Journal of Financial Economics 00 (2000) 000-000

Liquidity, investment ability, and mutual fund structure

Vikram Nanda[1], M. P. Narayanan[1,]*, Vincent A. Warther[2]

[1]*University of Michigan Business School, Ann Arbor, MI 48109, USA*
[2]*Lexecon, Inc. Chicago, IL 60604, USA*

(Received 7 May 1999; final version received 4 January 2000)

Abstract

We develop a model of the mutual fund industry in which the management fees and loads charged by actively managed open-end funds and average fund returns are determined endogenously in a competitive market setting. It is shown that heterogeneity in managerial skills at investing and minimizing costs, and the existence of investor clienteles with differing liquidity and marketing needs, gives rise to a variety of open-end fund structures that differ in the average return delivered to investors. Managers choose a fund's structure to maximize the rents they capture from their ability, taking into account the effect on investor flows. In equilibrium, funds that constrain liquidity withdrawals may have to charge lower fees and share some profits in the form of higher investor returns, when there is relative scarcity of investors with low liquidity needs.

JEL classification: G23; L14; L22.

Keywords: Mutual funds, liquidity, loads, fund performance, fund returns.

---

*Corresponding author. Tel.: 734 763 5936 fax: 734 936 0274

E-mail address: mpn@umich.edu

We wish to thank an anonymous referee whose suggestions have improved the paper significantly. We would also like to thank Ronen Israel, Bill Jennings, Raj Singh, Neal Stoughton, Anjan Thakor and participants at the Finance seminars at the University of Michigan, Washington University, University of Miami, Michigan State University, University of Notre Dame, and the 1999 AFA meetings in New York, for helpful suggestions.

0304-405X/99/ $19.00

1. Introduction

Individual investors seeking liquidity, portfolio diversification, and investment expertise at a low cost are increasingly choosing mutual funds as their the investment vehicle.[1] Investors differ, however, in their preferences and liquidity needs. To target specific investor clienteles, funds have developed a variety of structures, fee arrangements, and distribution channels. Our objective is to understand the diversity in the structure and performance of open-end mutual funds. To that end, we develop a model of the mutual fund industry in which the management fees and the loads charged by actively managed open-end funds are determined endogenously in a competitive market setting. The model shows how heterogeneity, in terms of managerial ability and the existence of multiple investor clienteles with differing liquidity and marketing needs, gives rise to a variety of open-end fund structures that differ in terms of the average return delivered to investors.

A starting premise of the model is that some managers are skilled at selecting investments and managing costs and are, thereby, able to generate excess returns compared to individual investors. In this setting, managers seek to maximize the rents from their ability by optimally choosing the fund's load structure and setting the management fee. An important consideration is that the funds are exposed to the stochastic liquidity demands of investors, a situation which adversely affects their performance. Unexpectedly high or low liquidity demands affect the return of a fund as managers must either liquidate or acquire assets at inopportune times. This demand reduces the average profits made by the fund and, hence, the rents that the fund manager can capture. Because of this situation, managers who form open-end funds have an incentive to attract investors with low anticipated liquidity needs. Managers can achieve this goal by structuring the fund as a load fund to discourage investors

---

[1] As of December 1999, there were 7791 mutual funds managing $6.8 trillion in assets, as reported by the Investment Company Institute. This amount is comparable to the assets of all commercial banks ($6 trillion), and 36% more than the assets of all private pension plans ($5 trillion). As of the end of 1999, equity mutual funds held 18% of all U.S. equities, by value (Flow of Funds of the United States, Board of Governors, Federal Reserve System).

with high liquidity needs. The disadvantage of this type of load fund, however, is that if investors with low liquidity needs are relatively scarce, the manager will have to share some of the rent by offering lower management fees and, in equilibrium, higher investor returns. We show that higher ability managers have a comparative advantage in attracting investors with low liquidity needs and are, therefore, more likely to form load funds. An interpretation of this result is that liquidity shocks impose a relatively greater burden on managers with higher ability. Hence, in equilibrium, lower ability managers emerge as the providers of liquidity, while higher ability managers are willing to pay a premium in order to mitigate the liquidity shocks.

While the basic model emphasizes the potential role of exit fees and other loads in screening out investors with high liquidity needs, loads have traditionally been used to pay for distribution costs, such as broker fees. This practice suggests that, in order to interpret the evidence on the performance of load vs. no-load fund types, it may be important to consider the effect of distribution methods and costs. As we discuss, some funds rely on the direct marketing approach, while others rely on the more costly method of using brokerage firms and other financial intermediaries to sell shares to investors. We explain how the model can be extended in a simple way to allow for additional investor clienteles. Loosely following Gruber (1996), we allow for investors who are sophisticated, who require relatively little by way of information and marketing effort, and those who are unsophisticated, who are relatively uninformed and require a substantial marketing effort.

The model predicts that, among open-end funds targeting the more sophisticated investors, the average return on load funds will be higher than that of no-load funds. However, when we consider the universe of all open-end funds, the average rate of return on load funds need not be greater than that of no-load funds. The empirical evidence on the performance of different fund types, in particular the difference in performance between load and no-load funds, is discussed in light of this prediction. In our view, the evidence is generally supportive of the model's predictions.

The model also predicts that management fee size will not necessarily be related to fund performance. This result provides a rationale for the empirical evidence, and is a consequence of the fact that funds catering to investors with low liquidity needs attract investors by offering lower management fees and higher expected returns. Another prediction is that the minimum loads charged by open-end funds will be positively related to the rate of return investors expect from such funds. This situation arises because the greater the expected rate of return offered to investors, the greater the exit fee needed to discourage investors with high liquidity needs from investing in the fund.

While the paper models open-end funds, an alternative approach to insulating a fund from liquidity shocks is to structure it as a closed-end fund. In a closed-end structure, investors meet their liquidity needs by selling shares in a secondary market, rather than withdrawing assets directly from the fund. We provide an intuitive discussion of the costs and benefits of the closed-end structure, relative to the open-end form. Liquidity shocks do not affect the assets managed by a closed-end fund, lending that structure an apparent advantage. However, the closed-end form has a significant disadvantage as well, in that it does not have the built-in monitoring mechanism that comes from allowing investors to freely withdraw their money or invest more as new information arrives about managerial ability.

In the mutual fund literature, the paper closest to ours is Chordia (1996), which develops a model of risk-averse investors facing stochastic liquidity needs. A monopolist fund is shown to enhance investor welfare because the fund is able to lower liquidity costs by diversifying across investors with less than perfectly correlated liquidity needs. The main feature that the papers have in common is that load types can be useful in separating investors with different liquidity needs. However, the present paper differs from Chordia (1996) in underlying approach and implications. The focus in our paper is on a competitive mutual fund industry with risk-neutral investors and heterogeneous managerial ability. In contrast to Chordia (1996), in which managerial fees are exogenously determined and the size of the monopolistic fund is limited only by the availability of investor funds, fees in our model are endogenously determined by competitive managers who anticipate the effects of the fees on the size of fund inflows and on their earnings. This structure enables us to derive cross-sectional implications regarding the mutual fund industry that relate fund profits, returns, management fees, liquidity costs, and loads. Also, in contrast to Chordia (1996), the cost of providing liquidity results in an optimal fund size that is positively related to managerial ability, providing a link between managerial ability and fund size. In our model, loads separate investor types by imposing direct penalties on investors that withdraw funds early. In contrast, in Chordia (1996), loads and liquidity costs are assumed to be borne equally by all investors in a fund, whether or not they withdraw their investment. A consequence of this difference is that, in our model, it is possible for more talented managers to form load funds, attracting investors with low liquidity needs by offering lower fees and higher returns. This consequence predicts the absence of a monotonic relation between performance and fees, a prediction that cannot be obtained using the approach in Chordia (1996).

Our paper draws upon and discusses the evidence from studies on the performance of mutual funds. Edelen (1999) develops and tests a model of the relation between liquidity trading and fund performance. This paper provides empirical support for the notion underlying our analysis, that liquidity costs can be significant for mutual funds. Particularly interesting is the finding that fund performance is adversely affected, regardless of the direction of the trade.

Other studies that are of interest in terms of the predictions of our model include Ippolito (1989), Elton, Gruber, Das, and Hlavka (1993), Gruber (1996), Carhart (1997), and Zheng (1999). These papers also explore the performance of mutual funds and compare the performance of load and no-load funds. The evidence from these studies is discussed in terms of the model's predictions. A number of papers have examined the flow of money into mutual funds [see, e.g., Ippolito (1992), Sirri and Tufano (1998), Hendricks, Patel, and Zeckhauser (1994), Warther (1995), Gruber (1996), and Zheng (1999)]. Our paper contributes to this literature by modeling the interaction of flows, performance, and load structure.

The paper proceeds as follows. Section 2 develops the basic model. Section 3 extends the basic model to allow for investor clienteles with different liquidity requirements and for heterogeneity in managerial ability. Empirical implications of the model are discussed in Section 4. The use of alternative distribution channels is considered in Section 5. Section 6 provides a brief discussion of other structures, such as closed-end funds, that can mitigate the cost of liquidity shocks. Section 7 concludes.

2. Basic model of an open-end mutual fund

In this section, we develop a basic model of an actively managed open-end mutual fund. The model analyzes how the optimal management fee and the resulting size of the fund are affected by the existence of stochastic liquidity demands, and by the ability of the fund's manager.

2.1. Description of the model

We consider a risk-neutral world with a large number of identical investors. The risk-neutral assumption simplifies the model and allows us to abstract from issues that are outside the focus of our analysis. Each investor has one dollar, which can be invested in either a risk-free asset or an open-end mutual fund. The rate of return on the risk-free asset is normalized to zero.

A starting premise of the model is that some investment managers have the ability to invest in risky assets and earn returns in excess of the risk-free rate. The managers have no resources, however, and are forced to raise cash from outside investors. The investment ability of these managers can be described in a variety of ways. Their superior ability may be from being able to trade more efficiently than the typical investor, to identify misvalued securities, or to predict changes in economic conditions or industry prospects more accurately than individual investors. Managers may also possess the ability to improve operating efficiency by minimizing processing and bid-ask spread costs, or the ability to improve tax efficiency.

The notion that some fund managers have superior investing ability is supported by recent empirical studies. Chevalier and Ellison (1996) finds that fund managers who attended colleges whose students achieved higher average SAT scores earn higher returns than managers from less selective colleges. Chalmers, Edelen, and Kadlec (1999) finds that funds with lower spread costs and better tax efficiency have higher returns, implying that trading efficiency is an aspect of managerial ability. In addition, a number of recent studies have identified subsets of fund managers who appear to outperform the market. Grinblatt and Titman (1992), Hendricks, Patel, and Zeckhauser (1993), Wermers (1996), and Gruber (1996) find that investors can earn higher returns by investing in the subset of funds that have done well in the past.[2] For our purpose, it is only necessary that there exists a subset of fund managers with superior ability.

There are F fund managers, which is defined to be a small number compared to the aggregate funds available from investors. Hence, in equilibrium, funds can always achieve their desired size without having to offer investors an expected return greater than zero. It is assumed for now that all managers have the same ability, denoted by $a > 0$, and that this ability is common knowledge. Investors are assumed to have no ability, and their only alternative to investing in mutual funds is to invest in securities with a zero rate of return.

The sequence of events in the formation of funds, investment of money acquired, and disposal of returns is shown in Figure 1. At Time 1, the fund is formed and the fund manager sets the management fee $g > 0$ as a fraction of the assets under management. The linear fee structure is consistent with the legal requirement in the U.S. on the fee structures for mutual funds.[3] Following the formation of the fund, investors purchase shares, such that each investor contributes a dollar and receives a unit share in the fund. The money acquired by the fund at Time 1 is denoted V_1, and is observable to market participants. After the funds flow in, the manager optimally chooses a positive proportion, w, of V_1 that is initially invested in risky assets. The remaining funds are invested in liquid risk-free assets.

[2] Early studies generally concluded that, on average, fund managers did not deliver superior after-fee returns to investors, a conclusion that may be sustained even before accounting for fees (e.g., Sharpe, 1966; Jensen, 1968; Malkiel, 1995). Some recent studies, however, have challenged this conclusion (e.g., Ippolito, 1989; Ferson and Schadt, 1996; Ferson and Warther, 1996).

[3] In the U.S., mutual funds are not legally permitted to set performance-based management fees. For most funds, the fee is set as a fixed percentage of the assets under management. See Das and Sundaram (1998) for an analysis of the optimality of fulcrum fee structures, which are legally permissible in the U.S. Also, see Lynch and Musto (1998) for a discussion of the optimality of various fee structures.

2.2. Stochastic liquidity withdrawals

At Time 2, a fraction of investors suffers a "liquidity shock," and this subset redeems their investment from the fund. The notion of a liquidity shock is that the investor receiving such a shock places a relatively low value on future income compared to current income, or, in other words, has a high discount factor. This liquidity shock is systematic. Therefore, the risk imposed on the mutual fund industry by this shock cannot be mitigated even if investors diversified their investments across different mutual funds. The fact that some investors place a relatively high value on current rather than future cash flows is precisely the reason the funds are structured to satisfy liquidity needs of investors, rather than to prohibit liquidity withdrawals, despite the costs imposed by stochastic withdrawals. Warther (1995) presents a study of unexpected aggregate net flows of mutual funds and their market impact.

The fraction of investors that stay on in the fund, m, is a random variable that is given by

$$m = \overline{m}h \tag{1}$$

where $0 < \overline{m} < 1$, and h is a random variable distributed over the support $[0, 1/\overline{m})$ with mean 1 and standard deviation s. At Time 1, investors and managers only have information about the distribution of m, and not its actual level. The proportion of investors who stay on in the fund is distributed with a mean $\overline{m}$ and standard deviation $\overline{m}\, s$ over the support [0, 1).

An important notion underlying the model is that the provision of liquidity can be costly and can induce some funds to choose structures that moderate the impact of liquidity shocks. The cost of providing liquidity can include the transactions costs of selling and buying fund assets to meet unexpected withdrawals, the tilting of the portfolio toward liquid, but otherwise unattractive, investments, or both. The cost of providing liquidity also includes the opportunity cost of holding liquid risk-free assets in anticipation of the withdrawals. Other costs of liquidity withdrawals include adverse selection costs and higher taxes due to unexpected capital gains. An implicit assumption here is that it is too costly for mutual funds to borrow funds to fully satisfy the liquidity withdrawals. This assumption is, by and large, consistent with the observed behavior of funds, though they do use bank borrowing on occasion or, if they belong to a fund family, borrow from other funds in the family. Consistent with these notions, Edelen (1999) finds that trades enacted to address liquidity needs reduce the performance of funds, regardless of the direction of the trade. In the context of our model, this finding is equivalent to assuming that it is costlier to purchase risky assets at Time 2 rather than Time 1 and that it is costlier to sell risky assets at Time 2 rather than Time 3. In particular, it is assumed that if the amount of liquidity withdrawals differs from the amount of liquid risk-free securities held by the fund, the fund incurs a cost, and the average cost is proportional to the absolute value of the differential. This assumption captures the reasonable notion that the fund is able to find relatively inexpensive ways to respond to small liquidity shocks by, for example, selling the most liquid securities or borrowing funds. For larger shocks, however, marginal costs, and, therefore, average costs, will be higher as the fund is forced to liquidate the less liquid securities in its portfolio. The notion that the cost of the trade is positively related to the size of the trade is consistent with the findings of Chan and Lakonishok (1997) and Kiem and Madhavan (1997). Specifically, d, the average cost per unit of unexpected withdrawals, is taken to be

$$d = b\left|(1-w)V_1 - (1-m)V_1\right| = bV_1\left|m - w\right|, \tag{2}$$

where b, a positive constant, is the cost factor. The above specification assumes that the average cost of unexpected withdrawals is symmetric with respect to the sign of the deviation from the value of the risk-free asset held. This assumption simplifies the analysis without compromising the general applicability of the model. Therefore, the total cost of unexpected withdrawals is given by

$$d\, V_1\left|m - w\right| = bV_1\left|m - w\right| \times V_1\left|m - w\right| = bV_1^{\,2}(m - w)^2 . \tag{3}$$

For simplicity, it is assumed that the cost of unexpected withdrawals is paid from the proceeds of the fund's investment at Time 3.

If the liquidity shock can be easily reversed by funds deposited by investors at Time 2, the cost of such shocks would be effectively eliminated. To ensure against this possibility, it is assumed that the magnitude of the liquidity withdrawal from a fund is not publicly observed, at least within the relevant time frame in which investment decisions are to be made. This assumption is reasonable because, in practice, fund flows take place continuously over time, and the information on fund flows is available to investors only with a lag, and that too imperfectly. It should be noted that, although we propose the strict assumption that fund flows are not observed by investors, a weaker assumption, that flows are not perfectly observable, is sufficient. Implicitly, it is being assumed that the manager cannot convey his information in the relevant time frame in which investment decisions are required to be made. Alternatively, a reasonable argument can be made that announcements by the manager will lack credibility, and will not be sufficient to resolve the information problem. At this stage, the management fee has already been set, and all managers have the incentive to maximize the funds under management. Hence, regardless of the actual liquidity shock, all managers will want to claim that withdrawals are large, if such a statement would trigger greater cash inflows. Therefore, in equilibrium, such announcements would be completely discounted by investors, at least within the relevant time frame. It is also assumed for now that investors are not penalized for liquidity withdrawals at Time 2.

2.3. Investment return

Let V_2 represent the money in the fund at Time 2 after liquidity withdrawals, that is, $V_2 = mV_1$. At Time 3, returns from the investment are realized, net of any liquidity costs. From these proceeds, the management fee is deducted, and the amount V_3 is paid out to investors. V_3 is given by

$$V_3 = V_2(1 + a + e) - gV_2 - bV_1^2(m - w)^2 . \tag{4}$$

The first term represents the return stemming from managerial ability a, and e is a random error with a mean of zero. Eq. (4) implicitly assumes that any amount in liquid risk-free securities remaining after redemption is optimally reinvested in risky assets. The second and third terms represent, respectively, the management fees and the costs imposed by the stochastic liquidity shocks. We analyze the equilibrium by first examining the manager's asset allocation decision at Time 1, and then work backwards to determine the quantity of funds that investors will provide to the fund manager.

2.4 Portfolio decision at Time 1

At Time 1, after the management fee has been set and investors have deposited the amount V_1 in the fund, the fund manager allocates V_1 between the risky assets and the risk-free asset. Since the management fee has already been set, it is reasonable to assume that the manager acts in the interest of investors and chooses the allocation that maximizes expected proceeds at Time 3. Taking expectations with respect to Time 1, we have from Eq. (4) that

$$EV_3 = (1+a)EV_2 - gEV_2 - bV_1^2 E(m-w)^2 , \tag{5}$$

where $E(.)$ is the expectation operator and expectations are taken at Time 1 after V_1 has been determined. Eq. (5) can be rewritten as,

$$EV_3 = (1+a-g)EV_2 - bV_1^2[\bar{m}^2 s^2 + (\bar{m}-w)^2]. \tag{6}$$

From Eq. (6), given V_1 and g, the optimal portfolio allocation is given by

$$w^* = \bar{m} . \tag{7}$$

In other words, the proportion of the assets invested in the risk-free asset, $1 - w$, equals $1 - \bar{m}$, the expected liquidity withdrawal at Time 2. Under the assumption that the manager allocates the portfolio optimally, we can now determine the size of the fund, which is the quantity of funds that investors will be willing to provide the manager, in equilibrium.

2.5. The Equilibrium fee and fund size

At Time 1, fund managers set the management fee, g, to maximize their expected profit of $p_1 \equiv gEV_2$. Let the expected average rate of return to investors in the fund, net of management fees, be denoted by r, where r is given by

$$1 + r = \frac{E[(1-m)V_1 + V_3]}{V_1} . \tag{8}$$

As a group, investors provide funds V_1 and receive return V_3 if they do not withdraw their money at Time 2. If they do withdraw their investment, investors receive only their original investment. The size of the fund is then determined by investors who purchase shares in the fund until the expected average rate of return, as opposed to marginal rate of return, is zero. This relation holds because all investors get the same rate of return. As each investor expects to withdraw her investment with probability (1 – $\bar{m}$), the net expected investment in risky assets is EV_2, and investors will invest in the fund to the point that the expected average rate of return, *r*, is zero. Setting $r = 0$ in Eq. (8), we get

$$EV_3 = \bar{m}V_1 = EV_2 \,. \tag{9}$$

The choice of g, the management fee, results from the following tradeoff. As g increases, the fund's expected profit tends to increase. However, the return to investors falls and, consequently, the resources made available to the manager are reduced. By assumption, the number of funds is small relative to the number of investors, and therefore each fund is able to attract sufficient investors to attain its optimal size. The manager's maximization problem is therefore

$$\underset{g}{Max}\, p_1 \equiv gEV_2 \,, \tag{10}$$

such that

$$r = 0, \text{ and} \tag{11}$$

$$w = w^* = \bar{m} \,. \tag{12}$$

Using Eqs. (6) and (9), the constraint represented by Eq. (11) can be replaced by

$$EV_2 = (1 + a - g)EV_2 - bV_1^2[\bar{m}^2 s^2 + (\bar{m} - w)^2] \,. \tag{13}$$

Using the constraint represented by Eq. (12), and replacing $\bar{m}^2 V_1^2$ by $(EV_2)^2$, the maximization problem reduces to

$$\underset{g}{Max}\, p_1 \equiv \frac{g\,(a - g)}{b\,s^2} \,. \tag{14}$$

The first order condition to the manager's problem is given by

$$g = a/2, \tag{15}$$

suggesting that the management fee should reflect the manager's ability. It is easily verified that the second-order condition for a maximum is satisfied. The expected value of the fund after liquidity withdrawals, EV_2, and the optimal size, V_1, of the fund at Time 1 are, therefore, given by

$$EV_2 = \frac{a}{2bs^2}, \tag{16}$$

and

$$V_1 = \frac{a}{2\overline{m}bs^2}, \tag{17}$$

respectively. From Eqs. (16) and (17) it follows that the optimal fund size, V_1, and the amount expected to be invested after liquidity withdrawals, EV_2, are both directly related to managerial ability, and inversely related to the liquidity cost factor (b) and the volatility of liquidity withdrawals (s). This relation reflects the fact that higher managerial ability and lower liquidity costs will attract more funds. The manager's expected profit is given by

$$p_1 = \frac{a^2}{4bs^2}. \tag{18}$$

The manager's expected profit is, therefore, increasing and convex in ability, and decreasing in the liquidity costs. It is also decreasing in the volatility of liquidity needs.

A brief explanation about the consistency between Eq. (4) and the constraint given in Eq.(11) is necessary. As mentioned previously, Eq. (4) assumes that all investment at Time 2, after the liquidity withdrawals, is in risky assets. For this assumption to be consistent with investors' expectation of a zero rate of return at Time 1, it must be that the return expected from investments at Time 2 must be negative in some states of the world. This result is achieved if there is the possibility of large unexpected withdrawals, such that the rate of return from risky assets in those states is negative. This result is also achieved if we assume that the fund incurs costs to process the accounts of investors. These costs include, but are not limited to, shareholder servicing costs, custodian and transfer-agent fees, legal and auditing fees, and directors' fees. The fund subtracts these costs from the fund inflows, and invests only the remaining funds. This setting results in investors receiving a negative rate of return, in terms of the expected return at Time 1, from any investment by the fund in risk-free assets. Therefore, even if all excess cash is optimally invested in only risky assets at Time 2, as assumed in Eq. (4), the constraint given in Eq. (11) can be satisfied if investors expect a zero rate of return at Time 1.

The manager, then, bears the anticipated costs imposed by liquidity shocks on the overall profitability of the fund. Further, the manager captures all of the rents arising from his ability. Hence, any impact on those rents due to liquidity shocks will, in equilibrium, be borne by the fund manager.

3. Investor clienteles and heterogeneous managerial ability

The model presented in the previous section provides insight into the costs of liquidity provision. Further, it shows how managerial ability and the liquidity needs of investors affect fund size and management fees. Given that the costs of providing boundless liquidity are borne by the manager, we would expect funds to seek ways to minimize exposure to liquidity shocks. To analyze the ways in which funds might compete for investors with low liquidity needs and the equilibrium that results from such competition, we now extend the model to include heterogeneity in investor liquidity needs and managerial ability. We show that, if there is heterogeneity in liquidity needs, mutual funds will compete by charging lower fees to attract investors with low liquidity needs, while using devices such as exit fees to deter investors with greater liquidity needs. In this competition, higher ability managers are able to outbid other managers for investors with low liquidity needs and, in equilibrium, these investors can capture some of the rents arising from managerial investing ability.

We introduce two classes of investors, types *L* and *H*, based on their potential liquidity needs. Investors know their own type, but no one else does. A fraction $1 - m_j$ of type j investors, $j \in \{L, H\}$, will encounter liquidity needs at Time 2 and will have to withdraw their investment from the fund. As in the previous section, only the distribution of m_j, with mean $\overline{m}_j$ and standard deviation $\overline{m}_j\, s_j$, is known at Time 1. It is assumed that

$$\overline{m}_L > \overline{m}_H > 0, \tag{19}$$

and

$$\overline{m}_L > \overline{m}_H > 0, \text{ and } q\, s_L = s_H, \quad q > 1. \tag{20}$$

The total number of type L investors is N_L and, since each investor has one dollar, N_L is also the total dollar investment by this group. The number of type H investors is assumed to be large relative to the funds sought by managers with positive ability.

We also allow fund managers to differ in terms of their ability, a. Specifically, the distribution of fund manager ability in the population can be represented by a continuous distribution function with support $[\underline{a}, A]$, $\underline{a} > 0$, and density function $f(a)$. As before, each individual manager's ability is public knowledge. We analyze the equilibrium in which some funds specialize in attracting type L investors, while other funds exclusively attract type H investors.

It can be seen from Eq. (18) that a manager's expected profit is higher if the fund can attract only type L investors, because those investors impose lower liquidity costs on the fund. One way to deter type H investors is to structure the fund as a load fund, charging an exit fee at the time of a liquidity withdrawal. This assumption can be made without loss of generality. If, for instance, exit fees were retained in the fund and distributed at Time 3 to the remaining investors, equilibrium outcomes would be unchanged regarding features such as management fees or funds invested. The fee is a fraction, x, of the amount withdrawn at Time 2, and is assumed to be retained by the manager. The exit fee is endogenously determined so that no type H investor finds it optimal to invest in a fund with an exit fee. Later, we characterize the size of the exit fee and show that alternative fee structures, such as front-end loads, are equivalent to exit fees.

We will refer to funds with exit fees as load funds, consistent with industry terminology. In practice, funds also charge loads to meet sales and marketing costs. The effect of these costs, and the use of alternative distribution channels, will be discussed in a subsequent section. We use the subscript P to denote no-load, or, "plain" funds and subscript X to denote funds with exit fees.

Since no-load funds cater solely to type H investors, who are assumed to be very large in number relative to the size of the funds, the analysis in the previous section applies for these funds. Let the optimal management fee, expected investment at Time 2, optimal size, and expected profit of no-load funds be denoted by g_P, EV_{P2}, V_{P1}, and p_{P1}, respectively. The expressions for these are given by Eqs. (15), (16), (17), and (18), respectively, replacing s with s_H.

The optimal management fee, size, rate of return to investors, and exit fees of load funds depend on the number of type L investors, N_L, relative to the number of such funds. The interesting case to consider is the one in which type L investors are scarce relative to the demand for funds by managers, enabling these investors to capture some of the rents from managerial ability. That is,

$$N_L < F\int_{\underline{a}}^{A} \frac{a}{2\overline{m}_L b s_L^2} f(a)da\,, \tag{21}$$

where the integrand in Eq. (21) is the size of the fund if the return to investors was zero, as given by Eq. (17), and F is the number of fund managers. For any given value of N_L/F, if s_L is low enough, then Eq. (21) is satisfied. Therefore, it follows that, for any given values of N_L/F and s_H, there will exist a q^* such that for all $q > q^*$, Eq. (21) is satisfied. If $q \le q^*$, and Eq. (21) is not satisfied, all funds will be load funds, allowing fund managers to capture all the rents from all investors. Therefore, the interesting case, in which more than one fund type can exist, requires that Eq. (21) to be satisfied. Hence, in the discussion that follows and for the various results below, it is implicitly assumed that this condition is satisfied.

With type L investors relatively scarce, managers of load funds must choose a combination of management fees and exit penalties that results in a positive rate of return to attract type L investors. In equilibrium, all load funds will provide the same rate of return, which is the rate that clears the market for type L investors. At this rate, the least able manager offering load funds, that is, the marginal manager, is indifferent between offering a load and a no-load fund. Denote this equilibrium rate of return to type L investors by r_X. In a competitive equilibrium, the manager of a load fund faces the problem of choosing the optimal combination of management and exit fees to maximize his expected profit, taking r_X as given. Let $\hat{g}_X$ denote the equilibrium management fee charged by a load fund, and let V_{X1} and EV_{X2} denote the corresponding fund size at Time 1 and the expected value of the fund at Time 2, after withdrawals. The load fund manager's expected profit, p_{X1}, is then given by

$$p_{X1} \equiv E\left[(1-m_L)xV_{X1} + \hat{g}_X V_{X2}\right] = \left[(1-\overline{m}_L)x + \overline{m}_L \hat{g}_X\right]V_{X1}. \tag{22}$$

Eq. (22) shows that the load fund manager's profit is the weighted average of the proceeds from management and exit fees. A fraction $(1 - m_L)$ of investors withdraw and pay the exit fee $x V_{X1}$, while the assets under management, V_{X2}, generate a management fee of $\hat{g}_X V_{X2}$. The second equality follows, since $EV_{X2} = \overline{m}_L V_{X1}$. Using g_X to denote the weighted average of the management and exit fees, we have

$$g_X \equiv (1-\overline{m}_L)x + \overline{m}_L \hat{g}_X . \tag{23}$$

The expected rate of return to investors, r_X, is given by

$$1 + r_X = E\left[\frac{(1-m_L)(1-x)V_{X1} + V_{X3}}{V_{X1}}\right]. \tag{24}$$

The fraction $(1 - m_L)$ of the funds invested at Time 1 is withdrawn at Time 2 for liquidity needs, with investors receiving $(1 - m_L)(1 - x)V_{X1}$ after exit fees. The remaining funds are invested, and the proceeds available to investors at Time 3 is V_{X3}. Substituting from Eq. (4) for V_{X3}, with suitable changes in subscripts, taking expectations, and using the fact that $w = w^* = \overline{m}_L$, we get,

$$r_X = \overline{m}_L a - g_X - b\overline{m}_L^2 s_L^2 V_{X1}. \tag{25}$$

From Eqs. (22) and (23), the load fund manager's optimization problem is then

$$\underset{g_X}{Max}\ p_{X1} \equiv g_X V_{X1} \tag{26}$$

subject to the constraint given in Eq. (25). The problem is well-defined only for non-negative fund size and fees.

The following lemma provides the solution to the above program for $a \geq r_X / \overline{m}_L$, the setting facing fund managers seeking type L investors. Managers with ability less than $r_X / \overline{m}_L$ will not form load funds since their expected profits will be negative. Proofs for all lemmas are included in the Appendix.

Lemma 1: For $a \geq r_X / \overline{m}_L$, the optimal fee, the fund size, and the expected profit of a load fund are given by

$$g_X = \frac{\overline{m}_L (a - \frac{r_X}{\overline{m}_L})}{2}, \tag{27}$$

$$V_{X1} = \frac{a - \frac{r_X}{\overline{m}_L}}{2 b s_L^2 \overline{m}_L}, \tag{28}$$

and

$$p_{X1} = \frac{(a - \frac{r_x}{\overline{m}_L})^2}{4 b s_L^2}. \square \tag{29}$$

Eq. (27) determines the weighted average of the load fund's management and exit fees. There is a minimum exit fee, derived below, that the load fund must charge to dissuade investors with high liquidity needs. Beyond this threshold, the load fund can charge any combination of management and exit fees. Note that the expected profit is monotonic in ability.

For r_X to be the unique equilibrium rate of return to investors, two conditions must be satisfied. First, if both load and no-load funds exist, the lowest ability manager offering load funds must be indifferent between offering a load fund and a no-load fund at this rate. Second, at this rate, the market must clear, such that the total amount of funds invested in load funds must equal the total amount of funds available to all type L investors. Let the marginal manager's ability be denoted by a_M. Then, the first condition implies

$$p_{X1}(a_M) = p_{P1}(a_M). \tag{30}$$

Substituting for the profit expressions of no-load and load funds from Eqs. (18) and (29), respectively, we get

$$r_X = \overline{m}_L a_M \left(1 - q^{-1}\right). \tag{31}$$

Eq. (31) shows that a unique equilibrium rate exists if a_M is unique, as we show below. Also, since expected profits of both load and no-load funds are monotonically increasing in ability, as set out in Eqs. (18) and (29), all managers with ability $a > a_M$ will form load funds, and the rest will form no-load funds.

The second condition implies that

$$F\int_{a_M}^{A} V_{X1} f(a)da = F\int_{a_M}^{A} \frac{a - \frac{r_X}{\overline{m}_L}}{2\overline{m}_L b s_L^{\ 2}} f(a)da = N_L . \tag{32}$$

We now show that there will be an unique a_M. Substituting the expression for r_X from Eq. (31), we get the following equation, the solution of which yields a_M:

$$F\int_{a_M}^{A} \frac{a - a_M\left(1 - q^{-1}\right)}{2\overline{m}_L b s_L^{\ 2}} f(a)da = N_L . \tag{33}$$

Note that the integral in the above equation is monotonic and decreasing in a_M. As $a_M \rightarrow A$, the left hand side of Eq. (33) approaches zero. Also, as $a_M \rightarrow \underline{a}$, all managers offer load funds, and, from Eq. (17), the left hand side tends to $F\int_{\underline{a}}^{A} \frac{a}{2\overline{m}_L b s_L^2} f(a)da$. Hence, as long as the condition specified in Eq. (21) is satisfied, a unique value of a_M exists to satisfy Eq. (33) in the range ($\underline{a}$, A). From Eq. (31, note that $a_M > r_X/\overline{m}_L$, and hence Lemma 1 applies to all $a > a_M$. The next two lemmas follow directly from Eqs. (31) and (33).

Lemma 2: The ability of the marginal manager offering a load fund, a_M, is decreasing in the number of type L investors relative to fund managers (N_L/F) and in the relative volatility of the liquidity needs of type H investors (q).

Lemma 3: The rate of return for investors in load funds, r_X, is decreasing in the number of type *L* investors relative to fund managers (N_L/F) and increasing in the relative volatility of the liquidity needs of type *H* investors (q).

The intuition behind these results can be presented as follows. As the number of type *L* investors increases relative to fund managers, a_M decreases for two reasons. First, more load funds are necessary to absorb the greater supply of type *L* investors, resulting in a marginal manager of lower ability. Second, the increased supply of *L* investors lowers the rate of return r_X. This setting makes it attractive for managers of lower ability to choose forming load funds over forming no-load funds. As the relative volatility of the liquidity needs of type *H* investors increases, no-load funds, which are held by type *H* investors, become less profitable to form, and more managers begin forming load funds. To avoid the increased cost arising from the liquidity needs of type *H* investors, load fund managers are willing to pay a higher rate of return to type *L* investors.

The above analysis yields several interesting results. The first result follows from the fact that the profit functions in Eqs. (18) and (29) are monotonic in managerial ability.

Proposition 1: Higher-ability managers, or managers with ability $a \geq a_M$, will form load funds and attract investors with low liquidity needs by offering them an expected return greater than zero. Investors with high liquidity needs will obtain an expected return of zero from load funds. Lower-ability managers, or managers with ability $a < a_M$, will form no-load funds and attract investors with high liquidity needs who will receive an expected return of zero.

A more complete proof of Proposition 1 is included in the Appendix. An interpretation of this result is that liquidity shocks impose a relatively greater burden on managers with greater ability. Hence, lower-ability managers emerge as the providers of liquidity, while higher-ability managers are willing to pay a premium to reduce exposure to liquidity shocks. Some indirect evidence exists to show that higher-ability managers form load funds. Chalmers, Edelen, and Kadlec (1999) find that load funds have greater sensitivity to bid-ask spread costs and are more tax efficient. If we believe that operational efficiency forms one dimension of managerial ability, then the above empirical result is consistent with Proposition 1.

Corollary 1.1 follows from the fact that the load funds are offered by more able managers.

Corollary 1.1: Load funds are more profitable than no-load funds.

The expected profit of the marginal manager is the same, regardless of the type of fund she offers. Since expected profits of both types of funds are monotonic in manager ability, expected profit of any load fund manager is greater than that of a no-load fund manager.

The next result compares the returns to investors from load and no-load funds. It follows from the fact that, in equilibrium, load fund managers share some of the rents arising from their ability to attract scarce type L investors. Interpreting the results in terms of the costs of liquidity provision, the results indicate that managers with greater ability find it relatively more costly to provide liquidity, compared to managers with lower ability. Hence, in equilibrium, lower-ability managers are more willing to bear the costs of liquidity provision, while higher-ability managers compete aggressively to reduce such liquidity costs by paying a premium in terms of expected investor return.

Proposition 2: The rate of return to investors in load funds is greater than that available from no-load funds. This rate of return is decreasing in the number of type L investors relative to fund managers, and increasing in the relative volatility of the liquidity needs of type H investors.

Next, we examine the size of the exit fee necessary to discourage type H investors from entering a load fund. To discourage type H investors, the fee must be set so that their expected rate of return from investing in load funds is less than the expected rate of return from investing elsewhere, which is zero. A type H investor expects with probability $(1 - \overline{m}_H)$ to withdraw her investment to meet liquidity needs, thereby earning a rate of return of $(1-x)$. With probability $\overline{m}_H$, she leaves her money in the fund. The expected return on a dollar invested at Time 2 is given by EV_{X3}/EV_{X2} = $EV_{X3}/\overline{m}_L V_{X1}$. From Eq. (24), it can be seen that

$$\frac{EV_{X3}}{\overline{m}_L V_{X1}} = \frac{1}{\overline{m}_L}\left[(1+r_X) - (1-\overline{m}_L)(1-x)\right].$$

Therefore, the expected return to a type *H* investor if she invests in a load fund is given by

$$(1-\overline{m}_H)(1-x) + \frac{\overline{m}_H}{\overline{m}_L}\left[(1+r_X) - (1-\overline{m}_L)(1-x)\right]. \quad (34)$$

This return must be less than one to dissuade a type *H* investor from investing in a load fund. Simplifying the expression in Eq. (34) and setting it equal to less than one, we get the minimum exit fee that must be charged,

$$x > \frac{\overline{m}_H}{\overline{m}_L - \overline{m}_H} r_X . \quad (35)$$

Eq. (35) suggests that the minimum exit fee is positively related to the return investors receive from load funds. Using Lemma 2, this relation implies that the minimum exit fee is also positively related to q, the relative volatility of the liquidity needs of type *H* investors, and to the relative scarcity of type *L* investors. The next proposition summarizes these results.

Proposition 3: The minimum loads imposed by load funds are, other factors kept the same, positively related to (i) the rate of return investors expect from such funds; (ii) the relative volatility of the liquidity needs of type *H* investors; and (iii) the relative scarcity of type *L* investors.

While we have chosen to model the load as an exit fee, there are a number of ways in which front-end or back-end loads could be imposed to screen out investors with high liquidity needs. For example, suppose the management fee is g_X and the exit fee is x for withdrawals at Time 2. An equivalent way of screening out investors with high liquidity needs is to impose a front-end load of x at Time 1, which is retained by the fund manager, while reducing the management fee to $(g_X - x)$. To see that the two fee structures are equivalent, note that, under either structure, investors withdrawing funds at Time 2 will receive only $(1-x)$ dollars for every dollar invested at Time 1. Investors staying with the fund until Time 3 will pay g_X to the fund manager under either structure. Hence, both structures will yield the same fund size and profitability.

4. Empirical implications

In this section, we discuss some of the model's empirical implications and their relation to the existing empirical literature. The existing empirical work classifies funds based on whether or not they charge a load. To relate the implications of our model to the empirical literature, we employ this classification of funds in the discussion below.

Implication 1: On average, load funds will tend to attract investors with low liquidity needs, while no-load funds will tend to attract investors with high liquidity needs.

This implication follows directly from Proposition 1 and the analysis in Section 3 above. Consistent with our view that loads can be used to screen investors with different anticipated liquidity needs, Chordia (1996) documents that front-end and back-end loads dissuade redemption. In this connection, it is noteworthy that load funds will often waive loads when the funds are unlikely to be withdrawn for liquidity needs, such as for IRA accounts.

Implication 2: Expected profits of load funds will be greater than those of no-load funds.

This implication follows from the fact that, in equilibrium, load funds are managed by more able managers (Corollary 1.1).

Implication 3: Load funds will offer higher rates of return to investors than no-load funds.

The empirical evidence on this issue needs to be interpreted with some care because load charges are often imposed as a means of paying for distribution costs to brokers. We discuss the model's predictions and the empirical evidence more fully in the next section, after we discuss the role of different distribution channels in the context of our model.

Implication 4: The greater the relative uncertainty about investor liquidity needs, the greater the rate of return received by investors in load funds.

Implication 4 arises because increased relative uncertainty about investor liquidity needs decreases the profit fund managers expect from no-load funds (Eq. (18)), thus providing them greater incentive to attract investors with low liquidity needs (Lemma 3). Fund mangers attract these investors by offering load funds with higher expected rates of return, and the ability of the marginal manager offering a load fund is higher.

This implication could be tested with a time-series analysis. Proxies for the relative uncertainty about investor liquidity needs, such as the ratio of measured volatility of liquidity flows to no-load mutual funds in a given month or year to the corresponding figure for load funds, should be positively correlated with returns to open-end load funds over time. A cross-sectional test of the implication may also be possible. Insofar as segmentation exists in the fund industry, the relation between uncertainty in investor redemptions and returns to open-end load funds may vary across fund segments, such that segments with higher liquidity uncertainty have higher returns.

Implication 5: Mutual fund fees are not necessarily related to fund performance.

If one equates managerial ability with a fund's return, this implication might seem surprising, because it implies that high-ability managers do not charge higher fees. While our model implies that load funds will provide a higher rate of return, it does not necessarily follow that their fees are higher. As can be seen from Eq. (28), the combination of management and exit fee is positively related to managerial ability and negatively related to the rate of return. Even though, on average, load fund managers have higher ability, they must also share some of the rents from their ability with investors in the form of higher returns. Thus, the fee of load funds, which demonstrate higher performance, is not necessarily greater than that of no-load funds, which demonstrate lower performance. This result is consistent with the empirical evidence. Gruber (1996) and Carhart (1997) find that fees are unrelated to fund performance.

Implication 6: Minimum loads charged by open-end funds are positively related to the rate of return investors expect from such funds.

From Proposition 3, we see that the greater the rate of return to investors, the greater the exit fee needs to be to discourage investors with high liquidity needs from investing in the fund. As with Implication 4, this implication could be tested with a time-series analysis. Changes in loads over time should be positively correlated with the difference between the returns to load funds and no-load funds.

Implication 7: Lower costs of managing liquidity shocks, or a decrease in the uncertainty associated with investor liquidity needs, will lead to a lower proportion of funds charging loads.

As Eq. (32) demonstrates, a decrease in the liquidity cost factor, b, or in the volatility of liquidity withdrawals, s_L, will result in an increase in a_M, implying that fewer funds would be organized as load funds. Developments such as an increase in the efficiency of security markets and markets for short-term financing would, therefore, be predicted to result in fewer funds charging loads. This result suggests that the large increase in no-load funds compared to loads funds in recent years may be, at least partly, the result of a relative decrease in the cost of liquidity provision, as financial markets have become more efficient and complete.

5. Investor clienteles based on sophistication and cost of access

While we have emphasized the potential role of exit fees and other loads in screening out investors with high liquidity needs, loads are also used to pay for distribution costs, such as broker fees. This structure suggests that, in order to interpret the evidence on the performance of different fund types, it may be important to consider the effect of distribution methods and costs. In this section, we discuss an extension of the model to allow for investor clienteles that differ in terms of access and service costs, resulting in the use of different distribution channels.

Broadly speaking, mutual funds have tended to use one of two channels (see Sirri and Tufano, 1998) to distribute their product to investors. Some funds rely on the direct marketing approach. This approach is usually taken by no-load funds, and relies on relatively lower cost distribution methods like advertising and direct mailings. Other funds rely on brokerage firms and other financial intermediaries to sell shares to investors. The distribution costs associated with these funds are higher. Brokers are often compensated out of the funds raised from a combination of front-end loads, and higher 12b-1, and higher service fees. Brokerage firms provide services to their customers, such as information about suitable funds, advice, and adjunct services like financial planning (see Sirri and Tufano, 1998). Therefore, despite the higher fees, these funds survive because some investors find the broker's services valuable. To distinguish the loads charged to meet the sales and service costs from those levied to discourage liquidity withdrawals, we call the former "sales loads" and the latter "liquidity loads."

The model developed above can be extended to allow for investors who are sophisticated, requiring relatively little information and marketing effort, and those who are unsophisticated, requiring a substantial marketing effort. This terminology is loosely derived from Gruber (1996), which argues that a significant fraction of mutual fund investors are "disadvantaged" or "unsophisticated" investors whose money flows are, at least in part, affected by advertising and advice from brokers. For our purposes, it is simplest to assume that the only relevant difference between these two classes of investors is that the sales and servicing costs are higher for the less sophisticated investors. These less sophisticated investors are, however, perfectly rational in terms of their investment decisions. An example of differences in levels of clientele sophistication might be the difference between institutional and retail investors. While we assume that all investors require the same rate of return, the model can easily accommodate Gruber's (1996) hypothesis that the reservation rate of return of less sophisticated investors is less than that offered by index funds.

While the intuition behind our extension of the previous model is uncomplicated, a presentation of the details of the model extension would be cumbersome. Therefore, rather than providing a full rendering of those details here, we provide an outline of the results and a fuller discussion of the implications of such an extension.

Allowing for different levels of investor sophistication results in there being four types of investor clienteles, differentiated on the basis of their liquidity needs and their sophistication. We show that, under appropriate conditions, each clientele is served by a fund that caters to its needs, resulting in four classes of funds. Among these fund classes, the only class that is constituted of no-load funds is the one that caters to sophisticated investors with high liquidity needs. The other three classes have load charges for sales, liquidity, or both. The class of funds catering to sophisticated investors with low liquidity needs will levy liquidity loads. The funds catering to unsophisticated investors with high liquidity needs will levy sales loads, while those catering to unsophisticated investors with low liquidity needs will levy liquidity and sales loads. Figure 2 summarizes the loads and the expected returns of the four classes of funds.

In terms of performance, funds catering to sophisticated investors with low liquidity needs will provide the highest expected return. These funds want to attract low liquidity investors who are relatively scarce in number and impose no sales and service costs on the fund. Funds that cater to sophisticated investors with high liquidity needs, which are structured as no-load funds, and funds that attract unsophisticated investors with low liquidity needs will also provide positive expected returns. These positive expected returns reflect investor scarcity as well as savings in sales and service costs. The first of the above two fund classes does not incur sales and service charges, while liquidity costs are absent in the funds of the second class. The expected return to unsophisticated investors with low liquidity needs is lower than that offered to sophisticated investors with low liquidity needs, due to the sales and service charges incurred by funds seeking to access and service the former types. The only class of funds expected to provide a zero rate of return is the one catering to unsophisticated investors with high liquidity needs, of whom there is assumed to be a large number.

In summary, two classes of load funds and the class of no-load funds will offer positive expected rates of return, while one class of load funds offers a zero rate of return. Therefore, when we consider all open-end funds, catering to both sophisticated and unsophisticated investors, it is not obvious whether load or no-load funds have the higher average expected rate of return, since returns depend on the proportions of the different clienteles in the economy. Implication 3, developed in the earlier section, can be interpreted as applying only to sophisticated investors. Therefore, Implication 3 can be modified as follows.

Implication 3a: Among open-end funds that target sophisticated investors, load funds will offer higher rates of return to investors than no-load funds. However, when we consider the universe of open-end funds, the average rate of return on load funds need not be greater than that of no-load funds.

The empirical evidence on mutual fund performance is broadly consistent with the model's implication. While Ippolito (1989) reports that load funds have significantly higher returns on average than no-load funds, this finding is challenged by Elton, Gruber, Das, and Hlavka (1993), who find problems with the paper's methodology. Carhart (1997) also finds, contrary to Ippolito (1989), that load funds under-perform no-load funds, on average, due to their higher expenses. Further, Gruber (1996) finds no significant difference, on average, between load and no-load fund performance. Consequently, in the overall sample, there is little evidence that, on average, load funds outperform no-load funds. Zheng (1999) adjusted for industry-wide flows in her study of the flow between funds, and provides additional evidence on this implication. She finds that, on average, adjusted for industry-wide flows, load funds with positive inflows outperform no-load funds with positive inflows. Drawing on the insight of Gruber (1996), that flows between funds are probably from sophisticated investors reacting to evidence on managerial ability, adjusted inflows may identify funds that are attracting flows from a sophisticated clientele. The evidence of Zheng (1999), then, is consistent with the implication that load funds targeting sophisticated investors should perform better, on average, than no-load funds that do the same.

The implications in Section 4 continue to hold, even when there is heterogeneity in investor sophistication. Load funds now include funds that levy either a liquidity load, a sales or service charge, or both. Since sophisticated investors with high liquidity needs are the only ones that invest in no-load funds, on average, load funds will tend to attract investors with low liquidity needs (see Implication 1). Both types of funds that cater to unsophisticated investors are load funds, and their expected profit is at least as much as that of no-load funds. Therefore, on average, expected profits of load funds will be greater than that of no-load funds (see Implication 2).

Implication 6 now applies only to funds that charge liquidity loads. As noted earlier, liquidity loads may be designed as front-end or back-end loads. Similarly, sales or service charges can also be front-end or back-end loads. It is observed, however, that some mutual funds levy a back-end load that declines over time, such that the longer the investment remains in the fund, the lower the exit fee. It may be reasonable to interpret such back-end loads and redemption restrictions as liquidity loads. If so, we should see the returns from such funds increase with the load charged. The other implications are unaffected by this extension. Implication 7 can be readily extended to include sales loads, as well.

6. A brief discussion on closed-end funds and liquidation costs

We have seen that exit fees can be used to screen investors and reduce the exposure of an open-end fund to liquidity shocks. A somewhat different manner of insulating a fund may be to structure it as a closed-end fund. In a closed-end structure, investors meet their liquidity needs by selling their shares in a secondary market, rather than withdrawing their assets directly from the fund. By making certain additional assumptions, the current model of open-end fund structure choice can be modified to include closed-end funds. A detailed discussion of closed-end funds is beyond the scope of this paper. An intuitive discussion of the costs and benefits of the closed-end structure, relative to the open-end form, is provided below.

The clear benefit of a closed-end structure is that the investment strategy of the manager and the expected returns on the assets under management are not subject to investor liquidity shocks. On the cost side, there are the transaction charges associated with trading closed-end fund shares. The more significant problem with the closed-end form, however, may be that it does not have the built-in monitoring mechanism that comes from the ability of investors to freely withdraw their money or invest more as new information arrives about managerial ability. In some cases, the inability to withdraw money from the fund may result in money being left in the hands of low-quality managers. In other cases, higher-ability managers may receive substantially less funds than could be invested profitably. Hence, the likelihood that funds may not be of optimal size, once new information on managerial ability arrives, may make the closed-end form less attractive when there is substantial uncertainty or learning about managerial ability over time.[4] An implication of this setting is that, other things equal, we would expect managers of closed-end funds to be well-established professionals about whose ability investors are fairly certain. In such cases, the benefit from investors retaining the flexibility to withdraw funds in response to new information may be less than the benefit from insulating the fund from liquidity shocks.

A natural implication is that closed-end funds will specialize in holding assets with high liquidation costs. Casual empiricism is consistent with this implication, given that most funds specializing in holding foreign securities, which are generally less liquid than U.S. securities, are closed-end. Similarly, most funds that hold real estate, which is inherently more illiquid than traded securities, tend to be real estate investment trusts, which are similar to closed-end funds. At the other extreme, the overwhelming majority of funds that invest in highly liquid exchange-traded securities, where liquidation costs are low, are formed as open-end funds, suggesting the relative importance of learning

[4] Another cost of closed-end funds derives from private benefits appropriated by blockholders (see Barclay, Holderness, and Pontiff, 1993).

about managerial ability over time. When assets are relatively liquid, only those managers who can convince investors of their ability, through, for example, a strong track record, will tend to form closed-end funds. Consistent with this observation, note that many closed-end funds holding relatively liquid assets are founded by "big-name" investors. An example of this case is Warren Buffet, whose firm, Berkshire Hathaway, has the essential features of a closed-end fund although it is technically not a closed-end fund.

7. Conclusion

We have developed a model of the mutual fund industry in which the management fees and the loads charged by actively managed open-end funds are determined endogenously in a competitive market setting. In the model, managers choose a fund's structure to maximize the rents they can capture from their ability, taking into account the effect their decisions have on investor flows. The setting is one in which fund performance is affected by liquidity costs caused by the stochastic liquidity demands of investors. The model yields several interesting and testable implications regarding fund performance, loads, management fees, and fund profits of open-end funds.

References

Barclay, M., Holderness, C., Pontiff, J., 1993. Private benefits from block ownership and discounts on closed-end funds. Journal of Financial Economics 33, 263-292.

Carhart, M., 1997. On the persistence in mutual fund performance. Journal of Finance 52, 57-82.

Chan, L., Lakonishok, J., 1997. Institutional equity trading costs: NYSE versus NASDAQ. Journal of Finance 52, 713-735.

Chalmers, J., Edelen, R., Kadlec, G., 1999. Evaluating mutual fund managers by the operational efficiency of their trades. Unpublished working paper. Wharton School of Business, Philadelphia, PA.

Chevalier, J., Ellison, G., 1999. Are some mutual fund managers better than others? Cross-sectional patterns in behavior and performance. Journal of Finance 54, 875-899.

Chordia, T., 1996. The structure of mutual fund charges. Journal of Financial Economics 41, 3-39.

Das, S., Sundaram, R., 1998. On the regulation of mutual fund fee structures. Unpublished working paper. New York University, NY.

Edelen, R., 1999. Investor flows and the assessed performance of open-end mutual funds. Journal of Financial Economics 53, 439-466.

Elton, E., Gruber, M., Das, S., Hlavka, M., 1993. Efficiency with costly information: A reinterpretation of evidence from managed portfolios. The Review of Financial Studies 6, 1-22.

Ferson, W., Schadt, R., 1996. Measuring fund strategy and performance in changing economic conditions. Journal of Finance 51, 425-462.

Ferson, W., Warther, V., 1996. Evaluating fund performance in a dynamic market. Financial Analysts Journal 52, 20-28.

Grinblatt, M., Titman, S., 1992. The persistence of mutual fund performance. Journal of Finance 47, 1977-1984.

Gruber, M., 1996. Another puzzle: The growth in actively managed mutual funds. Journal of Finance 51, 783-810.

Hendricks, D., Patel, J., Zeckhauser, R., 1993. Hot hands in mutual funds: short-run persistence of relative performance, 1974-1988. Journal of Finance 48, 93-130.

Hendricks, D., Patel, J., Zeckhauser, R., 1994. Investment flows and performance: Evidence from mutual funds, cross-border investments, and new issues. In: Sato, R., Levich, R., Ramachandran, R. (Eds.), Japan, Europe and International Financial Markets: Analytical and Empirical Perspectives. Cambridge University Press, New York, NY, 51-72.

Ippolito, R., 1989. Efficiency with costly information: A study of mutual fund performance. Quarterly Journal of Economics 104, 1-23.

Ippolito, R., 1992. Consumer reaction to measures of poor quality: Evidence from the mutual fund industry. *Journal of Law and Economics* 35, 45-70.

Jensen, M., 1968. The performance of mutual funds in the period 1945-1964. Journal of Finance 23, 389-416.

Kiem, D., Madhavan, A., 1997. Transaction costs and investment style: An inter-exchange analysis of institutional equity trades. Journal of Financial Economics 46, 265-292.

Lynch, A., Musto, D., 1998. Understanding fee structures in asset management business. Unpublished working paper. New York University, NY.

Malkiel, B., 1995. Returns from investing in equity mutual funds, 1971-1991. Journal of Finance 50, 549-572.

Sharpe, W., 1966. Mutual fund performance. Journal of Business 39, 119-138.

Sirri, E., Tufano, P., 1993. Competition and change in the mutual fund industry. In: Hayes III, S. L. (Ed.), Financial Services: Perspectives and Challenges. Harvard Business School Press, Boston, MA.

Sirri, E., Tufano, P., 1998. Costly search and mutual fund flows. Journal of Finance 53, 1589-1622.

Warther, V., 1995. Aggregate mutual fund flows and security returns. Journal of Financial Economics 39, 209-235.

Wermers, R., 1996. Momentum investment strategies of mutual funds, performance persistence, and survivorship bias. Unpublished working Paper. University of Colorado at Boulder.

Zheng, L., 1999. Is money smart? A study of mutual fund investors' fund selection ability. Journal of Finance 54, 901-933.

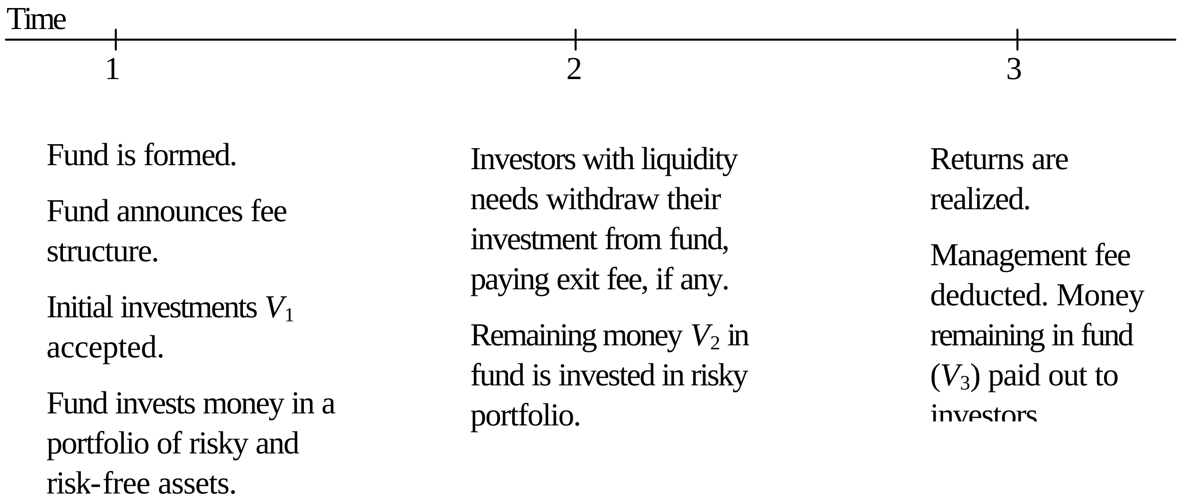

Figure 1. Sequence of events for mutual fund creation and competition

Is the investor sophisticated?

		Yes	No
Investor liquidity needs	High	No load fee Medium return	Sales load Low return
	Low	Liquidity load High return	Liquidity & Sales load Medium return

Figure 2. Loads and expected returns of funds catering to different clienteles.

Liquidity loads are fees charged to discourage liquidity withdrawals by fund investors. Sales loads are fees charged to meet the sales and service costs incurred to attract investors.

Appendix

A.1. Proof of Lemma 1

From Eq. (25) it follows that

$$V_{X1} = \frac{1}{b\overline{m}_L s_L^2}(\overline{m}_L a - g_X - r_X)\,. \tag{A1}$$

Multiplying the above expression by g_X yields the expected profit. The first order condition with respect to g_X yields Eq. (27). Substituting the expression for g_X from Eq. (27) in Eq. (A1), we get Eq. (28). Using Eqs. (27) and (28), we get Eq. (29).

A.2. Proof of Lemma 2

We provide a sketch of the proof here. As N_L/F increases, a_M cannot increase or stay the same. If it does, the value of the integral in Eq. (33) will decrease, violating the equality. Therefore, a_M must decrease. Using the same line of reasoning, it can be seen that a_M decreases as q increases.

A.3. Proof of Lemma 3

We provide a sketch of the proof here. As N_L/F increases, we know that a_M decreases (Lemma 2). It follows from Eq. (31) that r_X decreases.

As q increases, we know that a_M decreases (Lemma 2). As the lower limit of the integral in Eq. (33) decreases, the integrand must decrease to maintain the equality. This relation in turn implies that $a_M(1-q^{-1})$, and hence, r_X must increase.

A4. Proof of Proposition 1

To prove the proposition, it needs to be shown that, for $a \geq a_M$, $p_{X1}(a) \geq p_{P1}(a)$, and for $r_X < a < a_M$, $p_{X1}(a) < p_{P1}(a)$. Note that we are only interested in the range $a > r_X / \overline{m}_L$, because the fund size positive only in this range. Comparing Eqs. (18) and (29) for $p_{P1}(a)$ and $p_{X1}(a)$, respectively, and substituting for r_X from Eq. (31), it can be seen that

$$p_{X1}(a) \geq p_{P1}(a) \Leftrightarrow \frac{\left(a - a_M(1-q^{-1})\right)^2}{4b\,s_L^2} \geq \frac{a^2}{4b\,s_H^2} \Leftrightarrow a \geq a_M \,. \qquad \text{(A2)}$$

It also follows that the reverse inequality is true, as long as $a > r_X$.