the turnover component of activeness also emerge strongly in the data: funds that trade less are larger and cheaper, both across funds and over time.

A fund's scale is typically equated to its size, measured as assets under management (AUM). Our study implies a new concept of scale, which depends not only on the fund's size but also on its activeness. In our model, funds face decreasing returns to scale, but the implied measure of scale is size times activeness, not simply size. This idea makes sense. If two funds manage equal amounts of money, but one of them deploys its money more actively, it seems reasonable to view that fund as operating at a larger scale, essentially leaving a bigger footprint in the market.

In deriving our measure of portfolio liquidity, we apply a familiar concept: less-liquid assets are costlier to trade. We extend this concept to portfolios, viewing a portfolio as an asset and thereby considering the cost of trading the portfolio as a given basket of securities. When assessing portfolio liquidity, it seems natural to consider the average liquidity of the portfolio's constituents. For example, portfolios of small-cap stocks tend to be less liquid than portfolios of large-cap stocks. While this assessment is a useful starting point, it is

incomplete. We show that a portfolio's liquidity depends not only on the liquidity of the stocks held in the portfolio but also on the degree to which the portfolio is diversified:

$$\text{Portfolio Liquidity} = \text{Stock Liquidity} \times \text{Diversification} \ . \tag{1}$$

The more diversified a portfolio, the less costly is trading a given fraction of it. For example, a fund trading just 1 stock will incur higher costs than a fund spreading the same dollar amount of trading over 100 stocks, even if all of the stocks are equally liquid. Throughout, we focus on equity portfolios, but these ideas are more general.

Our measure of portfolio liquidity is easy to calculate from the portfolio's composition. Following equation (1), our measure has two components. The first, stock liquidity, is the average liquidity of the portfolio's holdings. The second component, diversification, has its own intuitive decomposition:

$$\text{Diversification} = \text{Coverage} \times \text{Balance} \ . \tag{2}$$

Coverage reflects the number of stocks in the portfolio. Portfolios holding more stocks have greater coverage. Balance reflects how the portfolio weights the stocks it holds. Portfolios with weights closer to market-cap weights have greater balance.

Diversification's role in portfolio liquidity is important empirically. We compute our measures of portfolio liquidity and diversification for the mutual funds in our sample. We find that fund portfolios have become more liquid over time, from 1979 through 2014. Average portfolio liquidity almost doubled over the sample period, driven by diversification. Diversification quadrupled, as both of its components in equation (2) rose steadily. Coverage rose because the number of stocks held by the average fund grew from 54 to 126. Balance rose because funds' portfolio weights increasingly resembled market-cap weights.[2]

Our model predicts tradeoffs between diversification and other fund characteristics. In equilibrium, funds with more-diversified portfolios should be larger and cheaper, they should trade more, and their stock holdings should be less liquid. We find strong empirical support for all four predictions. The negative relation between diversification and stock liquidity implies that these components of portfolio liquidity are substitutes: funds holding less-liquid stocks make up for it by diversifying more, and vice versa. The components of diversification, coverage and balance, are also substitutes: portfolios with lower coverage tend to be better balanced, and vice versa. Both substitution effects are predicted by our model.

Funds trading less-liquid portfolios will likely find greater profit opportunities, before trading costs, but less-liquid portfolios are also costlier to trade. This role of liquidity

[2]The increased resemblance of active funds' portfolios to the market benchmark is also apparent from measures such as active share and tracking error (e.g., Cremers and Petajisto, 2009, and Stambaugh, 2014).

in the profit-cost tradeoff has implications for the turnover-performance relation analyzed by Pástor, Stambaugh, and Taylor (2017). Specifically, the time-series relation between a fund's turnover and its subsequent performance should be stronger for funds with less-liquid portfolios. We find this implication holds strongly in the data.

Our focus on fund characteristics, and the tradeoffs among them, seems novel. Some of the concepts we address are familiar, however. One strand of related literature studies returns to scale in active management. This literature explores the hypothesis that as a fund's size increases, its ability to outperform its benchmark declines (Berk and Green, 2004).[3] This hypothesis is motivated by liquidity constraints. Being larger erodes performance because a larger fund trades larger dollar amounts, and trading larger dollar amounts incurs higher proportional trading costs. The hypothesis has received some empirical support. Several studies report that fund size negatively predicts fund performance, but the evidence is somewhat sensitive to the methodology applied, as discussed earlier. Unlike this prior literature, we do not examine fund performance. Our analysis of tradeoffs among fund characteristics reveals different evidence of decreasing returns to scale. We find that larger funds tend to have lower turnover and higher portfolio liquidity. This evidence is in line with our model, in which diseconomies of scale lead larger funds to trade less and hold more-liquid portfolios, either by holding more-liquid stocks or by diversifying more. Our results represent strong evidence of decreasing returns to scale, with a refined notion of scale, as explained earlier.

Two other studies provide related evidence on returns to scale. Pollet and Wilson (2008) find that mutual funds respond to asset growth mostly by scaling up existing holdings rather than by increasing the number of stocks held. But the authors also find that larger funds and small-cap funds are less reluctant to diversify in response to growth, exactly as our theory predicts. In their comprehensive analysis of mutual fund trading costs, Busse et al. (2017) report that larger funds trade less and hold more-liquid stocks. This evidence, which overlaps with our findings, also supports our model. In the language of equation (1), Busse et al. show that larger funds have higher stock liquidity; we show they also have higher diversification. The evidence of Busse et al. is based on a sample much smaller than ours (583 funds in 1999 through 2011), dictated by their focus on trading costs. Neither Busse et al. nor Pollet and Wilson do any theoretical analysis.

Our study is also related to the literature on portfolio diversification. The implications of diversification for risk are well understood. We show that diversification also has impor-

[3]This is the hypothesis of fund-level decreasing returns to scale. A complementary hypothesis of industry-level decreasing returns to scale is that as the size of the active mutual fund industry increases, the ability of any given fund to outperform declines (see Pástor and Stambaugh, 2012, and Pástor, Stambaugh, and Taylor, 2015). In this paper, we focus on the fund-level hypothesis.

tant implications for transaction costs and diseconomies of scale in active management. We propose a new measure of diversification that has strong theoretical motivation. Our measure blends features of two ad-hoc measures, the number of stocks held and the Herfindahl index of portfolio weights. By using our measure, we show that mutual funds have become substantially more diversified over time, yet their diversification remains relatively low.[4] We also derive predictions for the determinants of diversification. Funds with more-diversified portfolios should be larger and cheaper, they should trade more, and their holdings should be less liquid, on average. We find strong empirical support for all of these predictions.

The rest of the paper is organized as follows. Section 2 introduces our measure of portfolio liquidity. Section 3 examines the tradeoffs among fund characteristics. Section 4 analyzes tradeoffs that involve the components of portfolio liquidity, including diversification. Section 5 addresses tradeoffs that involve fund activeness. Section 6 explores fund tradeoffs in the form of simple correlations. Section 7 rethinks the concept of scale. Section 8 contrasts our setting with that of Berk and Green (2004). Section 9 analyzes the turnover-performance relation. Section 10 concludes. Additional material is in Appendices A and B. Further empirical results are in the Internet Appendix, which is available on the authors' websites.

2. Introducing Portfolio Liquidity

The definition of portfolio liquidity is based on trading costs: If one trades the same dollar amounts of two portfolios, the portfolio generating lower trading costs has greater liquidity. We show that this fundamental concept is captured by the following measure:

$$L = \left(\sum_{i=1}^{N} \frac{w_i^2}{m_i} \right)^{-1}, \tag{3}$$

where N is the number of stocks in the portfolio, w_i is the portfolio's weight on stock i, and m_i denotes the weight on stock i in a market-cap-weighted benchmark portfolio. The latter portfolio can be the overall market, the most familiar benchmark, or it can be the portfolio of all stocks in the sector in which the portfolio is focused, such as large-cap growth. We apply both choices in our empirical analysis of active mutual funds.[5]

To derive our measure, we begin with the familiar concept that less-liquid assets are costlier to trade. We apply this concept to portfolios by considering the cost of trading the

[4]Low diversification by institutional investors is also reported by Kacperczyk, Sialm, and Zheng (2005), Pollet and Wilson (2008), and others. Household portfolios also exhibit low diversification, as shown by Blume and Friend (1975), Polkovnichenko (2005), Goetzmann and Kumar (2008), and others.

[5]In our main regression results, reported in Tables 1 through 5, L is always defined with respect to the fund's sector benchmark, such as large-cap growth, because we include sector-quarter fixed effects.

portfolio as a whole, as if it were just another asset. That is, if D is the total dollar amount traded of the portfolio, the dollar amount traded of stock i is

$$D_i = Dw_i. \tag{4}$$

The corresponding total trading cost is

$$C = \sum_{i=1}^{N} D_i \, C_i \ , \tag{5}$$

where C_i is the cost per dollar traded of stock i. We assume C_i is larger when trading a larger fraction of M_i, the market capitalization of stock i. Specifically,

$$C_i = c \, \frac{D_i}{M_i} \ , \tag{6}$$

where the positive constant c is identical across the stocks in the benchmark. For example, if the benchmark is a small-cap value style index, c is the same for all stocks in that style, but stocks of different styles can have different values of c.[6] Equation (6) reflects the basic idea that larger trades have higher proportional trading costs, such as price impact. This idea has strong empirical support (e.g., Keim and Madhavan, 1997). We also consider two other ways of capturing the same idea: in Section 3.4, we modify equation (6) to allow for either nonlinearity or dependence on the stock's trading volume.

Combining equations (4) through (6), we can rewrite the total trading cost as

$$C = (c/M) \ D^2 \ \underbrace{\left(\sum_{i=1}^{N} \frac{w_i^2}{m_i} \right)}_{L^{-1}} \ , \tag{7}$$

where $m_i = M_i/M$. We define M as the market capitalization of all stocks in the benchmark portfolio, which allows L to be compared across portfolios having the same benchmark. Equation (7) shows that the expression for portfolio liquidity, given by equation (3), arises from trading costs. Trading a given dollar amount, D, of a portfolio with lower liquidity, L, incurs a greater total cost, C.

Portfolio liquidity is a characteristic that does not hinge on the trading behavior of whoever might hold the portfolio. For example, we do not assume any mutual fund actually trades each stock according to equation (4). The same portfolio can be held by two different funds trading in different ways, yet there is only one portfolio-specific value of L. In this respect we maintain the perspective on liquidity that is widely accepted for individual stocks. Measures of a stock's liquidity, such as its bid-ask spread or its turnover, do not hinge on

[6] Even within the same style, liquidity varies across stocks, as we explain in Section 4.1.

the behavior of whoever is trading the stock. Two investors trading equal amounts of the same stock often incur different costs, depending on how patiently they trade, how they execute their trades, etc. Nevertheless, less-liquid stocks are generally assumed to be costlier to trade. We simply assume the same about portfolios. In our analysis of fund tradeoffs in the following section, we assume that funds holding portfolios with lower L incur higher trading costs, ceteris paribus, regardless of how these funds trade.

Our measure of portfolio liquidity takes values between 0 and 1. The least liquid portfolio is fully invested in a single stock, the one with the smallest market capitalization among stocks in the benchmark. The liquidity of this portfolio is equal to the benchmark's market-cap weight on that smallest stock, so L can be nearly 0. A portfolio can be no more liquid than its benchmark, for which $L = 1$. This statement is proven in Appendix A, but its simple intuition follows from equation (6). When trading a given dollar amount of the benchmark portfolio, which has market-cap weights, the proportional cost of trading each stock is equal across stocks. With this cost denoted by κ, the proportional cost of the overall trade is also κ. If the benchmark portfolio is perturbed by buying one stock and selling another, then more weight is put on a stock whose proportional cost is now greater than κ, and less weight is put on a stock whose proportional cost is now smaller than κ. Therefore, the proportional cost of trading the same dollar amount of this alternative portfolio exceeds κ.

Another important property of our portfolio liquidity measure L in equation (3) is that it is increasing in the portfolio's diversification, as indicated in equation (1). Better-diversified portfolios are more liquid. We clarify the role of diversification in Section 4.

3. Tradeoffs Among Fund Characteristics

In this section, we examine the relations among key characteristics of active funds: portfolio liquidity, fund size, expense ratio, and turnover. We first derive such relations theoretically, from optimizing behavior of fund managers and investors. We then verify these relations empirically.

3.1. Fund Characteristics in Equilibrium

Consider an actively managed mutual fund over a single period. We assume the fund's expected total trading cost over this period is given by the function

$$C(A, T, L) = \theta A^{\gamma} T^{\lambda} L^{-\phi} \ , \tag{8}$$

where A is the fund's size (AUM), T is its turnover (dollar amount traded divided by AUM), and L is the liquidity of the fund's portfolio. The constants θ, γ, λ, and ϕ are positive. We assume $\gamma > 1$, so trading costs are increasing and convex in A, capturing this familiar role of fund size modeled by Berk and Green (2004). Unlike that study, we let trading costs depend not only on A but also on T and L. The relevance of T for equation (8) is clear, because a fund that trades more incurs higher trading costs, holding A and L constant.

The presence of portfolio liquidity, L, in equation (8) is motivated by the arguments from Section 2. Recall from equation (7) that a portfolio with a lower L incurs higher trading costs, holding fixed the fund's traded dollar amount, $D = AT$. Saying that it is costlier to trade a portfolio with a lower L is like saying it is costlier to trade a stock with a higher bid-ask spread or any other illiquidity measure. Saying that such a stock is costlier to trade does not assume one trades the stock in a particular way, e.g., trading at the bid and ask quotes. In the same vein, saying that a fund with a lower L incurs higher trading costs does not assume the fund trades its portfolio in a particular way, e.g., trading stocks in proportion to portfolio weights. We assign a negative but flexible role to L in determining trading costs (as well as flexible positive roles to A and T) by treating the exponents in equation (8) as free parameters. That equation nests equation (7) as a special case when $\theta = c/M$, $\gamma = 2$, $\lambda = 2$, and $\phi = 1$. In other words, the simple trading-cost function derived earlier motivates our assumption that costs increase convexly in A, increase in T, and decrease in L.

The roles of A, T, and L in equation (8) receive some empirical support from Edelen, Evans, and Kadlec (2013), who construct a proxy for fund trading costs, "position-adjusted turnover." This proxy is increasing in both T and A and decreasing in the number of holdings, a component of L. The proxy also takes into account the fund's market-cap category. The authors show that position-adjusted turnover is related to other fund characteristics in ways consistent with its interpretation as a measure of trading costs.

The fund's expected gross return, before fees and trading costs, depends on the fund's skill as well as how actively that skill is applied. To capture this interaction, we model the expected benchmark-adjusted gross return as

$$a = \mu\, g(T, L) \ , \tag{9}$$

where μ is a fund-specific positive constant reflecting skill in identifying profitable trading opportunities.[7] How actively the fund applies that skill is represented by the function $g(T, L)$. The roles of T and L in equation (9) are discussed in more detail in Section 5. Like Berk and

[7]While it is simplest to think of fund skill as known and equal to μ, our framework easily accommodates unknown skill. In general, μ can be interpreted as skill perceived by both fund managers and investors.

Green (2004), we assume a fund's size directly affects its trading costs but not its expected gross return, so A does not appear in equation (9).

An actively managed fund chooses its portfolio liquidity, L, turnover, T, and expense ratio, f. The fund's objective is to maximize its total revenue,

$$F = fA \ , \tag{10}$$

where f is the fund's expense ratio, i.e., the fee rate charged by the fund to its investors, expressed as a fraction of AUM. The fund's size, A, is chosen by investors. Our basic equilibrium relation among fund characteristics follows from the fund's first-order condition for f. We address the optimal choice of T and L in Section 5. Given T and L, denote the fund's proportional trading cost as a function of A by

$$q(A;T,L) = C(A,T,L)/A \ . \tag{11}$$

The fund's expected benchmark-adjusted net return to investors equals

$$\alpha = a - q(A) - f \ , \tag{12}$$

where we temporarily suppress T and L in $q(\ )$ for notational ease.

We assume that, given the fund's decisions, investors choose A so that α in equilibrium is a constant, known to the fund. That constant could be $\alpha = 0$, as assumed by Berk and Green (2004), or any other $\alpha \neq 0$. A non-zero alpha could, for example, reflect compensation for search costs, as in Gârleanu and Pedersen (2018). The equilibrium value of α can also vary across funds. For example, α can vary across investment styles if the degree of competition among funds varies across styles (e.g., Hoberg, Kumar, and Prabhala, 2018).

Define $h = a - \alpha - f$, and note from equation (12) that $A = q^{-1}(h)$. Differentiating the right-hand side of equation (10) with respect to f gives

$$\begin{aligned} \frac{dF}{df} &= A + f\left[\frac{d}{dh}\left(q^{-1}(h)\right)\right]\frac{dh}{df} \\ &= A + f\left[\frac{1}{q'\left(q^{-1}(h)\right)}\right](-1) \\ &= A - f\left[\frac{1}{q'(A)}\right] \end{aligned} \tag{13}$$

Setting $\frac{dF}{df} = 0$ then gives

$$f = Aq'(A). \tag{14}$$

For the cost function in equation (8), it is easily verified that equation (14) implies

$$f = (\gamma - 1)q(A;T,L), \tag{15}$$

with T and L restored in the notation. Taking logs and rearranging, again using equation (8), delivers the equilibrium relation:

$$\ln L = \frac{1}{\phi}\ln[(\gamma-1)\theta] + \left(\frac{\gamma-1}{\phi}\right)\ln A - \frac{1}{\phi}\ln f + \left(\frac{\lambda}{\phi}\right)\ln T \ , \tag{16}$$

or

$$\ln L = b_0 + b_1 \ln A - b_2 \ln f + b_3 \ln T \ . \tag{17}$$

Funds may differ in their skill level, μ, and activeness function, $g(T, L)$, but we assume they face the same cost function in equation (8). This assumption implies that b_0, b_1, b_2, and b_3 are positive constants across funds. Equation (17) then captures the cross-sectional tradeoffs among the four fund characteristics: portfolio liquidity L, fund size A, expense ratio f, and turnover T. Funds that are larger, cheaper, and trade more have more liquid portfolios. These predictions stem from the novel result in equation (15) that links an optimizing fund's expense ratio to its equilibrium proportional trading cost, q. From this result, we can easily understand the implication of changing one variable on the right-hand side of equation (17) while holding the other two variables constant across funds. Holding f constant, thus keeping the proportional trading cost constant, and holding T constant, a larger fund size, A, dictates a more liquid portfolio to offset the otherwise higher proportional costs of trading larger amounts. Because a higher f implies a higher proportional trading cost in equilibrium, that higher cost must arise from lower liquidity if A and T are held constant. Finally, holding A and f constant, the greater cost of heavier trading (i.e., larger T) must be offset by holding a more liquid portfolio in order to maintain the same proportional trading cost.

Equation (17) is quite general. Its derivation specifies neither a functional form for $g(T, L)$ nor how T and L are chosen by the fund. In Section 5, we consider a specification for $g(T, L)$ that provides insights into the choices of T and L and implies tradeoffs between a fund's activeness and its size and fee rate.

3.2. Empirical Evidence

We analyze a sample of 2,789 actively managed U.S. domestic equity mutual funds covering the 1979–2014 period. To construct this sample, we begin with the dataset constructed by Pástor, Stambaugh, and Taylor (2015, 2017), which combines and cross-validates data from the Center for Research in Securities Prices (CRSP) and Morningstar. We add three years of data and merge in the Thomson Reuters dataset of fund holdings. We restrict the sample to fund-month observations whose Morningstar category falls within the traditional 3×3 style

box (small/mid/large-cap interacted with growth/blend/value). This restriction excludes non-equity funds, international funds, and industry-sector funds. We exclude index funds because our model is designed for active funds trying to outperform a benchmark. We also exclude funds of funds and funds smaller than $15 million. We classify each fund into one of nine sectors corresponding to Morningstar's 3×3 style box.[8]

For each fund, we compute its characteristics at each quarter-end. To compute fund size, we cross-verify monthly AUM between CRSP and Morningstar, as described in Pástor, Stambaugh, and Taylor (2015). We obtain annual data on funds' expense ratios and turnover from CRSP. Expense ratio includes the fund's administrative and management fees. Expense ratio does not include brokerage commissions or other transaction costs, so funds with higher trading costs do not mechanically charge a higher expense ratio. Turnover is the minimum of the fund's dollar purchases and sales during the fiscal year, scaled by the fund's average total net assets. This measure aims to exclude trades induced by fund inflows and outflows, thus capturing trades that are largely discretionary. This is the measure of turnover that funds report to the SEC. Following Pástor, Stambaugh, and Taylor (2017), we winsorize turnover at the 1st and 99th percentiles. A more detailed description of our sample is in Appendix B.

For each fund and quarter-end, we compute portfolio liquidity from the fund's quarterly holdings data. Initially, we compute portfolio liquidity by using the market portfolio as the benchmark. Our definition of the market portfolio includes ordinary common shares (CRSP share code with first digit equal to 1) and REIT shares of beneficial interest (CRSP share code of 48). This definition is guided by the end-of-sample holdings of the world's largest mutual fund, Vanguard's Total Stock Market Index fund, as we explain in Appendix B.

To test the predictions from equation (17), we estimate this equation as a panel regression of $\ln(L)$ on the other fund characteristics in our mutual fund dataset:

$$\ln L_{i,t} = b_{0,s,t} + b_1 \ln A_{i,t} - b_2 \ln f_{i,t} + b_3 \ln T_{i,t} + \epsilon_{i,t} \ , \tag{18}$$

where the subscripts denote fund i and quarter t (i.e., the unit of observation is fund/quarter). The regression specification (18) modifies equation (17) in several ways: it takes a stand on which quantities vary and which do not, it adds an error term, $\epsilon_{i,t}$, and it allows the intercept, $b_{0,s,t}$, to be sector-quarter specific. We explain these choices in the following paragraphs.

Fund characteristics A, f, T, and L vary across both funds and time. Some of this

[8]Morningstar assigns funds to style categories based on the funds' reported portfolio holdings, and it updates these assignments over time. Since the assignments are made by Morningstar rather than the funds themselves, there is no room for benchmark manipulation of the kind documented by Sensoy (2009). The benchmark assigned by Morningstar can differ from that reported in the fund's prospectus.

variation comes from heterogeneity in the exogenous skill parameter μ. For a given μ, the fund chooses its f, T, and L to maximize its fee revenue. Depending on the form of $g(T, L)$, there may also be multiple optimal choices of f, T, and L for a given value of μ.[9] Such a setting is presented in Section 5. In general our model allows the choices of f, T, and L to vary flexibly across funds. These choices dictate each fund's A, as investors allocate money to funds to achieve a given equilibrium α. We assume the cost function parameters γ, λ, and ϕ from equation (8) are constant across funds and time, and thus so are the regression slopes b_1, b_2, and b_3, by equations (16) and (17).

We add an error term, $\epsilon_{i,t}$, to the model-predicted equation (17), for two reasons. First, no model is perfect. Second, L could be measured with error because it is computed from fund holdings data, which omit small holdings. To avoid the attenuation bias associated with this potential measurement error, we include L on the left-hand side of our regression. Another reason to include L on the left-hand side is that it is the novel quantity, whose determinants we wish to investigate. We assume that $\epsilon_{i,t}$ is uncorrelated with the right-hand side variables in equation (18).

The intercept $b_{0,s,t}$ represents sector-quarter fixed effects, where s is the sector to which fund i is assigned by Morningstar. Including these fixed effects offers three benefits. First, we treat our model's predictions as cross-sectional, and the fixed effects isolate variation across funds in the same sector and quarter. In principle, one could also view our model as describing a given fund solving a series of single-period problems. However, applying the model to a fund's time series would confront the problem that two fund characteristics, expense ratio and turnover, are measured in a way that is poorly suited for time-series analysis: they are measured only annually, and expense ratios vary little even across years. Second, by including sector-quarter fixed effects, we effectively use L defined with respect to a sector-specific benchmark rather than the market.[10] Third, from equations (16) and (17), the intercept $b_{0,s,t}$ depends on the cost-function parameter θ, which reflects market liquidity. The fixed effects allow market liquidity to vary both over time and across sectors.

Finally, note that equation (17) does not represent a causal relation. Instead, it captures the equilibrium relation among the jointly determined, endogenous fund characteristics.

[9] In that case, a fund's specific choices of f, T, and L can be determined by various factors that are outside of, and yet consistent with, our model. For example, funds may choose a fee rate that is customary for their sector or family. Also, a given manager's knowledge and training may be best suited to a high-turnover strategy, so the manager would choose high T, which then dictates a high L or high f.

[10] Sector-benchmarked L is equal to market-benchmarked L divided by the fraction of the total stock market capitalization accounted for by the sector. Since that fraction is sector-specific within a given quarter, sector-benchmarked $\ln(L)$ is equal to market-benchmarked $\ln(L)$ minus a sector-quarter-specific constant that is absorbed by our fixed effects. Nonetheless, estimates of equation (17) that use only quarter fixed effects, equivalent to using market-benchmarked L, are quite similar (see the Internet Appendix).

Testing our model does not require that we estimate any causal relations.

Table 1 provides strong support for the model's predictions in equation (17). The slope coefficients on all three regressors have their predicted signs, not only for the multiple regression, which is implied by the model, but also for simple regressions. Moreover, all three slopes are highly significant in the multiple regression. The slope on fund size ($t = 13.76$) shows that larger funds tend to have more-liquid portfolios. A one-standard-deviation increase in the logarithm of fund size is associated with a 0.22 standard-deviation increase in $\ln(L)$ (sector- and quarter-adjusted). The slope on expense ratio ($t = -11.26$) shows that cheaper funds tend to have more-liquid portfolios. The economic significance of expense ratio is comparable to that of fund size: a one-standard-deviation increase in $\ln(f)$ is associated with a 0.24 standard-deviation decrease in $\ln(L)$. Finally, the slope on turnover ($t = 4.93$) shows that funds that trade more tend to have more-liquid portfolios. A one-standard-deviation increase in $\ln(T)$ is associated with a 0.10 standard-deviation increase in $\ln(L)$. We conclude that funds with less-liquid portfolios trade less and are smaller and more expensive, fully in line with our theory.

The results in Table 1 are not obvious. For example, one might surmise that funds whose portfolios are closer to passive benchmarks trade less because tracking an index requires little turnover. This hypothesis predicts a negative relation between L and T. Our finding of a positive relation rejects this hypothesis, but it supports our model.

3.3 Robustness

The results in Table 1 survive a number of robustness exercises that check consistency across subsamples. We construct subsamples by forming subperiods or by separating funds using characteristics not included in the model. Details for all of the subsample results are reported in the Internet Appendix. We divide the sample period into two subperiods, 1979 through 2004 and 2005 through 2014, which contain roughly the same number of fund-quarter observations. The counterparts of Table 1 for both subperiods look very similar to the original.

Next, we form subsamples by separating the fund universe based on variables that proxy for the extent to which a fund faces an investment clientele less subjected to marketing or more sensitive to return performance. Prior studies find such variables are associated with alpha (Del Guercio and Reuter, 2014), but recall that our model is robust to cross-sectional differences in equilibrium alphas. We therefore expect to find similar results in both subsamples. In all cases, we split the universe at essentially a variable's median value. First, we separate funds based on the fraction of AUM generated through broker marketing,

as Del Guercio and Reuter (2014) find that investors in broker-sold funds are less responsive to returns, and Christoffersen, Evans, and Musto (2013) find that payments to brokers predict fund flows. For both the high- and low-broker-sold subsamples, we obtain estimates very similar to those in Table 1. Second, we separate funds based on the 12b-1 fee, another marketing proxy (e.g., Barber, Odean, and Zheng (2005)). Both the high- and low-fee subsamples produce estimates very similar to those in Table 1. Third, we separate funds into those with and without an institutional share class, as Evans and Fahlenbrach (2012) find that flows from institutional investors are more sensitive to a fund's performance and fees. Both subsamples produce results very similar to those in Table 1.

We also form subsamples by separating funds according to the number of stocks a fund holds, again splitting the fund universe at the median. We use this variable to proxy crudely for the extent to which an active fund employs quantitative strategies. Abis (2018), for example, finds that quantitative funds hold more stocks. Again, both subsamples exhibit similar results.

We also investigate the robustness of our results to including other fund characteristics that may impact a fund's liquidity. In our empirical tests, we compute L as the liquidity of the fund's equity portfolio, omitting cash holdings. The fraction of a fund's total AUM held in cash may impact the choice of the equity portfolio's liquidity. A cash buffer may allow a fund to trade in and out of equity positions more patiently or to deal more efficiently with investor flows to and from the fund. Those abilities may allow the fund to have a less liquid equity portfolio. Such an effect can enter our model through the value of θ that multiplies the cost function in equation (8). A higher cash weight can imply a lower θ, i.e., lower trading costs for the fund's equity portfolio, given that portfolio's characteristics. A lower value of θ implies a lower value of b_0 in equation (17), but the implied values of b_1, b_2, and b_3 are unchanged. Consistent with these predictions, when we add the fund's cash weight as an independent variable to the regression in (18), that variable enters with a significantly negative coefficient, but the coefficients on A, T, and f are virtually unchanged. Motivated by the above fund-flow consideration, we also include the fund's flow volatility as a control in equation (18), measured as the standard deviation of the fund's 12 monthly flows during the previous calendar year, winsorized at the 1st and 99th percentiles. Again the coefficients on A, T, and f are very similar to those in Table 1. Results for these regressions are reported in the Internet Appendix.

We consider robustness to methodology as well. We re-estimate our regressions at the annual rather than quarterly frequency. We compute standard errors using two-way clustering by both fund and quarter, versus our main specification in which we cluster by fund to

account for the persistence of a fund's characteristics over time. These alternatives produce results very similar to those in Table 1. See the Internet Appendix for details.

3.4 Portfolio Liquidity Under Different Trading Costs

We also consider alternative measures of portfolio liquidity that are derived as in Section 2 but with a stock's trading cost specified differently from equation (6). One modification of equation (6) replaces the stock's market capitalization by its trading volume:

$$C_i = c\,\frac{D_i}{V_i}\ , \tag{19}$$

where V_i is dollar trading volume for stock i. Under this specification, what matters for the proportional cost of a trade is the size of the trade relative to the stock's total trading. Denote $v_i = V_i/V$, where V is total dollar trading volume summed across all stocks. The total trading cost in equation (7) then becomes

$$C = (c/V)\ D^2\ \underbrace{\left(\sum_{i=1}^{N}\frac{w_i^2}{v_i}\right)}_{L_V^{-1}}\ . \tag{20}$$

In this case, our portfolio liquidity measure from equation (3) is modified as indicated in equation (20). To compute this alternative version of L, which we denote by L_V, we use each stock's trading volume, V_i, over the previous three months. Table 2 reports results when Table 1 is recomputed by using L_V. All three fund characteristics (size, expense ratio, turnover) again enter significantly in the predicted directions, with coefficient estimates that are quite similar to the original values.

In equation (6), the cost per dollar traded of a stock increases linearly with the ratio of the dollar amount traded to the stock's market capitalization. Trading, say, 1% of a stock's market capitalization then costs twice as much per dollar traded compared to trading 0.5% of the stock's capitalization. We can allow a nonlinear relation by generalizing equation (6) as

$$C_i = c\,\left(\frac{D_i}{M_i}\right)^{\eta}\ , \tag{21}$$

where $\eta > 0$. The trading cost in equation (7) then becomes

$$C = \left(\frac{c}{M^{\eta}}\right) D^{1+\eta}\ \underbrace{\left(\sum_{i=1}^{N}\frac{w_i^{1+\eta}}{m_i^{\eta}}\right)}_{L_{\eta}^{-1}}\ , \tag{22}$$

so that our portfolio liquidity measure is modified as labeled by L_η in equation (22). The linear specification in equation (6) sets $\eta = 1$. Kyle and Obizhaeva (2016) examine costs of portfolio transition trades and conclude that a linear function fits the data slightly less well than a square-root relation, i.e., $\eta = 0.5$. For values of η ranging from 0.1 to 0.9, Table 3 reports results of re-estimating the original multiple regression reported in the last column of Table 1. Size, expense ratio, and turnover consistently enter significantly in the predicted directions, with t-statistics quite similar to those in Table 1. While the t-statistics are little changed, the coefficients on all three variables become smaller in absolute magnitude as η decreases. The latter effect is consistent with ϕ becoming larger in magnitude as L becomes less sensitive to trading larger fractions of a stock's total shares. Recall that ϕ is minus the exponent on L in the fund's cost function given by equation (8), and note from equations (16) and (17) that ϕ is also the denominator of the coefficients on size, expense ratio, and turnover. If the fraction of a stock's market cap that a fund trades is assigned a weaker role (lower η) in trading costs, then it makes sense that the resulting measure of portfolio liquidity must exert a stronger role (higher ϕ) in order to capture liquidity's effects.

As explained at the outset, a portfolio's liquidity depends not only on the liquidity of its securities but also on the portfolio's diversification. The above alternative versions of L still incorporate diversification, because their underlying trading-cost derivations still treat the portfolio as an asset. Our main measure of L in equation (3) has a correlation (in logs) of 0.96 with the L in equation (20) and at least 0.98 with the L in equation (22) for $\eta \geq 0.1$. In contrast, if we re-compute our measure of portfolio liquidity with the diversification component excluded, using the decomposition in the next section, the correlation with our original measure is just 0.71.

4. Liquidity Tradeoffs

In this section, we examine fund tradeoffs that involve components of portfolio liquidity. After identifying these components in Section 4.1, we present empirical evidence of their tradeoffs in Section 4.2.

4.1. Components of Portfolio Liquidity

We show in Appendix A that portfolio liquidity from equation (3) can be decomposed as

$$L = \underbrace{\frac{1}{N}\sum_{i=1}^{N}\mathcal{L}_i}_{\text{Stock Liquidity}} \times \underbrace{\left(\frac{N}{N_M}\right)\left[1+\text{Var}^*\left(\frac{w_i}{m_i^*}\right)\right]^{-1}}_{\text{Diversification}}. \tag{23}$$

The first component of L is the equal-weighted average of $\mathcal{L}_i = M_i/\overline{M}$, with $\overline{M}$ denoting the average market capitalization of stocks in the benchmark. That is, $\overline{M} = \frac{1}{N_M}\sum_{j=1}^{N_M} M_j$, where N_M is the number of stocks in the benchmark. We label this component "stock liquidity" because $\mathcal{L}_i$ captures the liquidity of stock i relative to all stocks in the benchmark. Stock liquidity is larger (smaller) than 1 if the portfolio's holdings have a larger (smaller) average market capitalization than the average stock in the benchmark.

Using a stock's market capitalization to measure its liquidity follows from our assumption (6), which implies that trading \$1 of stock i incurs a cost proportional to $1/m_i$. This implication is intuitive—trading a given dollar amount of a small-cap stock (whose m_i is small) incurs a larger price impact than trading the same dollar amount of a large-cap stock (whose m_i is large). Moreover, market capitalization is closely related to other measures of stock liquidity in the data. For example, we calculate the correlations between the log of market capitalization and the logs of two popular measures, the Amihud (2002) measure of illiquidity and dollar volume, across all common stocks. The two correlations average -0.91 and 0.90, respectively, across all years in our sample period. In a robustness analysis, we show that alternative measures of stock liquidity, namely the Amihud measure, dollar volume, and the bid-ask spread, produce similar tradeoff results, especially with quarter fixed effects (see the Internet Appendix). Dollar volume is the appropriate measure of a stock's liquidity in the alternative version of L presented in equation (20). Finally, it makes little difference whether market capitalization is float-adjusted or not: the correlation between the logs of float-adjusted and unadjusted market capitalization is 0.98.[11] We use unadjusted market capitalization in our empirical analysis to maximize data coverage.

The second component of L is "diversification." We choose this label for the second factor in equation (23) because that factor includes several elements that are commonly used to judge the extent to which a portfolio is diversified, as explained below.

Broadly speaking, diversification refers to spreading one's wealth across many assets in a balanced fashion. The implications of diversification for portfolio risk are well understood.

[11] We compute this correlation using data on the Russell 3000 stocks from 2011 to 2014. Data on stocks' shares outstanding are from CRSP. Data on float-adjusted shares outstanding are from Russell.

We show that diversification also has implications for transaction costs: better-diversified portfolios are cheaper to trade. Such portfolios are more liquid because they incur lower trading costs than more concentrated portfolios with the same size and turnover.

Diversification is a foundational concept in finance, yet there is no accepted standard for measuring it. In an important early contribution, Blume and Friend (1975) use two measures. The first one is the number of stocks in the portfolio. This measure is also used by Goetzmann and Kumar (2008), Ivković, Sialm, and Weisbenner (2008), Pollet and Wilson (2008), and others. The idea is that portfolios holding more stocks are better diversified. While this idea is sound, the measure is far from perfect. Consider two portfolios holding the same set of 500 stocks. The first portfolio weights the stocks in proportion to their market capitalization. The second portfolio is 99.9% invested in a single stock while the remaining 0.1% is spread across the remaining 499 stocks. Even though both portfolios hold the same number of stocks, the first portfolio is clearly better diversified.

The second measure of diversification used by Blume and Friend is the sum of squared deviations of portfolio weights from market weights, essentially a market-adjusted Herfindahl index. The Herfindahl index measures portfolio concentration, the inverse of diversification. Studies that have designed related portfolio measures include Kacperczyk, Sialm, and Zheng (2005), Goetzmann and Kumar (2008), and Cremers and Petajisto (2009), among others.

Our measure of portfolio diversification blends the ideas from both of the above measures. As one can see from equation (23), our measure can be further decomposed as

$$\text{Diversification} = \underbrace{\left(\frac{N}{N_M}\right)}_{\text{Coverage}} \times \underbrace{\left[1 + \text{Var}^*\left(\frac{w_i}{m_i^*}\right)\right]^{-1}}_{\text{Balance}}. \tag{24}$$

The first component, "coverage," is the number of stocks in the portfolio (N) divided by the number of stocks in the benchmark (N_M). Dividing by N_M makes sense. If all firms in the benchmark were to merge into one conglomerate, a portfolio holding only the conglomerate's stock would be perfectly diversified despite holding only a single stock. Given N_M, portfolios holding more stocks have larger coverage. Coverage is always between 0 and 1, with the maximum value reached if the portfolio holds every stock in the benchmark.

The second component, "balance," reflects the extent to which the portfolio's weights resemble market-cap weights, regardless of the number of stocks in the portfolio. The term $\text{Var}^*\left(w_i/m_i^*\right)$ is the variance of w_i/m_i^* with respect to the probability measure defined by $m_i^* = m_i/\sum_{i=1}^{N} m_i$, so that $\sum_{i=1}^{N} m_i^* = 1$.[12] If portfolio weights equal market-cap weights,

[12]Note that $\sum_{i=1}^{N_M} m_i = 1$, but $\sum_{i=1}^{N} m_i \leq 1$, because $N \leq N_M$. $\text{Var}^*(.)$ can be easily computed using the

so that $w_i/m_i^* = 1$, then $\text{Var}^*(w_i/m_i^*) = 0$ and balance equals 1. Like coverage, balance is always between 0 and 1.

Equation (24) shows that a portfolio is well diversified if it holds a large fraction of the benchmark's stocks and if its weights are close to market-cap weights. Given the ranges of coverage and balance, diversification is always between 0 and 1. The benchmark portfolio has coverage and balance both equal to 1.[13]

Figure 1 provides some history of L and its components for active mutual funds. Panel A plots the time series of the cross-sectional means of L across all funds, relative to the market benchmark. Average L doubled between 1980 and 2000, indicating that fund portfolios became substantially more liquid relative to the market benchmark. To understand this pattern, we plot in Panel B the time series of the two components of L: stock liquidity and diversification. Stock liquidity rose sharply in the late 1990s, explaining the contemporaneous increase in L observed in Panel A, but it declined steadily in the 21st century.[14] Judging by this large decline, one might expect fund portfolios to have become less liquid in the 21st century, but that is not the case, as shown in Panel A. The reason is that fund portfolios have become much more diversified, with diversification almost tripling between 2000 and 2014. The two opposing effects, the decrease in stock liquidity and the increase in diversification, roughly cancel out, resulting in a flat pattern in L since 2000.

The sharp increase in diversification after 2000 is striking. To shed more light on it, we plot in Panel C the components of diversification: balance and coverage. Both components rise steadily, especially after 2000.[15] The increase in coverage, equal to N/N_M, is dissected in Panel D. The average N rises essentially linearly from 54 in 1980 to 126 in 2014, indicating that funds hold an increasingly large number of stocks. In contrast, the number of stocks in the market plummets from about 8,600 in the late 1990s to fewer than 5,000 in 2014. The observed increase in coverage is thus driven by a combination of a rising N and falling N_M. Together, Panels C and D show that active mutual funds have become more index-like.

The time-series patterns in Figure 1 could be driven by fund entry or exit. Instead,

expression $\text{Var}^*\left(w_i/m_i^*\right) = \sum_{i=1}^{N} w_i^2/m_i^* - 1$. Details are in Appendix A.

[13]Our measure of diversification is easy to calculate from equation (24). Those wishing to circumvent the calculation of variance with respect to the m^* probability measure can follow a simple two-step approach: first compute L from equation (3) and then divide it by stock liquidity, following equation (23).

[14]This decline indicates that the average stock held by mutual funds became smaller relative to the average stock in the benchmark. Either funds tilted their portfolios toward smaller stocks or the average benchmark stock increased in size. Evidence of the former effect is provided by Blume and Keim (2017), who show that institutional investors increased their holdings of smaller stocks in recent decades.

[15]The upward trends in both components of diversification, as well as the resulting upward trend in portfolio liquidity, are statistically significant, as we show in the Internet Appendix.

we find these patterns are driven largely by within-fund variation, though the increase in diversification is also significantly affected by fund entry and exit. In particular, the increase in N over the past decade is explained mainly by entry and exit. One possible channel is the entry of quantitative funds, which tend to hold many stocks and whose numbers swelled over the past decade (see Abis, 2018). We show these results in the Internet Appendix. That Appendix also provides cross-sectional descriptive evidence on L and its components.

4.2. Empirical Evidence of Tradeoffs

In addition to the main fund tradeoffs implied by equation (17), our model also implies tradeoffs that involve the components of portfolio liquidity. Equation (23) implies that

$$\ln(L) = \ln(\text{Stock Liquidity}) + \ln(\text{Diversification}) \ . \tag{25}$$

Combined with equation (17), this equation implies

$$\ln(\text{Diversification}) = b_0 + b_1 \ln A - b_2 \ln f + b_3 \ln T - \ln(\text{Stock Liquidity}) \ , \tag{26}$$

where b_0, b_1, b_2, and b_3 are the same constants as before. Equation (26) makes strong predictions about the determinants of portfolio diversification. In equilibrium, funds with more-diversified portfolios should be larger and cheaper, they should trade more, and their stock holdings should be less liquid, on average.

Column 1 of Table 4 provides strong support for these predictions. Fund size, expense ratio, and turnover help explain diversification with the predicted signs, and the slopes have magnitudes similar to those in column 4 of Table 1. The new regressor, stock liquidity, also enters with the right sign and is highly significant, both statistically ($t = -21.61$) and economically. A one-standard-deviation increase in ln(Stock Liquidity) is associated with a 0.95 decrease in ln(Diversification), for example, a decrease in diversification from 0.26 to 0.10. Stock liquidity and diversification are thus substitutes: funds tend to make up for the low liquidity of their holdings by diversifying more. This evidence fits our model.

The estimated slope on stock liquidity, -0.621, is significantly different from its model-predicted value of -1. Our model is clearly not perfect, but no model is, and even an imperfect model can be useful. The usefulness of our model in this context lies in its novel prediction that funds face a tradeoff between their diversification and the average liquidity of their holdings. This prediction is strongly supported in the data as the aforementioned slope, -0.621, is significantly negative. Our model may be too simple to nail down the slope's magnitude, but it makes a new and correct prediction about its sign.

The tradeoffs involving diversification are very robust. They obtain not only for our theoretically motivated measure of diversification from equation (24) but also for three ad-hoc measures: the Herfindahl index of portfolio weights, the number of stocks in the portfolio, and the R-squared from the regression of fund returns on benchmark returns. Moreover, the tradeoffs obtain not only with sector-quarter fixed effects, as in Table 4, but also with quarter fixed effects. Finally, the tradeoffs also emerge from simple correlations. For example, the within-sector cross-sectional correlation between diversification and stock liquidity is -41%. See the Internet Appendix for details.

Next, we drill deeper by decomposing diversification following equation (2):

$$\ln(\text{Diversification}) = \ln(\text{Coverage}) + \ln(\text{Balance}) \ . \tag{27}$$

Combined with equation (26), this equation implies

$$\ln(\text{Coverage}) = b_0 + b_1 \ln A - b_2 \ln f + b_3 \ln T - \ln(\text{Stock Liquidity}) - \ln(\text{Balance}) \tag{28}$$

and

$$\ln(\text{Balance}) = b_0 + b_1 \ln A - b_2 \ln f + b_3 \ln T - \ln(\text{Stock Liquidity}) - \ln(\text{Coverage}) \ . \tag{29}$$

These equations make predictions about the determinants of portfolio coverage and balance.

Columns 2 and 3 of Table 4 support those predictions. In both regressions, all the variables enter with the predicted signs. Most of the variables are highly significant; only turnover in column 3 is marginally significant. The slopes on balance in column 2 and coverage in column 3 are both negative, indicating that coverage and balance are substitutes. Funds that are less diversified in terms of coverage tend to be more diversified in terms of balance, and vice versa.

Finally, column 4 of Table 4 tests the prediction analogous to that in equation (26), except that diversification and stock liquidity switch sides: the former appears on the right-hand side and the latter on the left-hand side of the regression. The evidence again supports the model, though a bit less strongly. Three of the four slopes have the right sign and are significant. The slope on turnover is insignificant. The slope on diversification has the predicted sign, but it differs from its model-predicted value of -1. A similar discrepancy appears in columns 2 and 3. As noted earlier, our model is too simple to capture the magnitudes of all slope coefficients, but at least it makes correct predictions about their signs.

The results in Table 4 survive a variety of robustness checks, similar to those discussed earlier in the context of Table 1. For example, we conduct subsample analysis, change data frequency, and double-cluster standard errors. For details, see the Internet Appendix.

5. Fund Activeness

Funds actively apply their skill in an effort to reap profits. Recall that the function $g(T, L)$ in equation (9) captures how actively skill is applied. We refer to this function as "activeness." We hypothesize that $g(T, L)$ is increasing in turnover, T, and decreasing in portfolio liquidity, L. That is, a fund is more active if it trades more and if it holds a less-liquid portfolio.

We hypothesize that $g(T, L)$ is increasing in T because we find it reasonable for the expected benchmark-adjusted gross return in equation (9) to be increasing in T. An active fund cannot expect to generate profits without trading. A positive role for T emerges also from the theory and evidence of Pástor, Stambaugh, and Taylor (2017), who establish a positive link between a fund's turnover and its performance.[16] Intuitively, higher turnover means the fund is more frequently applying its skill in identifying profit opportunities.

We hypothesize that $g(T, L)$ is decreasing in L because we expect both components of L, stock liquidity and diversification, to be negatively related to activeness. The role of stock liquidity in activeness reflects evidence that mispricing is greater among less-liquid and smaller stocks (e.g., Sadka and Scherbina, 2007, and Stambaugh, Yu, and Yuan, 2015), consistent with arguments that arbitrage is deterred by higher trading costs and greater volatility (e.g., Shleifer and Vishny, 1997, Pontiff, 2006). A fund tilting toward such stocks is more actively pursuing mispricing where it is most prevalent.

Both components of diversification—coverage and balance—explain diversification's role in activeness. By holding fewer stocks (i.e., lower coverage), a fund can focus on its best trading ideas, leading to higher expected gross profits. By deviating more from market-cap weights (i.e., lower balance), a fund can place larger bets on its better ideas, again boosting performance. Theoretical settings in which portfolio concentration (i.e., lower diversification) arises optimally include Merton (1987), van Nieuwerburgh and Veldkamp (2010), and Kacperczyk, van Nieuwerburgh, and Veldkamp (2016). Empirical evidence linking portfolio concentration to performance includes results in Kacperczyk, Sialm, and Zheng (2005), Ivković, Sialm, and Weisbenner (2008), and Choi et al. (2017).

[16] Additional evidence on the turnover-performance relation in mutual funds is presented by Wermers (2000), Chen, Jagadeesh and Wermers (2001), Kacperczyk, Sialm, and Zheng (2005), and others. Turnover is related to the fund's investment horizon, whose relation to fund performance is analyzed by Yan and Zhang (2009), Cremers and Pareek (2016), and Lan, Moneta, and Wermers (2018). The latter study also reports that large-cap funds tend to have longer investment horizons than small-cap funds, which is consistent with our model to the extent that a longer investment horizon indicates low turnover.

5.1. Choosing Activeness: Implications

We begin by assuming activeness takes the simple form $g(T, L) = T^{\psi} L^{\beta}$. Also, to simplify the analysis in this subsection, we assume $\alpha = 0$ in equilibrium (Berk and Green, 2004). If there exists a maximum for fee revenue with respect to T, L, and f, then $\psi = \lambda/\gamma$ and $\beta = -\phi/\gamma$ (as shown in the Appendix). Therefore, we specify activeness as

$$g(T, L) = T^{\lambda/\gamma} L^{-\phi/\gamma} \ . \tag{30}$$

This specification provides an interesting set of tradeoffs for a fund. With activeness given by equation (30), a fund does not have unique values of T, L, and f that maximize its fee revenue, F. The fund instead trades off different values of those quantities that produce the same maximized F, given the fund's skill, μ. Specifically, the fund can choose any values of T, L, and f that satisfy equation (30) and the relation,

$$\mu g(T, L) = \frac{\gamma}{\gamma - 1} f, \tag{31}$$

which rearranges equation (12), setting $\alpha = 0$, replacing a by the right-hand side of equation (9), and replacing $q(A)$ by its implied value from equation (15). For example, the fund can arbitrarily choose its fee rate, f, which then dictates its activeness, $g(T, L)$, via equation (31). Given that level of activeness, the fund can arbitrarily choose its portfolio liquidity, L, which then dictates its turnover, T, via equation (30).

A higher fee rate dictates greater activeness but does not produce greater fee revenue. It simply dictates a smaller fund. To see this, first recall from Section 3.1 that $A = q^{-1}(h)$, where $q(\ )$ is the proportional cost in equation (11), and $h = \mu g(T, L) - \alpha - f$. It follows from equation (8) that $q^{-1}(h) = \left[(1/\theta) T^{-\lambda} L^{\phi} h\right]^{1/(\gamma-1)}$. Multiplying by f and substituting $\mu g(T, L) - f$ for h (setting $\alpha = 0$) gives fee revenue, $F = fA$, as

$$F = f \left(\frac{1}{\theta} T^{-\lambda} L^{\phi} \left[\mu g(T, L) - f \right] \right)^{\frac{1}{\gamma-1}} \ . \tag{32}$$

We then substitute from equations (30) and (31) into equation (32) to obtain

$$F = \mu^{\frac{\gamma}{\gamma-1}} (\gamma - 1) \left(\frac{1}{\theta \gamma^{\gamma}} \right)^{\frac{1}{\gamma-1}} \ . \tag{33}$$

A fund's equilibrium fee revenue is pinned down by the fund's skill μ, holding the cost parameters θ and γ constant. It makes sense for more skilled funds to earn higher fee revenue, and this prediction is not unique to our model.[17] The fee revenue, $F = fA$, does

[17] Skill also determines fee revenue in the model of Berk and Green (2004). See Section 8 for further discussion. Berk and van Binsbergen (2015) also discuss the relation between skill and fee revenue.

not depend on f because when f changes, A adjusts in the opposite direction to keep F constant. Irrelevance of f for F also occurs, for example, in the equilibrium models of Berk and Green (2004), Hugonnier and Kaniel (2010), and Stambaugh (2014).

The data confirm that when $g(T, L)$ is computed as in equation (30), it depends significantly on both T and L in the correct directions. Pairing the regression estimates of b_1, b_2, and b_3 in equation (17) with their corresponding functions of γ, λ, and ϕ in equation (16) delivers implied estimates of 0.138 and -1.367 for the exponents of T and L, respectively, in equation (30). Dividing those values by their standard errors, computed via the delta method, gives t-statistics of 4.58 and -13.98. Both T and L thus enter $g(T, L)$ significantly.

We obtain two implications for activeness: (i) larger funds choose to be less active, controlling for f, and (ii) higher-fee funds choose to be more active, controlling for A. To see these, we divide the right-hand side of equation (32), which equals fA, by f to obtain

$$A = \left(\frac{1}{\theta} T^{-\lambda} L^{\phi} \left[\mu g(T, L) - f\right]\right)^{\frac{1}{\gamma - 1}} \ . \tag{34}$$

We then substitute from equations (30) and (31) to obtain

$$A \ = \ g^{-\frac{\gamma}{\gamma - 1}} \left[\frac{f}{\theta(\gamma - 1)}\right]^{\frac{1}{\gamma - 1}} \ . \tag{35}$$

Taking logs and rearranging gives

$$\ln(g) = d_0 + d_1 \ln(f) - d_2 \ln(A) \ , \tag{36}$$

where d_1 and d_2 are positive and $d_0 = -(1/\gamma) \ln[\theta(\gamma - 1)]$. Thus, g is increasing in f and decreasing in A. Intuitively, larger funds, facing diseconomies of scale, optimally reduce their trading costs by reducing their activeness. Lower-fee funds also choose to be less active because they are less skilled, holding size constant.

Suppose we were to compute observations of g from equation (30), with the exponents on T and L implied by the coefficients from the regression corresponding to equation (17), as discussed above. Then the regression corresponding to equation (36) would deliver estimates of d_1 and d_2 that are simple transformations of our previously reported estimates of the coefficients in equation (17). In other words, the data would supply no additional information about the implied fund tradeoffs involving activeness. Therefore, we instead look for empirical confirmation of these tradeoffs when g is computed in a simpler way.

5.2. Computing Activeness: Evidence

Our simplified calculation of activeness is motivated by the same setting in which we derive our portfolio liquidity measure. In that setting, a portfolio is viewed as just another asset, effectively traded as such, with each stock's traded amount being proportional to its portfolio weight. Recall that if a fund trades its portfolio that way, its cost function is given by equation (8) with $\gamma = 2$, $\lambda = 2$, and $\phi = 1$. Applying those parameter values to equation (30) gives

$$g(T, L) = TL^{-1/2}, \tag{37}$$

our empirical measure of activeness.

With g computed as in equation (37), we estimate the regression corresponding to equation (36). The results are reported in Table 5. As our model predicts, activeness is related negatively to fund size and positively to expense ratio. Both relations are very strong, with t-statistics of about 10 in magnitude. These relations obtain not only in the multiple regression but also in simple regressions, with t-statistics exceeding 13 in magnitude. Like Tables 1 through 4, Table 5 reports results with sector-quarter fixed effects, but including just quarter fixed effects produces very similar results. We also find very similar results in two subsamples, 1979 through 2004 and 2005 through 2014. See the Internet Appendix.

Our activeness measure from equation (37) has a correlation of 55%, in logs, with the popular active share measure of Cremers and Petajisto (2009). Active share is computed by using only portfolio weights of the fund and the benchmark, as is our portfolio liquidity measure, L. Both active share and L capture deviations of portfolio weights from benchmark weights, so it is not surprising that the correlation between active share and L is high, -79%, in logs. But our measure of activeness incorporates not only L but also T. This inclusion of turnover captures the intuitive notion that a fund is more active if it trades more. The presence of turnover in activeness, and its absence from active share, is the main difference between the two measures. Yet when we replace activeness by active share in Table 5, we obtain the same conclusions: smaller funds and higher-fee funds tend to be more active. Cremers and Petajisto report similar results. We also obtain the same conclusions when replacing activeness by another proxy, the inverse of the R-squared from the regression of fund returns on benchmark returns. See the Internet Appendix.

6. Tradeoffs: Simple Correlations

The tradeoffs implied by our model take the form of multiple regressions (e.g., equation (17)), but they emerge also from simple correlations. In this section, we analyze the correlations among the four main fund characteristics. To understand these correlations in the context of our model, we need additional assumptions.

6.1. Larger Funds Are Cheaper

Our model predicts a negative correlation between fund size, A, and expense ratio, f. We derive this prediction from equation (33), in which a fund's equilibrium fee revenue, $F = Af$, is determined by the fund's skill, μ. Holding μ constant, A and f are perfectly negatively correlated across funds. If μ varies across funds, the correlation between A and f is no longer perfect, but it remains negative as long as μ is not highly correlated with f across funds. Specifically, let $\beta_{\mu,f}$ denote the slope from the cross-sectional regression of $\ln(\mu)$ on $\ln(f)$. Our model implies a negative cross-sectional correlation between A and f as long as $\beta_{\mu,f} < (\gamma - 1)/\gamma$ (see Appendix A for the proof). It makes sense for $\beta_{\mu,f}$ to be positive, in that more skilled funds should be able to charge higher fee rates. Nonetheless, it seems plausible for $\beta_{\mu,f}$ to be small enough to satisfy the assumption because in practice, expense ratios have a variety of determinants beyond skill (marketing, distribution, etc.).

Empirical evidence strongly supports the prediction that larger funds are cheaper. Table 6 reports correlations between fund characteristics, again measured in logs. In our mutual fund dataset, the cross-sectional within-sector correlation between fund size and expense ratio is -31.5% ($t = -15.27$). Larger funds clearly charge lower expense ratios. This evidence is consistent with our model. Others have also reported a negative correlation between size and expense ratio for active mutual funds (e.g., Warner and Wu, 2011). One potential explanation is that funds' administrative costs tend to decline with AUM, but funds need not pass those savings along to investors. The negative correlation occurs in part because some fund advisory contracts set the management fee to decline in AUM (e.g., Deli, 2002). That practice is consistent with our model, which appears to be the first formal theory supporting a negative relation between size and expense ratio.

The correlation between fund size and expense ratio is also strongly negative in the time series for the typical fund, -25.1% ($t = -17.54$). In computing the time-series correlations in Panel B of Table 6, we need to account for the substantial growth in the dollar values of stocks that renders dollar AUM unappealing as a time-series measure of fund size: AUM

values in the 1980s are not comparable to those today. To address this fact, we divide each fund's AUM by the contemporaneous total stock market capitalization. Pástor, Stambaugh, and Taylor (2015) also deflate fund size by stock-market value when analyzing a time series of fund size.

6.2. Larger and Cheaper Funds Are Less Active

Recall from equation (36) that the fund's activeness, $g(T, L)$, is positively correlated with f, controlling for A, and negatively correlated with A, controlling for f. These correlations obtain also without controls, under additional assumptions. If skill (μ) is constant across funds, both simple correlations are perfect. The positive correlation between g and f follows directly from equation (31). The negative correlation between g and A obtains when we substitute for f from equation (31) into equation (35), yielding

$$A = \frac{1}{g}\,\mu^{\frac{1}{\gamma-1}}\,(\gamma\theta)^{-\frac{1}{\gamma-1}} \quad . \tag{38}$$

The product of fund size and activeness, Ag, is determined by μ. Holding μ constant, A is perfectly negatively correlated with g. If μ varies across funds, both correlations retain their signs as long as μ is not too highly correlated with f or A. Specifically, let $\beta_{\mu,A}$ denote the slope from the regression of $\ln(\mu)$ on $\ln(A)$. The model implies a negative correlation between g and A as long as $\beta_{\mu,A} < \gamma - 1$ and a positive correlation between g and f as long as $\beta_{\mu,f} < 1$ (see Appendix A for the proof). Empirical evidence strongly supports both of these predictions, as shown in columns 1 and 2 of Table 5. Funds that are larger and cheaper are less active, as the model predicts.

One way a fund can be less active is to trade less. The above predictions for $g(T, L)$ imply that, controlling for L, T should be negatively related to fund size and positively related to expense ratio. This is indeed true in the data, as we show in the Internet Appendix. Moreover, the relations hold even without controlling for L. In Table 6, T is negatively correlated with fund size, both in the cross section and in the time series: the correlations are -10.5% ($t = -6.00$) and -14.7% ($t = -12.11$), respectively. In addition, T is positively correlated with expense ratio: the correlation is 13.0% ($t = 6.34$) in the cross section and 10.5% ($t = 7.54$) in the time series. Deli (2002) also reports that funds with higher turnover charge more. In short, larger and cheaper funds trade less.

6.3. Funds with More-Liquid Portfolios Are Larger and Cheaper

Table 4 shows that the partial correlation between L and A is positive, and also that the partial correlation between L and f is negative. Both relations hold strongly even in simple correlations, as shown in Table 6. The correlations between L and A are 28.5% ($t = 17.77$) and 30.8% ($t = 18.00$) in the cross section and time series, respectively. The correlations between L and f are -29.1% ($t = -13.29$) and -11.8% ($t = -6.78$). These correlations also emerge from the simple-regression results reported in Table 1. In short, funds with more-liquid portfolios are larger and cheaper, as predicted by our model.

The cross-sectional correlations that involve L are extremely robust. The correlations in Panel A of Table 6 are computed from panel regressions with quarter-sector fixed effects, which isolate cross-sectional correlations within sectors.[18] Those correlations are therefore weighted averages of cross-sectional correlations, where the averaging is across all quarters in our sample. It turns out that the cross-sectional relations involving L hold not only on average, but also in every single quarter in our sample. This stunning fact is plotted in Figure 2. Both correlations involving L retain the same sign in every quarter between 1980 and 2014. In fact, in each quarter, their magnitudes exceed 20% in absolute value.

Two other cross-sectional correlations discussed earlier are similarly strong, which is why we plot their time series in Figure 2. The correlation between fund size and expense ratio, analyzed in Section 6.1, is negative in every single quarter, varying between -0.74 and -0.23 across quarters. The correlation between turnover and expense ratio, analyzed in Section 6.2, is positive in every quarter, varying between 0.10 and 0.36.

While Figure 2 plots cross-sectional correlations, the time-series correlations reported in Table 6 are of similar magnitudes. The time-series correlation between L and fund size, 30.8%, is particularly strong. It shows that when a fund gets larger, its portfolio becomes more liquid. This fact is easily interpreted in the context of our theory. Consider a fund that receives a large inflow. Cognizant of decreasing returns to scale, the fund's manager makes the fund's portfolio more liquid. And vice versa—after a large outflow, a fund can afford to make its portfolio less liquid.

To illustrate these effects, we pick the example of Fidelity Magellan, the largest mutual fund at the turn of the millenium. Figure 3 plots the time series of Magellan's AUM and

[18] We also compute plain cross-sectional correlations (i.e., including quarter fixed effects instead of sector-quarter fixed effects). The results are very similar to those in Panel A of Table 6 so we report them only in the Internet Appendix. In that Appendix, we also show the results from another robustness exercise, in which we recompute Table 6 for two subperiods containing roughly the same number of fund-month observations. The results in both subsamples look very similar to the full-sample ones.

its portfolio liquidity. The comovement between the two series is striking. Between 1980 and 2000, Magellan's assets grew rapidly, in large part due to the fund's stellar performance under Peter Lynch in 1977 through 1990. Over the same period, and especially after 1993, the liquidity of Magellan's portfolio also grew rapidly. From 1993 to 2001, Magellan's L grew from 0.1 to 0.4, a remarkable increase equal to nearly five standard deviations of the sample distribution of L. After 2000, though, Magellan's assets shrank steadily, and by 2014, they were down by almost 90%. Over the same period, Magellan's L was down also, back to about 0.1. A natural interpretation is that Magellan's large size around 2000 forced the fund's managers to increase the liquidity of Magellan's portfolio to shelter the fund from the pernicious effects of decreasing returns to scale.

7. Rethinking Scale

What is a fund's scale? Following Berk and Green (2004), active funds are typically viewed as facing decreasing returns to scale, with scale given by fund size, i.e., AUM. Our framework offers a new perspective on scale. In our setting, funds face decreasing returns to scale, but scale depends not only on size but also on activeness.

Let Π denote the fund's expected dollar profit net of trading costs (but before fees). From equations (8) and (9), this profit is given by

$$\begin{aligned} \Pi &= aA - C(A,T,L) \\ &= \mu g(T,L)A - \theta A^{\gamma} T^{\lambda} L^{-\phi} \ . \end{aligned} \tag{39}$$

When the fund chooses activeness, so that $g(T,L)$ is given by equation (30), then

$$\begin{aligned} \Pi &= \mu T^{\lambda/\gamma} L^{-\phi/\gamma} A - \theta A^{\gamma} T^{\lambda} L^{-\phi} \\ &= \mu S - \theta S^{\gamma} \ , \end{aligned} \tag{40}$$

where

$$\begin{aligned} S &= T^{\lambda/\gamma} L^{-\phi/\gamma} A \\ &= g(T,L)\, A \ . \end{aligned} \tag{41}$$

The net profit function given by equation (40) is concave in S (recall that $\gamma > 1$). That is, the fund faces decreasing returns to scale with respect to S.

The fund's scale, S, is activeness times size, not just size. This concept of fund scale makes intuitive sense. If two funds manage equal amounts of money, but one fund deploys its money more actively, that fund leaves a bigger footprint in the market.

If one considers size alone as a fund's scale, one must condition on a given level of the fund's activeness. The value of activeness, g, affects the degree to which the fund faces decreasing returns with respect to size, A. Decreasing returns are often discussed in the context of the fund's alpha, with α decreasing in A (for a given a fee rate, f). The strength of that negative relation, however, depends on the fund's g. For the latter given by equation (30) and for the cost function in equation (8), equation (12) becomes

$$\alpha = \mu g - \theta g^{\gamma} A^{\gamma-1} - f \ , \tag{42}$$

so

$$\frac{\partial \alpha}{\partial A} = -\theta(\gamma - 1) g^{\gamma} A^{\gamma-2} < 0 \ . \tag{43}$$

An increase in A depresses α more when the fund has a greater g. In other words, both size and activeness matter for the degree to which the fund faces decreasing returns to scale.

8. Relation to Berk and Green (2004)

Besides delivering a different concept of scale, our setting departs from that of Berk and Green (2004), hereafter BG, in other key respects. First, we incorporate both turnover and portfolio liquidity. The fund's choices of those characteristics, absent from BG, enter the fund's trading costs as well as its activeness. Our setting considers four fund characteristics: size, expense ratio, turnover, and portfolio liquidity. Only two of them, size and expense ratio, appear in the BG setting. Our richer setting allows us to obtain new insights into the tradeoffs involving turnover and portfolio liquidity, as well as the tradeoffs involving the components of portfolio liquidity: stock liquidity, diversification, coverage, and balance.

The BG setting can be shown to imply tradeoffs between characteristics for a given fund, but in a more limited way than ours. The most straightforward is the tradeoff between size and expense ratio. In the BG setting, a fund that cuts its fee rate attracts additional capital, which is indexed at low cost. The fund's size is thus inversely related to its expense ratio.

By adding mild assumptions to the BG setting, we can also derive tradeoffs between a fund's size and the two fund characteristics that do not explicitly appear in that setting. Assuming the fund's indexed portion has zero turnover, the BG setting implies a negative relation between a fund's size and its turnover. It also implies a positive relation between size and portfolio liquidity, our newly introduced measure.[19] Thus, although not discussed

[19]This relation follows from the following result, which we derive in Appendix A: If a portfolio with liquidity L is combined with the benchmark (index fund), the liquidity of the resulting combination, $\tilde{L}$, obeys $\tilde{L}^{-1} = 1 + \omega^2(L^{-1} - 1)$, where ω is the non-indexed (active) fraction of the fund.

by BG, it is possible to derive some relations among the four fund characteristics in their setting. Importantly, however, in the BG setting all four characteristics have a single quantity driving those relations—the fraction of the fund that is indexed. So, for example, one cannot consider the implications for size and expense ratio if the fund were to increase its turnover but not change portfolio liquidity. In the BG setting, an increase in turnover would have to reflect a lower fraction indexed, so it would have to be accompanied by a decrease in portfolio liquidity. In our setting, each fund characteristic can potentially trade off against independent variation in the others.

Finally, by providing a more complete specification of trading costs that incorporates turnover and liquidity, we are able to derive equation (17) and apply it cross-sectionally, whereas BG do not make cross-sectional predictions about fund characteristics. Instead, BG focus on the performance-flow relation induced by the updating of investors' beliefs.

9. Liquidity and the Turnover-Performance Relation

As discussed earlier, active funds that trade less-liquid portfolios will likely find greater profit opportunities, before trading costs, but less-liquid portfolios are also costlier to trade. This role of liquidity in the tradeoff between gross profit and trading costs has implications for the relation between a fund's turnover and its subsequent performance.

The basic turnover-performance relation has a simple motivation. A fund is likely to trade more in periods when it identifies more opportunities. In the framework of Pástor, Stambaugh, and Taylor (2017), a fund's trades in the current period establish positions that yield profits in subsequent periods, as prices correct. Time-varying profit opportunities then imply a positive time-series relation between a fund's turnover and its subsequent performance, as that study finds empirically.

Portfolio liquidity's role in the turnover-performance relation arises from its role in the profit-cost tradeoff noted above. If trading costs are high, then so too must be the gross profit opportunities that justify trading. The turnover-performance relation should therefore be stronger for funds whose trades are more costly, as Pástor, Stambaugh, and Taylor (2017) explain. That study confirms this implication empirically, using fund categories as rough proxies to identify funds having higher trading costs, such as funds investing in small-cap stocks. We investigate the same implication, but we instead use a fund characteristic linked directly to trading costs: portfolio liquidity. If a fund exploits profit opportunities by trading a less-liquid portfolio, the gross profit opportunities need to be larger, because a given amount

of trading is costlier when portfolio liquidity is lower. Applying the same reasoning as Pástor, Stambaugh, and Taylor (2017), we expect the turnover-performance relation to be stronger for less liquid portfolios.

We follow that study in regressing each fund's benchmark-adjusted return in month t on the fund's turnover for the most recent fiscal year ending before month t, adding back the expense ratio to the return. Also as in that study, we include fund fixed effects, thereby estimating the time-series relation between a fund's turnover and its subsequent performance. Unlike that study, our regression includes an independent variable that interacts turnover with the illiquidity of the fund's portfolio. This interaction term should enter positively if the turnover-performance relation is indeed stronger for a fund facing higher trading costs, i.e., a fund trading a less-liquid portfolio.

To capture the interaction of turnover with portfolio illiquidity, we use the quantity already defined in equation (37), namely our activeness measure, $TL^{-1/2}$. Because portfolio liquidity, L, enters this quantity inversely, activeness reflects the desired interaction between turnover, T, and portfolio illiquidity.[20] We regress a fund's performance on not only the fund's turnover but also on its activeness, expecting activeness to enter positively.

The results in Table 7 confirm this prediction. The relation between activeness and subsequent performance is positive and strongly significant.[21] For the multiple regression reported in column 3, which includes turnover as well as activeness, the t-statistic on activeness is 7.33. The slope on activeness is positive and highly significant also when we control for portfolio illiquidity, as we show in the Internet Appendix. This result provides further evidence of portfolio liquidity's empirical relevance for trading costs, given the role such costs should play in the turnover-performance relation.

10. Conclusions

We model and document strong tradeoffs among the most salient characteristics of active mutual funds: fund size, expense ratio, turnover, and portfolio liquidity. We find empirically that funds with smaller size, higher expense ratios, and lower turnover tend to hold less-liquid portfolios. They also hold less-diversified portfolios. All of these findings are predicted by our

[20]Since we do not take the logarithm of L in this turnover-performance regression, we winsorize $L^{-1/2}$ at the 1st and 99th percentiles before interacting it with T to compute activeness.

[21]This time-series relation complements prior results documenting positive cross-sectional relations between fund performance and other measures of fund activity, such as active share (Cremers and Petajisto, 2009) and R-squared (Amihud and Goyenko, 2013). Our focus is not on the activeness-performance relation per se but rather on the role of L in the turnover-performance relation found in earlier work.

equilibrium model, in which the key fund characteristics are jointly determined. Additional predictions also hold in the data. For example, larger funds are cheaper, funds that trade less are larger and cheaper, and funds that are less active are larger and cheaper. Table 8 lists the paper's 25 predicted relations between fund characteristics, along with the corresponding t-statistics reported earlier. Virtually all of the predictions receive strong empirical support. Our results provide novel and compelling evidence of decreasing returns to scale in active management. A fund's scale is captured by its activeness times AUM, not just AUM.

Another contribution of our study is to introduce the concept of portfolio liquidity. We show that a portfolio's liquidity depends not only on the liquidity of its holdings but also on its diversification. We derive simple measures of portfolio liquidity and diversification. Based on these measures, we find that active mutual funds' portfolios have become relatively more liquid over time, mostly as a result of becoming more diversified. We also find that the components of portfolio liquidity are substitutes: funds holding less-liquid stocks tend to diversify more, and funds holding fewer stocks choose portfolio weights closer to market-cap weights. Finally, we confirm empirically our prediction that the turnover-performance relation is stronger for funds with less-liquid portfolios.

Our empirical analysis focuses on U.S. equity mutual funds. Future research can apply our concepts and measures to portfolios held by other types of institutions, such as hedge funds, private equity funds, fixed income mutual funds, and pension funds. More research into relations among fund characteristics also seems warranted.

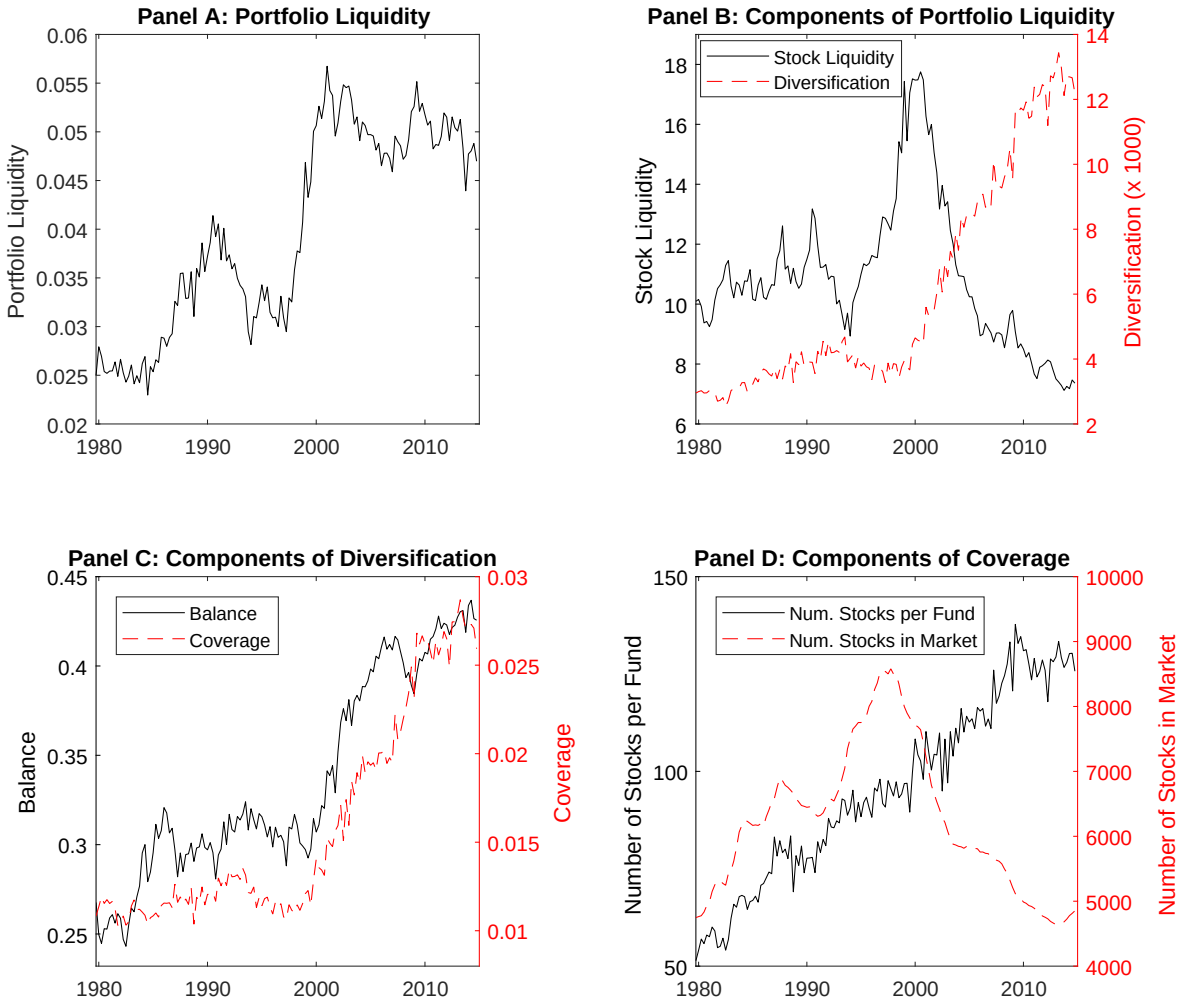

Figure 1. Time Series of Average Portfolio Liquidity and Its Components. This figure plots the quarterly time series of the cross-sectional means of portfolio liquidity, stock liquidity, diversification, coverage, balance, and the number of stocks held by each fund. Liquidity, diversification, and its components are computed with respect to the market benchmark. In Panel D we also plot the number of stocks in the market portfolio.

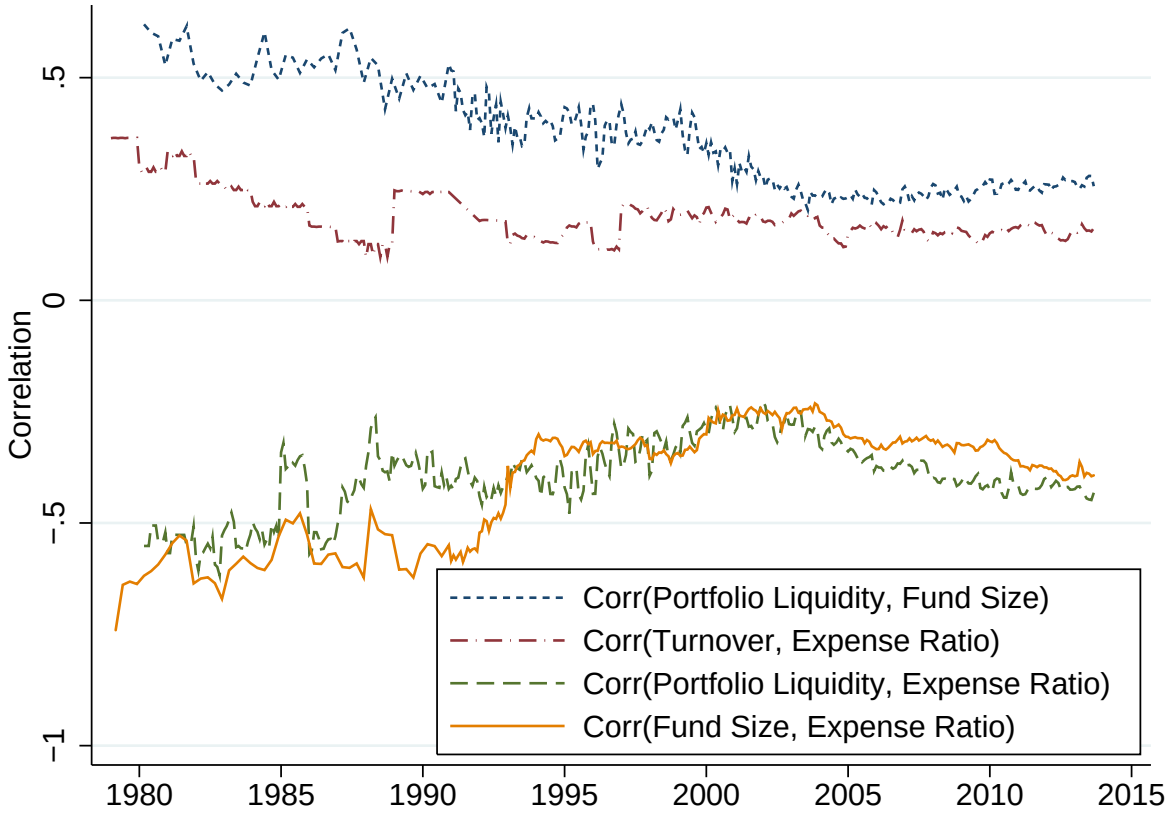

Figure 2. Cross-Sectional Correlations Over Time. This figure plots monthly time series of the cross-sectional correlation between the two variables noted in the legend. All variables are measured in logs. For each correlation, we drop months with fewer than 30 observations. To convert portfolio liquidity from a quarterly to a monthly variable, we take portfolio liquidity from the current month or, if missing, from the previous two months. Portfolio liquidity is computed with respect to the market benchmark.

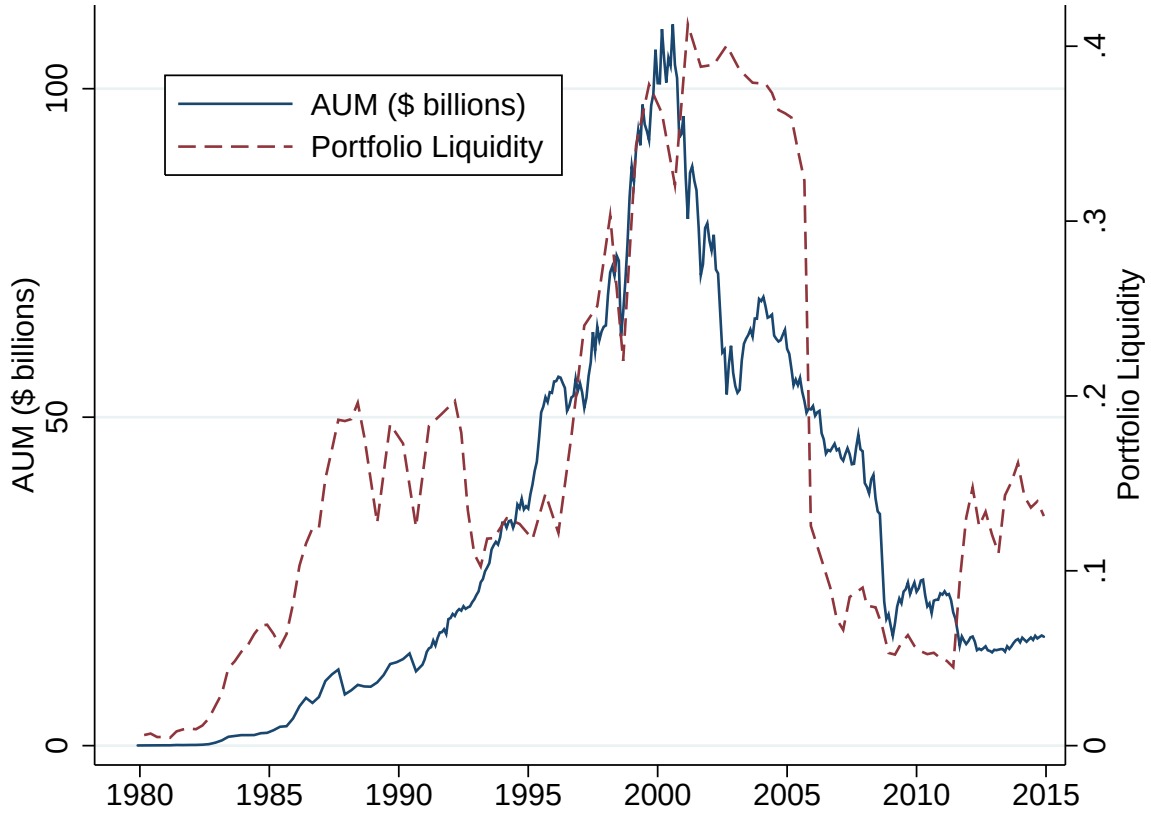

Figure 3. Fidelity Magellan Fund. This figure plots Magellan's assets under management (AUM) and portfolio liquidity, computed with respect to the market benchmark.

Table 1
Explaining Mutual Funds' Portfolio Liquidity

This table presents results from OLS panel regressions in which the dependent variable is a mutual fund's portfolio liquidity, L. The regressors—fund size, A, expense ratio, f, and fund turnover, T—are measured contemporaneously with the dependent variable. All variables are measured in logs. The unit of observation is the fund/quarter. All regressions include sector×quarter fixed effects (FEs) and cluster by fund. The R^2 values in the penultimate row include the FEs' contribution. The last row contains the R^2 values from the regression of the dependent variable on the FEs alone. t-statistics are in parentheses.

	(1)	(2)	(3)	(4)
Fund Size	0.157			0.124
	(17.77)			(13.76)
Expense Ratio		-0.766		-0.608
		(-13.29)		(-11.26)
Turnover			0.0408	0.101
			(1.93)	(4.93)
Observations	88925	89017	81892	76928
R^2	0.627	0.623	0.591	0.652
R^2 (FEs only)	0.594	0.588	0.591	0.598

Table 2
Explaining Volume-Based Portfolio Liquidity

This table presents results for the same regressions as in Table 1 except that portfolio liquidity is computed as L_V in equation (20), using dollar volume.

	(1)	(2)	(3)	(4)
Fund Size	0.159			0.133
	(15.62)			(13.18)
Expense Ratio		-0.743		-0.617
		(-11.51)		(-10.75)
Turnover			0.180	0.242
			(7.67)	(10.48)
Observations	88922	89014	81889	76925
R^2	0.596	0.589	0.574	0.631
R^2 (FEs only)	0.568	0.562	0.564	0.572

Table 3
Explaining Portfolio Liquidity Under Nonlinear Trading Costs

This table presents results corresponding to the last column in Table 1, except that portfolio liquidity is computed as L_η in equation (22), using a nonlinear function for proportional trading costs. The last column in Table 1 corresponds to $\eta = 1$, whereas this table reports results for lower values of η, as identified by the column headings.

	η=0.1	η=0.3	η=0.5	η=0.7	η=0.9
Fund Size	0.00922	0.0298	0.0534	0.0798	0.109
	(11.46)	(12.06)	(12.62)	(13.13)	(13.56)
Expense Ratio	-0.0519	-0.163	-0.283	-0.409	-0.541
	(-11.66)	(-11.70)	(-11.63)	(-11.49)	(-11.34)
Turnover	0.00870	0.0273	0.0473	0.0683	0.0901
	(4.77)	(4.86)	(4.91)	(4.93)	(4.93)
Observations	76928	76928	76928	76928	76928
R^2	0.724	0.710	0.694	0.677	0.660

Table 4
Explaining the Components of Portfolio Liquidity

This table presents results from four OLS panel regressions with dependent variables noted in the column headers. All regressors are measured contemporaneously with the dependent variable. All variables are measured in logs. The unit of observation is the fund/quarter. All regressions include sector×quarter fixed effects (FEs) and cluster by fund. The R^2 values in the penultimate row include the FEs' contribution. The last row contains the R^2 values from the regression of the dependent variable on the FEs alone. t-statistics are in parentheses.

	(1)	(2)	(3)	(4)
	Diversification	Coverage	Balance	Stock Liquidity
Fund Size	0.134	0.0940	0.0452	0.0122
	(15.00)	(12.08)	(7.54)	(2.35)
Expense Ratio	-0.622	-0.408	-0.238	-0.132
	(-11.00)	(-9.33)	(-6.95)	(-5.26)
Turnover	0.122	0.102	0.0247	-0.0146
	(5.96)	(6.37)	(1.92)	(-1.32)
Stock Liquidity	-0.621	-0.337	-0.308	
	(-21.61)	(-14.21)	(-14.90)	
Balance		-0.0447		
		(-2.08)		
Coverage			-0.0343	
			(-2.09)	
Diversification				-0.264
				(-24.49)
Observations	76928	76928	76928	76928
R^2	0.465	0.336	0.286	0.882
R^2 (FEs only)	0.240	0.163	0.172	0.857

Table 5
Explaining Fund Activeness

This table presents results from OLS panel regressions with the dependent variable equal to Activeness. Activeness equals $TL^{-1/2}$, where T is turnover and L is portfolio liquidity. All regressors are measured contemporaneously with the dependent variable. All variables are measured in logs. The unit of observation is the fund/quarter. All regressions include sector×quarter fixed effects (FEs) and cluster by fund. The last row contains the R^2 values from the regression of the dependent variable on the FEs alone. The R^2 values in the penultimate row include the FEs' contribution. t-statistics are in parentheses.

	(1)	(2)	(3)
Fund Size	-0.138		-0.100
	(-13.23)		(-9.53)
Expense Ratio		0.712	0.558
		(13.14)	(10.12)
Observations	76928	76928	76928
R^2	0.392	0.398	0.415
R^2 (FEs only)	0.356	0.356	0.356

Table 6
Correlations Among Fund Characteristics

This table reports correlations among the given fund characteristics, all measured in logs. Panel A reports correlations across funds within sector-quarters. Starting with our full panel dataset, we first de-mean each variable using the mean across all observations in the same sector and quarter, then we compute the full-sample correlation between the two de-meaned variables. Panel B reports time-series correlations within funds, which we compute analogously except that we de-mean each variable using each fund's time-series mean. Fund size is scaled by total stock market capitalization. Portfolio liquidity is defined with respect to the market benchmark. t-statistics are computed clustering by fund and adjusting for de-meaning.

	Fund Size	Expense Ratio	Portfolio Liquidity	Turnover
Panel A: Cross-Sectional Correlations Within Sectors				
Fund Size	1			
Expense Ratio	-0.315	1		
	(-15.27)			
Portfolio Liquidity	0.285	-0.291	1	
	(17.77)	(-13.29)		
Turnover	-0.105	0.130	0.039	1
	(-6.00)	(6.34)	(1.93)	
Panel B: Time-Series Correlations				
Fund Size	1			
Expense Ratio	-0.251	1		
	(-17.54)			
Portfolio Liquidity	0.308	-0.118	1	
	(18.00)	(-6.78)		
Turnover	-0.147	0.105	-0.109	1
	(-12.11)	(7.54)	(-6.76)	

Table 7
Activeness and the Turnover-Performance Relation

The dependent variable is the fund's net return plus expense ratio minus Morningstar's designated benchmark return in month t. Turnover and Activeness are measured for the most recent time period that ends before month t. For this analysis, we use L computed with respect to the market benchmark. Following Pástor, Stambaugh, and Taylor (2017), all regressions include fund fixed effects and cluster by Morningstar sector × month. Heteroskedasticity-robust t-statistics are in parentheses.

	(1)	(2)	(3)
Turnover (T)	0.00123		-0.000625
	(6.53)		(-2.26)
Activeness ($TL^{-1/2}$)		0.000162	0.000190
		(8.67)	(7.33)
Observations	320712	235337	235337
R^2	0.015	0.016	0.016

Table 8
Review of Predicted Relations

This table lists the paper's predicted relations between fund characteristics. The last column contains the corresponding t-statistics, reported in the previous tables. In most cases, the predicted relation between two characteristics controls for others, so the t-statistic is from a multiple regression. All predictions are cross-sectional except the last, which is a within-fund time-series prediction.

Fund Characteristics		Predicted Relation	t-statistic
Portfolio Liquidity	Fund Size	+	13.8
Portfolio Liquidity	Expense Ratio	−	−11.3
Portfolio Liquidity	Turnover	+	4.9
Diversification	Fund Size	+	15.0
Diversification	Expense Ratio	−	−11.0
Diversification	Turnover	+	6.0
Diversification	Stock Liquidity	−	−21.6
Coverage	Fund Size	+	12.1
Coverage	Expense Ratio	−	−9.3
Coverage	Turnover	+	6.4
Coverage	Stock Liquidity	−	−14.2
Coverage	Balance	−	−2.1
Balance	Fund Size	+	7.5
Balance	Expense Ratio	−	−7.0
Balance	Turnover	+	1.9
Balance	Stock Liquidity	−	−14.9
Stock Liquidity	Fund Size	+	2.4
Stock Liquidity	Expense Ratio	−	−5.3
Stock Liquidity	Turnover	+	−1.3
Activeness	Fund Size	−	−9.5
Activeness	Expense Ratio	+	10.1
Fund Size	Expense Ratio	−	−15.3
Fund Size	Turnover	−	−6.0
Expense Ratio	Turnover	+	6.3
Future Return	Activeness	+	7.3

REFERENCES

Abis, Simona, 2018, Man vs. machine: Quantitative and discretionary equity management, Working paper, Columbia University.

Amihud, Yakov, 2002, Illiquidity and stock returns: cross-section and time-series effects, *Journal of Financial Markets* 5, 31–56.

Amihud, Yakov, and Ruslan Goyenko, 2013, Mutual fund's R^2 as predictor of performance, *Review of Financial Studies* 26, 667–694.

Barber, Brad M., Terrance Odean, and Lu Zheng, 2005, Out of sight, out of mind: The effects of expenses on mutual fund flows, *Journal of Business* 78, 2095–2119.

Berk, Jonathan B., and Richard C. Green, 2004, Mutual fund flows and performance in rational markets, *Journal of Political Economy* 112, 1269–1295.

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

Blume, Marshall E., and Irwin Friend, 1975, The asset structure of individual portfolios and some implications for utility functions, *Journal of Finance* 30, 585–603.

Blume, Marshall E., and Donald B. Keim, 2017, The changing nature of institutional stock investing, *Critical Finance Review* 6, 1–41.

Bris, Arturo, Huseyin Gulen, Padma Kadiyala, and P. Raghavendra Rau, 2007, Good stewards, cheap talkers, or family men? The impact of mutual fund closures on fund managers, flows, fees, and performance, *Review of Financial Studies* 20, 953–982.

Busse, Jeffrey A., Tarun Chordia, Lei Jiang, and Yuehua Tang, 2017, Mutual fund trading costs, Working paper, Emory University.

Chen, Hsiu-Lang, Narasimhan Jegadeesh, and Russ Wermers, 2000, The value of active mutual fund management: An examination of the stockholdings and trades of fund managers, *Journal of Financial and Quantitative Analysis* 35, 343–368.

Chen, Joseph, Harrison Hong, Ming Huang, and Jeffrey Kubik, 2004, Does fund size erode mutual fund performance? *American Economic Review*, 94: 1276-1302.

Choi, Nicole, Mark Fedenia, Hilla Skiba, and Tatyana Sokolyk, 2017, Portfolio concentration and performance of institutional investors worldwide, *Journal of Financial Economics* 123, 189–208.

Christoffersen, Susan E.K., Richard Evans, and David K. Musto, 2013, What do consumers' fund flows maximize? Evidence from their brokers' incentives, *Journal of Finance* 68, 201–235.

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

Cremers, Martijn, and Ankur Pareek, 2016, Patient capital outperformance: The investment skill of high active share managers who trade infrequently, *Journal of Financial Economics* 122, 288–306.

Del Guercio, Diane, and Jonathan Reuter, 2014, Mutual fund performance and the incentive to generate alpha, *Journal of Finance* 69, 1673–1704.

Deli, Daniel N., 2002, Mutual fund advisory contracts: An empirical investigation, *Journal of Finance* 57, 109–133.

Edelen, Roger, Richard Evans, and Gregory Kadlec, 2013, Shedding light on "invisible" costs: Trading costs and mutual fund performance, *Financial Analysts Journal* 69, 33–44.

Evans, Richard B., and Rüdiger Fahlenbrach, 2012, Institutional investors and mutual fund governance: Evidence from retail-institutional twin funds, *Review of Financial Studies* 25, 3530–3571.

Gârleanu, Nicolae, and Lasse H. Pedersen, 2018, Efficiently inefficient markets for assets and asset management, *Journal of Finance* 73, 1663–1712.

Goetzmann, William N., and Alok Kumar, 2008, Equity portfolio diversification, *Review of Finance* 12, 433–463.

Harvey, Campbell R., and Yan Liu, 2017, Decreasing returns to scale, fund flows, and performance, Working paper, Duke University.

Hoberg, Gerard, Nitin Kumar, and Nagpurnanand Prabhala, 2018, Mutual fund competition, managerial skill, and alpha persistence, *Review of Financial Studies* 31, 1896–1929.

Hugonnier, Julien, and Ron Kaniel, 2010, Mutual fund portfolio choice in the presence of dynamic flows, *Mathematical Finance* 20, 187–227.

Ivković, Zoran, Clemens Sialm, and Scott Weisbenner, 2008, Portfolio concentration and the performance of individual investors, *Journal of Financial and Quantitative Analysis* 43, 613–655.

Kacperczyk, Marcin, Clemens Sialm, and Lu Zheng, 2005, On the industry concentration of actively managed equity mutual funds, *Journal of Finance* 60, 1983–2011.

Kacperczyk, Marcin, Stijn Van Nieuwerburgh, and Laura Veldkamp, 2016, A rational theory of mutual funds' attention allocation, *Econometrica* 84, 571–626.

Keim, Donald, and Ananth Madhavan, 1997, Transaction costs and investment style: An inter-exchange analysis of institutional equity trades, *Journal of Financial Economics* 46, 265–292.

Kyle, Albert S., and Anna A. Obizhaeva, 2016, Market microstructure invariance: Empirical hypotheses, *Econometrica* 84, 1345–1404.

Lan, Chunhua, Fabio Moneta, and Russ Wermers, 2018, Holding horizon: A new measure of active investment management, Working paper, University of Maryland.

Merton, Robert C., 1987, A simple model of capital market equilibrium with incomplete information, *Journal of Finance* 42, 483–510.

Pástor, Ľuboš, and Robert F. Stambaugh, 2012, On the size of the active management industry, *Journal of Political Economy* 120, 740–781.

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

Pástor, Ľuboš, Robert F. Stambaugh, and Lucian A. Taylor, 2017, Do funds make more when they trade more?, *Journal of Finance* 72, 1483–1528.

Polkovnichenko, Valery, 2005, Household portfolio diversification: A case for rank-dependent preferences, *Review of Financial Studies* 18, 1467–1502.

Pollet, Joshua, and Mungo Wilson, 2008, How does size affect mutual fund behavior? *Journal of Finance* 63, 2941–2969.

Pontiff, Jeffrey, 2006, Costly arbitrage and the myth of idiosyncratic risk, *Journal of Accounting and Economics* 42, 35–52.

Reuter, Jonathan, and Eric Zitzewitz, 2015, How much does size erode mutual fund performance? A regression discontinuity approach, Working paper, Boston College.

Sadka, Ronnie, and Anna Scherbina, 2007, Analyst disagreement, mispricing, and liquidity, *Journal of Finance* 62, 2367–2403.

Sensoy, Berk A., 2009, Performance evaluation and self-designated benchmark indexes in the mutual fund industry, *Journal of Financial Economics* 92, 25–39.

Shleifer, Andrei, and Robert W. Vishny, 1997, The limits of arbitrage, *Journal of Finance* 52, 35–55.

Stambaugh, Robert F., 2014, Presidential address: Investment noise and trends, *Journal of Finance* 69, 1415–1453.

Stambaugh, Robert F., Jianfeng Yu, and Yu Yuan, 2015, Arbitrage asymmetry and the idiosyncratic volatility puzzle, *Journal of Finance* 70, 1903–1948.

Van Nieuwerburgh, Stijn, and Laura Veldkamp, 2010, Information acquisition and underdiversification, *Review of Economic Studies* 77, 779–805.

Warner, Jerold B., and Joanna Shuang Wu, 2011, Why do mutual fund advisory contracts change? Performance, growth, and spillover effects, *Journal of Finance* 66, 271–306.

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

Yan, Xuemin, 2008, Liquidity, investment style, and the relation between fund size and fund performance, *Journal of Financial and Quantitative Analysis* 43, 741–768.

Yan, Xuemin, and Zhe Zhang, 2009, Institutional investors and equity returns: Are short-term institutions better informed?, *Review of Financial Studies* 22, 893–924.

Zhu, Min, 2018, Informative fund size, managerial skill, and investor rationality, *Journal of Financial Economics* 130, 114–134.

Appendix A. Proofs

Proof that the most liquid portfolio is the benchmark portfolio:

Starting from equation (3), we solve the following constrained minimization problem:

$$\min_{\{w_i\}} \sum_{i=1}^{N_M} \frac{w_i^2}{m_i} \qquad \text{subject to } \sum_{i=1}^{N_M} w_i = 1 \ , \tag{A1}$$

where N_M is the number of stocks in the benchmark. The problem is convex, so the first-order conditions describe the minimum. Denoting the optimal portfolio weights by $\tilde{w}_i$ and the Lagrange multiplier by ζ, the first-order conditions are $\frac{2\tilde{w}_i}{m_i} - \zeta = 0$, so that $\tilde{w}_i = \frac{\zeta m_i}{2}$. Substituting into the constraint yields $\sum_{i=1}^{N_M} \frac{\zeta m_i}{2} = 1$, which implies $\frac{\zeta}{2} \sum_{i=1}^{N_M} m_i = 1$, which in turn implies $\zeta = 2$, which then gives $\tilde{w}_i = m_i$.

A different proof, which is instructive in its own right, relies on a perturbation argument. Consider a portfolio with liquidity L. We perturb this portfolio by buying a bit of stock i and selling a bit of stock j, so the new portfolio weights are $w_i^* = w_i + u$ and $w_j^* = w_j - u$, where $u > 0$ and all other weights remain the same. The portfolio's illiquidity changes to

$$\begin{aligned} \left(L^{-1}\right)^* &= \sum_{n \notin \{i,j\}} \frac{w_n^2}{m_n} + \frac{(w_i+u)^2}{m_i} + \frac{(w_j-u)^2}{m_j} \\ &= L^{-1} + 2u\left(\frac{w_i}{m_i} - \frac{w_j}{m_j}\right) + u^2\left(\frac{1}{m_i} + \frac{1}{m_j}\right) \ . \end{aligned} \tag{A2}$$

If the original portfolio is the benchmark portfolio, for which $w_i/m_i = w_j/m_j = 1$, it follows immediately that any perturbation increases portfolio illiquidity: $(L^{-1})^* > L^{-1}$.

Proof of equation (23):

First, define $m = \sum_{i=1}^{N} m_i$ and note that

$$m = \sum_{i=1}^{N} m_i = \sum_{i=1}^{N} \frac{M_i}{M} = \frac{\sum_{i=1}^{N} M_i}{\sum_{i=1}^{N_M} M_i} = \frac{N}{N_M} \times \frac{1}{N} \sum_{i=1}^{N} \frac{M_i}{\frac{1}{N_M}\sum_{j=1}^{N_M} M_j} \ . \tag{A3}$$

Second, rearrange the inverse of portfolio liquidity from equation (3) as follows:

$$\begin{aligned} L^{-1} &= \sum_{i=1}^{N} \frac{w_i^2}{m_i} = \frac{1}{m}\sum_{i=1}^{N} \frac{w_i^2}{m_i^*} = \frac{1}{m}\sum_{i=1}^{N} m_i^* \left(\frac{w_i}{m_i^*}\right)^2 = \frac{1}{m}\mathrm{E}^*\left\{\left(\frac{w_i}{m_i^*}\right)^2\right\} \\ &= \frac{1}{m}\left[\left(\mathrm{E}^*\left\{\frac{w_i}{m_i^*}\right\}\right)^2 + \mathrm{Var}^*\left(\frac{w_i}{m_i^*}\right)\right] = \frac{1}{m}\left[1 + \mathrm{Var}^*\left(\frac{w_i}{m_i^*}\right)\right] \ , \end{aligned} \tag{A4}$$

where E^* is the expectation with respect to the m^* measure. Combining equations (A3) and (A4) yields equation (23).

Proof of the statement from Section 5.1 that leads to equation (30):

Substituting $g(T,L) = T^{\psi}L^{\beta}$ into equation (32), we obtain fee revenue as

$$F = \left(\frac{1}{\theta} f^{\gamma-1}\left[\mu T^{\psi-\lambda}L^{\beta+\phi} - fT^{-\lambda}L^{\phi}\right]\right)^{\frac{1}{\gamma-1}} \quad . \tag{A5}$$

The fund maximizes F by choosing not only f (see equation (31)) but also T and L:

$$\frac{\partial F}{\partial T} = \frac{1}{\gamma-1}\left(F^{\gamma-1}\right)^{\frac{1}{\gamma-1}-1}\left(\frac{1}{\theta}\right) f^{\gamma-1}\left[\mu(\psi-\lambda)T^{\psi-\lambda-1}L^{\beta+\phi} + f\lambda T^{-\lambda-1}L^{\phi}\right] = 0 \tag{A6}$$

$$\frac{\partial F}{\partial L} = \frac{1}{\gamma-1}\left(F^{\gamma-1}\right)^{\frac{1}{\gamma-1}-1}\left(\frac{1}{\theta}\right) f^{\gamma-1}\left[\mu(\beta+\phi)T^{\psi-\lambda}L^{\beta+\phi-1} - f\phi T^{-\lambda}L^{\phi-1}\right] = 0\,. \tag{A7}$$

Setting the bracketed terms equal to zero, we obtain

$$\mu\, g(T,L) = \frac{\lambda}{\lambda-\psi}\, f \tag{A8}$$

$$\mu\, g(T,L) = \frac{\phi}{\phi+\beta}\, f\ , \tag{A9}$$

where equation (A8) follows from equation (A6) and equation (A9) follows from equation (A7). Recall that maximizing F with respect to f leads to equation (31): $\mu\, g(T,L) = \frac{\gamma}{\gamma-1}\, f$. Combining that equation with equations (A8) and (A9), we see that for all three first-order conditions to have a solution, we must have

$$\frac{\lambda}{\lambda-\psi} = \frac{\phi}{\phi+\beta} = \frac{\gamma}{\gamma-1}\ , \tag{A10}$$

which implies $\psi = \lambda/\gamma$ and $\beta = -\phi/\gamma$, which in turn immediately implies equation (30).

Proofs of statements from Section 6:

First, we prove that the cross-sectional correlation between fund size and expense ratio is negative as long as $\beta_{\mu,f} < (\gamma-1)/\gamma$. Take logs in equation (33), so that

$$\ln(A) = -\ln(f) + \frac{\gamma}{\gamma-1}\ln(\mu) + \text{constant}\ , \tag{A11}$$

and note that

$$\begin{aligned}\text{Cov}(\ln(A),\ln(f)) &= \text{Cov}(-\ln(f) + \frac{\gamma}{\gamma-1}\ln(\mu),\ln(f)) \\ &= \frac{\gamma}{\gamma-1}\text{Cov}(\ln(\mu),\ln(f)) - \text{Var}(\ln(f))\ .\end{aligned} \tag{A12}$$

This covariance is negative if $\text{Cov}(\ln(\mu),\ln(f))/\text{Var}(\ln(f) < \frac{\gamma-1}{\gamma}$, or $\beta_{\mu,f} < \frac{\gamma-1}{\gamma}$.

Second, we prove that the correlation between g and f is positive as long as $\beta_{\mu,f} < 1$. Take logs in equation (31), so that

$$\ln(g) = \ln(f) - \ln(\mu) + \ln(\frac{\gamma}{\gamma-1})\ , \tag{A13}$$

and note that

$$\begin{aligned}\text{Cov}(\ln(g),\ln(f)) &= \text{Cov}(\ln(f)-\ln(\mu),\ln(f)) \\ &= \text{Var}(\ln(f))-\text{Cov}(\ln(\mu),\ln(f))\ . \end{aligned} \tag{A14}$$

This covariance is positive if $\text{Cov}(\ln(\mu),\ln(f))/\text{Var}(\ln(f)<1$, or $\beta_{\mu,f}<1$.

Finally, we prove that the correlation between g and A is negative as long as $\beta_{\mu,A}<\gamma-1$. Take logs in equation (38), so that

$$\ln(g)=\frac{1}{\gamma-1}\ln(\mu)-\ln(A)+\text{constant}\ , \tag{A15}$$

and note that

$$\begin{aligned}\text{Cov}(\ln(g),\ln(A)) &= \text{Cov}(\frac{1}{\gamma-1}\ln(\mu)-\ln(A),\ln(A)) \\ &= \frac{1}{\gamma-1}\text{Cov}(\ln(\mu),\ln(A))-\text{Var}(\ln(A))\ . \end{aligned} \tag{A16}$$

This covariance is negative if $\text{Cov}(\ln(\mu),\ln(A))/\text{Var}(\ln(A)<\gamma-1$, or $\beta_{\mu,A}<\gamma-1$.

Proof of the statement from footnote 19:

Suppose an active portfolio is blended with a passive benchmark so that $\omega\in[0,1]$ is the weight on the active portfolio and $1-\omega$ is the weight on the benchmark. The active portfolio has liquidity L and weights w_i; the benchmark has liquidity of one and weights m_i. The blended portfolio's weights are $\widetilde{w}_i=\omega w_i+(1-\omega)\,m_i$. Its illiquidity is

$$\begin{aligned}\widetilde{L}^{-1} &= \sum_i\frac{\widetilde{w}_i^2}{m_i} \\ &= \sum_i\frac{\omega^2w_i^2+2\omega\,(1-\omega)\,w_im_i+(1-\omega)^2\,m_i^2}{m_i} \\ &= \omega^2L^{-1}+2\omega\,(1-\omega)\left(\sum_i\frac{w_im_i}{m_i}\right)+(1-\omega)^2 \\ &= \omega^2L^{-1}+1-\omega^2. \end{aligned} \tag{A17}$$

In words, the blended portfolio's illiquidity is a weighted average of the illiquidities of the active portfolio and the benchmark, where the weights are ω^2 and $1-\omega^2$. Also note that $\widetilde{L}^{-1}\leq L^{-1}$: indexing a part of the portfolio reduces the portfolio's illiquidity.

Appendix B. Data

To construct our sample of actively managed U.S. domestic equity mutual funds, we begin with the 1979–2011 dataset constructed by Pástor, Stambaugh, and Taylor (2015), whose detailed description is in the online Data Appendix to that paper. We expand the dataset by merging it with the Thomson Reuters dataset of fund holdings and adding data from 2012 through 2014. We exclude funds identified by CRSP or Morningstar as index funds, funds whose name contains the word "index," and funds classified by Morningstar as funds of funds. We exclude fund-month observations with expense ratios below 0.1% per year since they are extremely unlikely to belong to actively managed funds. Finally, we exclude fund-month observations with lagged fund size below $15 million in 2011 dollars. We aggregate share classes belonging to the same fund.[22]

When computing portfolio weights w, we drop all fund holdings that are not included in our definition of the market portfolio, which is guided by the holdings of Vanguard's Total Stock Market Index fund. This fund tracks the CRSP US Total Market Index, which is designed to track the entire U.S. equity market. We find that 98.9% of the fund's holdings are either ordinary common shares (CRSP share code, *shrcd*, with first digit equal to 1) or REIT shares of beneficial interest ($shrcd = 48$). We therefore define the market as all CRSP securities with these share codes. This definition includes foreign-incorporated firms ($shrcd = 12$), many of which are deemed domestic by CRSP (they make up 1.4% of the Vanguard fund's holdings), but it excludes securities such as ADRs ($shrcd = 31$) and units or limited partnerships (*shrcd* first digit equal to 7). A fund's holding can fall outside the market if its CUSIP cannot be linked to the CRSP database (1.0% of the Vanguard fund's holdings), or if the security is in CRSP but outside our definition of the market (0.1% of the Vanguard fund's holdings). These holdings mainly represent cash, bonds, and other non-equity securities. For the median (average) fund/month observation in our sample, 2.3% (3.5%) of holding names and 1.9% (3.1%) of holding dollars are outside the market.

Monthly fund returns, net of expense ratio, are from CRSP and Morningstar. Following Pástor, Stambaugh, and Taylor (2015), we require that CRSP and Morningstar agree closely on a fund's return; otherwise we set it to missing.

For any fund-level variable requiring holdings data, we set the variable to missing if there is a large discrepancy in a fund's AUM between our CRSP/Morningstar database and the Thomson Reuters holdings database. We compute the ratio of the fund's AUM according

[22]Many mutual funds offer multiple share classes, which represent claims on the same underlying assets but have different fee structures. Different share classes of the same fund have the same Morningstar FundID. We aggregate all share classes of the same fund. Specifically, we compute a fund's size by summing AUM across the fund's share classes, and we compute the fund's expense ratio, returns, and other variables by asset-weighting across share classes.

to CRSP/Morningstar to the fund's AUM obtained by adding up all the fund's holdings from Thomson Reuters. If this ratio exceeds 2.0 or is less than 0.5, we set all holdings-based measures to missing. This filter drops the holdings-based variables for 3.4% of fund/quarter observations. We suspect that some of these large discrepancies are due to poor links between Thomson Reuters and CRSP/Morningstar.