Faculty Research

University of Michigan Business School

WORKING PAPER SERIES

Liquidity, Investment Ability, and Mutual Fund Structure

Vikram Nanda
University of Michigan Business School

M. P. Narayanan
University of Michigan Business School

Vincent A. Warther
Lexecon, Inc.

Working Paper 00-003

Liquidity, Investment Ability, and Mutual Fund Structure

Vikram Nanda*

M. P. Narayanan*

Vincent A. Warther**

April 1997

(Revised: January 2000)

*University of Michigan Business School, Ann Arbor, MI 48109. **Lexecon, Inc., Chicago, IL.

Contact author: M. P. Narayanan, Ph: (734) 763-5936; Fax: (734) 936-0274; email: 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 Michigan, Washington University, University of Miami, Michigan State University, University of Notre Dame, and the 1999 AFA meetings in New York, for helpful suggestions.

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.

JUN 27 2000

LIQUIDITY, INVESTMENT ABILITY, AND MUTUAL FUND STRUCTURE

I. Introduction

Mutual funds are increasingly the investment vehicle of choice for the individual investor seeking liquidity, portfolio diversification and investment expertise at a low cost.[1] Investors differ, however, in terms of their preferences and liquidity needs – with the result that funds appear to have developed a variety of structures, fee arrangements and distribution channels to target specific investor clienteles. Our objective in the paper is to understand the diversity in the structure and performance of open-end mutual funds. Towards 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 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 fund's performance is adversely affected by its exposure to the stochastic liquidity demands of investors. Unexpectedly high or low liquidity demands affect the return of a fund as managers must either liquidate or acquire assets at inopportune times.

---

[1] As of October 1998, there were 7294 mutual funds managing 5.1 trillion dollars in assets (as reported by the Investment Company Institute). This is comparable to the loans and investments of all commercial banks (5.5 trillion dollars), and 40% more than the assets of all private pension plans (3.6 trillion dollars). As of the third quarter of 1998, equity mutual funds held 16% of all U.S. equities, by value (Board of Governors, Federal Reserve System).

This reduces the average profits made by the fund and, hence, the rents that the fund manager can capture. Because of this, managers who form open-end funds have an incentive to attract investors with low anticipated liquidity needs. They can achieve this by structuring the fund as a load fund to discourage investors 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 high liquidity need investors, loads have also been used traditionally to pay for distribution costs such as broker fees. This suggests that, in order to interpret the evidence on the performance of different fund types (i.e., load vs. no-load), 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' and require relatively little by way of information and marketing effort and those who are 'unsophisticated', relatively uninformed and require a substantial marketing effort.

The model predicts that among open-end funds that target 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 fees 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 them 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 investors' expected rate of return from such funds. The reason is that the greater the expected rate of return 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, it is apparent that 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. Closed-end funds have the advantage that liquidity shocks do not affect the assets under management. However, the closed-end form has a significant disadvantage as well: 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 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 since, by diversifying across investors with less than perfectly correlated liquidity needs, the fund is able to lower liquidity costs. The main feature that the papers have in common is that loads can be useful in separating investors with different liquidity needs. The underlying approach and implications are, however, different. The focus in our paper is on a competitive mutual fund industry with risk-neutral investors and heterogeneity in managerial ability. In contrast to Chordia (1996) where managerial fees are exogenously determined and the size of the monopolistic fund is limited only by the availabilit 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, unlike in 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 early. In contrast, in Chordia (1996), loads and liquidity costs are assumed to be borne *equally* by all investors in a fund, whether they withdraw or not. A consequence is that in our model the more able managers form load funds, attracting low liquidity investors by offering lower fees and higher returns. This predicts the absence of a monotonic relation between performanc 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. An interesting paper by Edelen (1999) develops and tests a model of the relation between liquidity trading and fund performance. The 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, irrespective of the direction of the trade.

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 (1998), that explore the performance of mutual funds and comparing 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 (1998)]. Our paper contributes to this literature by modeling the interaction of flows, performance and load structure.

The paper proceeds as follows. Section II develops the basic model. Section III 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 IV. The use of alternative distribution channels is considered in section V. Section VI provides a brief discussion of other structures, such as closed-end funds, that can mitigate the cost of liquidity shocks. Section VII concludes.

II. 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.

a. Model set up

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 not the focus of our analysis. Each investor has one dollar which can be invested either in a risk-free asset or in 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 viewed in a variety of ways: they may be able to trade more efficiently than the typical investor, identify misvalued securities, or predict changes in economic conditions or industry prospects more accurately than investors. They 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 attende colleges with higher average SAT scores earn higher returns than managers from less selecti 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 (1993), Hendricks, Pa and Zeckhauser (1993), Wermers (1996), and Gruber (1996) find that investors can earn high returns by investing in the subset of funds that have done well in the past.[2] For our purpose, i is only necessary that there exists a subset of fund managers with superior ability.

There are F fund managers, which is 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 hence 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 $\gamma > 0$ as a fraction of the assets under management. The linear fee

[2] Early studies generally concluded that, on average, fund managers did not deliver superior returns to investors after fees, and possibly even before 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), and Ferson and Warther (1996)).

structure is consistent with the legal requirement in the U.S. on the fee structures for mutual funds.[3] Following this, investors buy shares in the fund, each investor contributing a dollar and receiving 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 the proportion $w > 0$ of V_1 that is initially invested in risky assets, the rest being invested in liquid risk-free assets.

b. Stochastic Liquidity Withdrawals

At time 2, a fraction of investors suffers a "liquidity shock" and redeem 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 (*i.e.*, has a high discount factor). This is a systematic liquidity shock and, therefore, cannot be diversified away by investors spreading their investment needs across mutual funds.[4]

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

$$\mu = \bar{\mu}\eta$$

where $0 < \bar{\mu} < 1$, and η is a random variable distributed over the support $[0, 1/\bar{\mu})$ with mean 1 and standard deviation σ. At time 1, investors and managers know only the distribution of μ which has mean $\bar{\mu}$ and standard deviation $\bar{\mu}\,\sigma$ 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 such shocks. The cost of

[3] In the U.S., mutual funds are not legally permitted to set performance based management fees and 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 the 'fulcrum' fee structures that are legally permissible in the U.S. Also, see Lynch and Musto (1998) for a model of optimality of various fee structures.

[4] The fact that some investors place a relatively high value on current rather than future cashflows is precisely the reason the funds are structured to *satisfy* investors liquidity needs, rather than to *prohibit* liquidity withdrawals – despite the costs imposed by stochastic withdrawals. See Warther (1995) for a study of unexpected aggregate net flows of mutual funds and their market impact.

providing liquidity can include the transactions costs of selling and buying fund assets to mee unexpected withdrawals and/or the tilting of the portfolio toward liquid, but otherwise unattractive, investments.[5] It also includes the opportunity cost of keeping liquid risk-free assets in anticipation of the withdrawals. Other costs of liquidity withdrawals include advers selection costs and higher taxes due to unexpected capital gains and losses. Consistent with these notions, Edelen (1999) finds that liquidity related trades reduce the performance of fund irrespective of the direction of the trade. In the context of our model, this 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 that the average cost is proportional to the absolute value c the differential. This captures the reasonable notion that the fund is able to find relatively inexpensive ways to deal with small liquidity shocks by, say, selling the most liquid securities or borrowing funds. For larger shocks, however, marginal (and average) costs are higher as th 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 trade size is consistent with the findings of Chan and Lakonishok (1997) and Kiem and Madhavan (1997). Specifically, δ, the average cost per unit of unexpected withdrawals, is taken to be

$$\delta = \beta\left|(1-w)V_1 - (1-\mu)V_1\right| = \beta V_1\left|\mu - w\right|, \quad (1$$

where $\beta > 0$ 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 o the risk-free asset held. This simplifies the analysis without loss of generality about the implications of the model. Therefore, the total cost of unexpected withdrawals is given by

---

[5] An implicit assumption here is that it is too costly for mutual funds to borrow funds to fully satisfy the liquidity withdrawals. This 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.

$$\delta V_1|\mu - w| = \beta V_1|\mu - w| \times V_1|\mu - w| = \beta V_1^2(\mu - w)^2. \tag{2}$$

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 fund flows from investors at time 2, this would effectively eliminate the cost of such shocks. 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 is a reasonable assumption since, 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 though we make the stylized assumption that fund flows are not observed by investors, the weaker assumption that they are not perfectly observable is sufficient.[6] It is also assumed for now that investors are not penalized for liquidity withdrawals at time 2.

c. Investment Return

Let V_2 represent the money in the fund at time 2 after liquidity withdrawals, *i.e.*, $V_2 = \mu V_1$. 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

$$V_3 = V_2(1 + a + \varepsilon) - \gamma V_2 - \beta V_1^2(\mu - w)^2. \tag{3}$$

The first term represents the return stemming from managerial ability a and ε is a zero-mean random term. Equation (3) implicitly assumes that any amount in liquid risk-free securities

[6] 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. The reason is that, at this stage, the management fee has already been set and all managers have the incentive to maximize the funds under management. Hence, irrespective of the actual liquidity shock, all managers will want to claim that withdrawals are large - if they could use this to trigger greater cash inflows. Therefore, in equilibrium, such announcements would be completely discounted – at least within the relevant time frame.

remaining after redemption is (optimally) reinvested in risky assets. The second and third term 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 working backwards to determine the quantity of funds that investors will provide the manager.

d. 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 Equation (3) that

$$EV_3 = (1+a)EV_2 - \gamma EV_2 - \beta V_1^2 E(\mu - w)^2, \tag{4}$$

where $E(.)$ is the expectation operator and expectations are taken at time 1 after V_1 has been determined. Equation (4) can be rewritten as,

$$EV_3 = (1+a-\gamma)EV_2 - \beta V_1^2[\bar{\mu}^2\sigma^2 + (\bar{\mu}-w)^2]. \tag{5}$$

It is clear from Equation (5) that, given V_1 and γ, the optimal portfolio allocation is given by

$$w^* = \bar{\mu}. \tag{6}$$

In other words, the proportion of the assets invested in the risk-free asset, $1 - w$, equals $1 - \bar{\mu}$, 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, *i.e.,* the quantity of funds that investors will be willing to provide the manager in equilibrium.

e. The Equilibrium Fee and Fund Size

At time 1, fund managers set the management fee γ to maximize their expected profit $\pi_1 \equiv \gamma EV_2$. Let the expected *average* rate of return (net of management fees) to investors in the fund be denoted by r, where r is given by

$$1 + r = \frac{E[(1-\mu)V_1 + V_3]}{V_1}. \tag{7}$$

As a group, investors provide V_1 and receive V_3 if they do not withdraw their money at time 2; if they do withdraw, they receive their original investment back. 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 since all investors get the same rate of return. As each investor expects to withdraw her investment with probability $(1 - \bar{\mu})$, the net expected investment in risky assets is EV_2 and investors invest in the fund to the point that the expected average rate of return r is zero. Setting $r = 0$ in Equation (7), we get

$$EV_3 = \bar{\mu}V_1 = EV_2. \tag{8}$$

The choice of γ results from the following tradeoff. As γ 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. Since (by assumption) the number of funds is small relative to the number of investors, each fund is able to attract sufficient investors to attain its optimal size. The manager's maximization problem is therefore,

$$\underset{\gamma}{Max}\ \pi_1 \equiv \gamma EV_2$$

such that

$$r = 0. \tag{9}$$

$$w = w^* = \bar{\mu}. \tag{10}$$

Using equations (5) and (8), constraint (9) can be replaced by

$$EV_2 = (1 + a - \gamma)EV_2 - \beta V_1^2[\bar{\mu}^2\sigma^2 + (\bar{\mu} - w)^2].$$

Using constraint (10), and replacing $\bar{\mu}^2 V_1^2$ by $(EV_2)^2$, the maximization reduces to

$$\underset{\gamma}{Max}\, \pi_1 \equiv \frac{\gamma(a-\gamma)}{\beta\sigma^2}. \tag{11}$$

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

$$\gamma = a/2. \tag{12}$$

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

$$EV_2 = \frac{a}{2\beta\sigma^2}, \tag{13}$$

and

$$V_1 = \frac{a}{2\bar{\mu}\beta\sigma^2}, \tag{14}$$

respectively. From equations (13) and (14) 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 (β) and the volatility of liquidity withdrawals (σ). This reflects the fact that higher managerial ability and lower liquidity costs will attract more funds. The manager's expected profit is given by

$$\pi_1 = \frac{a^2}{4\beta\sigma^2}. \tag{15}$$

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 Equation (3) and constraint (9) is in order. As mentioned, Equation (3) assumes that all investment at time 2 after the liquidity withdrawals is in risky assets. For this to be consistent with investors' expectation of 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 is achieved if there is the possibility of large unexpected withdrawals so that the rate of return from risky assets in those states is negative. It is also achieved if we assume that the fund incurs costs to process investors' account. These costs include shareholder servicing costs, custodian and transfer-agent fees, legal and auditing fees, directors' fees, etc. The fund subtracts these costs from the fund inflows and invests only the rest. This results in investors receiving a negative rate of return (in terms of *ex-ante* return at date 1), from any investment by the fund in the risk-free asset. Therefore, even if all excess cash is optimally invested only in risky assets at date 2, as assumed in Equation (3), constraint (9) can be satisfied with investors expecting a zero rate of return at date 1.

Observe that it is the manager who ends up bearing the *ex-ante* anticipated costs imposed by liquidity shocks on the overall profitability of the fund. The manager captures all of the rents from his ability; hence, any impact on those rents due to liquidity shocks will, in equilibrium, be borne by the fund manager.

III. Investor clienteles and heterogeneous managerial ability

The model presented in the previous section provides insight into the costs of liquidity provision and shows how managerial ability and the liquidity needs of investors affect fund size and management fees. Given that the costs of providing (unfettered) 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 resulting equilibrium, 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 liquidit needs. Investors know their own type but no one else does. A fraction $1 - \mu_j$ of type j investors, $j \in \{L, H\}$, will have liquidity needs at time 2 and will need to withdraw their investment from the fund. As in the previous section, only the distribution of μ_j, with mean $\bar{\mu}$ and standard deviation $\bar{\mu}_j \sigma_j$, is known at time 1. It is assumed that

$$\bar{\mu}_L > \bar{\mu}_H > 0, \text{and } \theta \sigma_L = \sigma_H, \quad \theta > 1.$$

The total number of type L investors is N_L and, since each investor has one dollar, this 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 attract type H investors exclusively.

It can be seen from Equation (15) that a manager's expected profit is higher if the fund can attract only type L investors since they impose lower liquidity costs on the fund. One way to deter type H investors is to structure the fund as a load fund; specifically, to charge an exit fee at the time of a liquidity withdrawal. This is without loss of generality. If, for instance, exi fees were retained in the fund and distributed at time 3 to investors, equilibrium outcomes would be unchanged in terms of management fees, funds invested etc. 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 i 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 and 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 funds ("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 γ_P, EV_{P2}, V_{P1}, and π_{P1}, respectively. The expressions for these are given by equations (12), (13), (14), and (15), respectively, replacing σ with σ_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,

$$N_L < F\int_{\underline{a}}^{A} \frac{a}{2\overline{\mu}_L \beta \sigma_L^2} f(a)da\,, \tag{16}$$

where the integrand in condition (16) is the size of the fund if the return to investors was zero, as given by Equation (14), and F is the number of fund managers. For any given value of N_L/F, condition (16) is satisfied if σ_L is low enough. Therefore, it follows that for any given values of N_L/F and σ_H, there will exist a θ^* such that for all $\theta > \theta^*$, condition (16) is satisfied. If $\theta \leq \theta^*$ and condition (16) is not satisfied, all funds will be load funds and fund managers will capture all the rents from all investors. Therefore, the interesting case, in which more than one fund type can exist, requires condition (16) 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. This is also the rate at which the least able manager offering load funds *i.e.,* 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 $\hat{\gamma}_X$ denote the equilibrium management fee charged by a load fund and 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 π_{X1} is then given by

$$\pi_{X1} \equiv E[(1-\mu_L)xV_{X1} + \hat{\gamma}_X V_{X2}] = [(1-\overline{\mu}_L)x + \overline{\mu}_L \hat{\gamma}_X]V_{X1}. \tag{17}$$

It can be seen from the equation above that the load fund manager's profit is the weighted average of the proceeds from management and exit fees. A fraction $(1-\mu_L)$ of investors withdraw and pay the exit fee xV_{X1}, while the assets under management, V_{X2}, generate a management fee of $\hat{\gamma}_X V_{X2}$. The second equality follows since $EV_{X2} = \overline{\mu}_L V_{X1}$. Using γ_X to denote the weighted average of the management and exit fees, we have:

$$\gamma_X \equiv (1-\overline{\mu}_L)x + \overline{\mu}_L \hat{\gamma}_X. \tag{18}$$

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

$$1 + r_X = E\left[\frac{(1-\mu_L)(1-x)V_{X1} + V_{X3}}{V_{X1}}\right]. \tag{19}$$

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

$$r_X = \overline{\mu}_L a - \gamma_X - \beta\overline{\mu}_L^2\sigma_L^2 V_{X1}. \tag{20}$$

From equations (17) and (18), the load fund manager's optimization problem is then:

$$\underset{\gamma_X}{Max}\ \pi_{X1} \equiv \gamma_X V_{X1}$$

subject to the constraint (20). The problem is well-defined only for non-negative fund size and fees.

The following lemma provides the solution to the above program for $a \geq r_X / \overline{\mu}_L$. Managers with ability less than $r_X / \overline{\mu}_L$ will not form load funds since their expected profits will be negative. Proofs are in the appendix.

Lemma 1: The optimal fee, the fund size, and the expected profit of a load fund are given by (for $a \geq r_X / \overline{\mu}_L$):

$$\gamma_X = \frac{\overline{\mu}_L(a - \frac{r_X}{\overline{\mu}_L})}{2}, \tag{21}$$

$$V_{X1} = \frac{a - \frac{r_X}{\overline{\mu}_L}}{2\beta\sigma_L^2\overline{\mu}_L}, \tag{22}$$

$$\pi_{X1} = \frac{(a - \frac{r_X}{\overline{\mu}_L})^2}{4\beta\sigma_L^2}. \square \tag{23}$$

Equation (21) 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 b 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. Secon at this rate the market must clear: 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

$$\pi_{X1}(a_M) = \pi_{P1}(a_M).$$

Substituting for the profit expressions of no-load and load funds from equations (15) and (23) respectively, we get

$$r_X = \overline{\mu}_L a_M \left(1 - \theta^{-1}\right). \qquad (2$$

It can be seen from Equation (24) that there is a unique equilibrium rate 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 (equations (15) and (23)), all managers with $a > a_M$ will form load funds and the rest will form no-load funds.

The second condition implies that

$$F\int_{a_M}^{A} V_{X1} f(a)da = F\int_{a_M}^{A} \frac{a - \frac{r_X}{\overline{\mu}_L}}{2\overline{\mu}_L \beta \sigma_L^{\ 2}} f(a)da = N_L. \qquad (2$$

We now show that there will be an unique a_M. Substituting the expression for r_X from Equation (24), we get the following equation, the solution of which yields a_M:

$$F\int_{a_M}^{A} \frac{a - a_M\left(1-\theta^{-1}\right)}{2\overline{\mu}_L \beta \sigma_L^{\ 2}} f(a)da = N_L. \qquad (26$$

Note that the integral in the above equation is monotonic and decreasing in a_M. As $a_M \to A$, the left hand side of Equation (26) tends to zero. Also, as $a_M \to \underline{a}$, all managers offer load funds and the left hand side tends to $F\int_{\underline{a}}^{A} \frac{a}{2\overline{\mu}_L \beta \sigma_L^2} f(a)da$ (from Equation (14)). Hence, as long as condition (16) is satisfied, there exists a unique value of a_M satisfying Equation (26) in the range $(\underline{a}, A)$. Note from Equation (24) that $a_M > r_X / \overline{\mu}_L$ and hence Lemma 1 applies to all $a > a_M$. The next two lemmas follow directly from equations (24) and (26).

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 type H investors' liquidity needs (θ). □

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 type H investors' liquidity needs (θ). □

The intuition behind these results is straightforward. As the number of type L investors increases relative to fund managers, a_M decreases for two reasons. First, more load funds are required 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 makes it attractive for lower ability managers to form load funds over no-load funds. As the relative volatility of type H investors' liquidity needs increases, no-load funds, which are held by type H investors, become less profitable to form and more managers switch to forming load funds. To avoid the increased cost arising from type H investor liquidity needs, load funds 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 equations (15) and (23) are monotonic in managerial ability.

Proposition 1: Higher-ability managers (managers with $a \geq a_M$) 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 (managers with $a < a_M$) will form no-load funds and attract investoı with high liquidity needs who will receive an expected return of zero. □

An interpretation of this result is that liquidity shocks impose a relatively greater burc 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. There is some indirect evidence that higher ability managers form load fun(Chalmers, Edelen, and Kadlec (1999) find that load funds have greater sensitivity to bid-ask spread costs and are more tax efficient. If we interpret that one dimension of managerial abil: is operational efficiency, then the above empirical result is consistent with Proposition 1.

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

This follows from the fact that the load funds are offered by more able managers. The expected profit of the marginal manager is the same irrespective of the type of fund she offers Since expected profits of both types of funds are monotonic in manager ability, expected prof: 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 their rents 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 t 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 o: expected investor return.

Proposition 2: The rate of return to investors in load funds is greater than that of 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 type H investors' liquidity needs. □

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 the expected rate of return to them from investing in load funds is less than the expected rate of return from investing elsewhere, *i.e.*, zero. A type H investor expects with probability $(1 - \overline{\mu}_H)$ to withdraw her investment to meet liquidity needs, thereby earning a rate of return of $(1-x)$. With probability $\overline{\mu}_H$ she leaves her money in the fund. The expected return on a dollar invested at time 2 is given by $EV_{X3}/EV_{X2} = EV_{X3}/\overline{\mu}_L V_{X1}$. From Equation (19), it can be seen that

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

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

$$(1-\overline{\mu}_H)(1-x) + \frac{\overline{\mu}_H}{\overline{\mu}_L}\left[(1+r_X) - (1-\overline{\mu}_L)(1-x)\right]. \tag{28}$$

This return must be less than one to dissuade a type H investor from investing in a load fund. Simplifying expression (28) and setting it less than one, we get the minimum exit fee that must be charged:

$$x > \frac{\overline{\mu}_H}{\overline{\mu}_L - \overline{\mu}_H} r_X. \tag{29}$$

It can be seen from the above expression that the minimum exit fee is positively related to investors' return from load funds. Using Lemma 2, this implies that the minimum exit fee is also positively related to θ, the relative volatility of type H investors' liquidity needs, and the relative scarcity of type L investors.

Proposition 3: The minimum loads imposed by load funds are, *ceteris paribus*, positively related to (i) investors' expected rate of return from such funds; (ii) the relative volatility of th type H investors' liquidity needs; 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- or back-end loads could be imposed to screen out high liquidity need investors. For example, suppose the management fee is γ_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-en load of x at time 1 (retained by the fund manager), while reducing the management fee to $(\gamma_X$ $x)$. To see that the two fee structures are equivalent, note that, under either structure, investors withdrawing at time 2 will receive only $(1 - x)$ dollars for every dollar invested at time 1. Investors staying with the fund till time 3 will pay γ_X to the fund manager under either structure. Hence, both structures will result in the same fund size and profitability.

IV. Empirical Implications

In this section, we discuss some of the model's empirical implications and their relatior to the existing empirical literature. The existing empirical work has tended to distinguish between funds based on whether they are load or no-load. To relate the implications of our model to the empirical literature, we often classify funds along similar lines in the discussion below.

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

This follows directly from Proposition 1 and the analysis in Section III 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 — *e.g.*, for IRA accounts.

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

This 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 since loads 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. □

The reason is that increased relative uncertainty about investor liquidity needs decreases fund managers' expected profit from no-load funds (Equation (15)), thus providing them greater incentive to attract investors with low liquidity needs (Lemma 3). Fund mangers do so 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 is potentially testable 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. The implication may also be testable cross-sectionally. 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 so 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 appear somewhat surprising because it would indicate 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 Equation (22), the combination of management and exit fee is positively related to managerial ability and negatively related to the rate of return. Even though load fund managers on average 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 (higher performance) is not necessarily greater than that of no-load funds (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 investors' expected rate of return 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 is potentially testable 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 can be seen from Equation (25), a decrease in the liquidity cost factor β or in the volatility of liquidity withdrawals σ_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 suggests that the large increase in no-load funds compared to loads funds in recent years may, at least partly, be the result of a relative decrease in the cost of liquidity provision, as financial markets have become more efficient and complete.

V. 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 high liquidity need investors, loads are also used to pay for distribution costs such as broker fees. This 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 (1993)) to distribute their product to investors. Some funds rely on the direct marketing approach. This is usually associated with no-load funds and reliance on relatively lower cost distribution methods such as 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 is higher and brokers are often paid through a combination of front-end loads and higher 12*b*-1 and service fees. Despite the higher fees, these funds survive because some investors find the broker's services valuable.[7] To distinguish the loads charged meet the sales and service costs from those levied to discourage liquidity withdrawals, we ca the former "sales loads" and the latter "liquidity loads."

The model developed above can be extended to allow for investors who are 'sophisticated' and require relatively little by way of information and marketing effort and those who are 'unsophisticated', relatively uninformed and require a substantial marketing effort. This terminology is loosely derived from Gruber (1996), which argues that a significar 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 (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 requi the same rate of return, the model can easily accommodate Gruber's (1996) hypothesis that th reservation rate of return of less sophisticated investors is less than that offered by index funds

While the intuition is straightforward, the details of the model extension are somewhat cumbersome. Therefore, rather than providing a full rendering, we provide an outline of the results and a fuller discussion of the implications of such an extension.[8]

Differences in 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,

---

[7] As Sirri and Tufano (1993) put it: "These services include marketing that reduces consumer need to search for funds, advice that simplifies buying decisions and possibly non-mutual fund services also provided by brokers an(financial planners."

[8] The full details of the extension are available from the authors.

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 sales and/or liquidity loads. 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. This reflects the funds' desire to attract the relatively scarce low liquidity investors who impose no sales and service costs on the fund. Funds that cater to sophisticated investors with high liquidity needs (no-load funds) and 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 as funds incur sales and service charges 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 (those that cater 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 this depends 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 load funds on average outperform no-load funds. Zheng (1998), in her study of the flow *between* funds (by adjusting for industry-wide flows), provides additional evidence on this implication. She finds that, on average, load funds with positive inflows (adjusted for industry-wide flows) outperform no-load funds with (similarly) adjusted 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. Then the evidence of Zheng (1998) is consistent with the implication that load funds that target sophisticated investors should perform better on average than no-load funds that do the same.

It is easily verified that the implications in Section IV continue to hold even when there is heterogeneity in investor sophistication. Load funds now include funds that levy a liquidity load and/or those that levy a sales/service charge. 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 (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 (Implication 2).

Implication 6 now applies only to funds that charge liquidity loads. As noted earlier, liquidity loads may be designed as front or back-end loads. Similarly, sales/service charges can also be front or back-end loads. It is observed, however, that some mutual funds levy a back-end load that declines over time: the longer the funds are kept invested, the less 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 also.

VI. 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, investor liquidity needs are met through selling of their shares in a secondary market, rather than the withdrawal of assets directly from the fund. An intuitive discussion of the costs and benefits of the closed end structure, relative to the open-end form, is provided below.[9]

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, of course, the transaction costs associated with trading

[9]By making certain additional assumptions, the current model of open-end fund structure choice can be modified to also include closed end funds. A detailed discussion of closed end funds is beyond the scope of this paper.

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 have 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 through time.[10] An implication, therefore, 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 as most funds that specialize 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 which invest in highly liquid exchange-traded securities, where liquidation costs are low, are formed as open-end funds, suggesting the relative importance of learning about managerial ability over time. When assets are relatively liquid, only those managers who can convince investors of their ability (perhaps through a

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

strong track record), will tend to form closed-end funds. Consistent with this, we observe that many closed-end funds that hold relatively liquid assets, are founded by "big-name" investors.[11]

VII. 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 on investor flows. The setting is one in which the 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.

[11] Warren Buffet is an example. While his Berkshire Hathaway is technically not a closed-end fund, for our purposes, it has the essential features of a closed-end fund.

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. Working paper, Wharton School.

Chevalier, J., Ellison, G., 1996. Are some mutual fund managers better than others? Cross-sectional patterns in behavior and performance. Working paper.

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. Working paper, NYU.

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

Elton, E., Martin, G., 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: Ryuzo Sato, Richard Levich and Rama Ramachandran, eds., Japan, Europe and international financial markets: analytical and empirical perspectives, (Cambridge University Press) 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. Working paper, New York University.

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

Sharpe, W., 1996. Mutual fund performance. *Journal of Business* 39, 119-138.

Sirri, E., Tufano, P., 1993. Competition and Change in the Mutual Fund Industry. in *Financial Services: Perspectives and Challenges*, ed. Samuel L. Hayes III, Harvard Business School Press, Boston.

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. Working Paper, University of Colorado at Boulder, Boulder, Colorado.

Zheng, L., 1998. Is Money Smart? A study of mutual fund investors' fund selection ability. *Journal of Finance* 54, 901-933.

1	2	3
Fund is formed. Fund announces fee structure. Initial investments V_1 accepted. Fund invests money in a portfolio of risky and risk-free assets.	Investors with liquidity needs withdraw their investment from fund, paying exit fee, if any. Remaining money V_2 in fund is invested in risky portfolio.	Returns are realized. Management fee deducted and rest (V_3) paid out to investors.

Figure 1: Sequence of events

		Investor Sophistication: Yes	Investor Sophistication: No
Investor liquidity Needs	High	Load: None Return: Medium	Load: Sales load Return: Low
	Low	Load: Liquidity load Return: High	Load: Liquidity & Sales load Return: Medium

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

Appendix-A

Proof of Lemma 1

From Equation (20) it follows that

$$V_{X1} = \frac{1}{\beta\overline{\mu}_L\sigma_L^2}(\overline{\mu}_L a - \gamma_X - r_X). \tag{A1}$$

Multiplying the above expression by γ_X yields the expected profit. The first order condition with respect to γ_X yields Equation (21). Substituting the expression for γ_X from Equation (21) in Equation (A1), we get Equation (22). Using equations (21) and (22) we get Equation (23).□

Proof of Lemma 2

We provide a sketch of the proof. As N_L/F increases, a_M cannot increase or stay the same. If it does, the value of the integral in Equation (26) 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 θ increases.□

Proof of Lemma 3

We provide a sketch of the proof. As N_L/F increases, we know that a_M decreases (Lemma 2). It follows from Equation (24) that r_X decreases.

As θ increases, we know that a_M decreases (Lemma 2). As the lower limit of the integral in Equation (26) decreases, the integrand in must decrease to maintain the equality. This in turn implies that $a_M(1-\theta^{-1})$, and hence r_X, must increase.□

Proof of Proposition 1:

To prove the proposition, it needs to be shown that for $a \geq a_M$, $\pi_{X1}(a) \geq \pi_{P1}(a)$ and for $r_X < a < a_M$, $\pi_{X1}(a) < \pi_{P1}(a)$. Note that we are only interested in the range $a > r_X/\overline{\mu}_L$ since only in this range is the fund size positive. Comparing expressions (15) and (23) for $\pi_{P1}(a)$ and $\pi_{X1}(a)$, respectively, and substituting for r_X from Equation (24), it can be seen that

$$\pi_{X1}(a) \geq \pi_{P1}(a) \Leftrightarrow \frac{\left(a - a_M(1-\theta^{-1})\right)^2}{4\beta\sigma_L^2} \geq \frac{a^2}{4\beta\sigma_H^2} \Leftrightarrow a \geq a_M .$$

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