POOLING EXTERNALITIES IN MUTUAL FUNDS

Woodrow Johnson†
Graduate School of Business
Columbia University

This draft: January 08, 2002

†I thank Charles Calomiris, Steve Cameron, Lawrence Glosten, Charles Himmelberg, Laurie Hodrick, Charles Jones, Joseph Stiglitz, Xiaoyan Zhang, and seminar participants the 2001 Canadian Economics Association annual meetings. Special thanks go to the anonymous mutual fund family that provided proprietary data for this study and to its many employees that helped in its interpretation. Financial support from the Center for the Marketing of Financial Services is acknowledged. Correspondence should be sent to Columbia University, 3022 Broadway, Uris Hall 311, New York, NY 10027. Email can be directed to wj40@columbia.edu.

Abstract

This article has three objectives. The first objective is to present a micro-theory model that discusses a new incentive to trade: the mutual fund pricing mechanism. Welfare theorems show that trading is socially wasteful and that there is an alternative, Pareto optimal pricing mechanism.

The second objective is to document the distribution of investment horizons in a large, proprietary panel of all investors in one no-load mutual fund family. This article is the first known direct source of evidence on no-load mutual fund holding periods. Moreover, it shows that there are observable shareholder characteristics that enable the fund to predict reliably which shareholders will be low duration and which will be high duration by applying a single-spell duration model.

The third objective is to measure the wealth transfer that results from trading activity. The shareholders are divided into two groups based on their expected duration in the fund. Simulations show that the high-duration shareholders receive 45 bp more each year in a fund restricted to their type while the low-duration shareholders receive 34 bp less in a fund restricted to their type. These pooling costs are conservative: splitting the shareholders into more than two groups substantially increases these estimates.

The article concludes with a discussion of why mutual funds choose to pool predictably different shareholders. Policy implications are noted.

1 Introduction

Mutual funds stand ready each day to sell (buy) an unlimited number of fund shares to (from) the investing public. They are virtual melting pots of investors: households, pension funds, government agencies, educational institutions, and hedge funds are among the economic agents known to invest in them. As financial intermediaries, non-money market funds are increasingly important in the U.S. economy, collectively pooling some $5,233 billion and comprising 9% of household wealth (or about a third of household financial wealth).[1] Over fifty million households own mutual funds, of which 71% own equity mutual funds. In 2000, mutual funds managed 20% of the U.S. retirement market, including nearly 50% of the IRA market.[2]

This article explores some of the economic consequences of pooled mutual fund investments from the perspective of the investor who can purchase equities directly. The term "mutual fund" is used synonymously for "no-load equity mutual fund," but the discussion readily extends to include non-equity funds. Load funds are very different, however, and are treated separately in section 9.

Suppose an investor has a one-day investment horizon. He cannot make individual equity purchases because a one-day direct investment in equities has a negative expected return after accounting for transaction costs such as bid-ask spreads and brokerage commissions.[3] A mutual fund, however, is a viable alternative for this investor because funds do not have bid-ask spreads and can be purchased without paying brokerage commissions. Why is the fund structure able to offer positive, after-cost expected returns to short-term investors when the equity market cannot? Is this a feature or a flaw of the fund?

A mutual fund is required to price fund shares *proportionately* by the 1940 Investment Company Act, as amended.[4] This means that all benefits and costs of fund ownership accrue equally to each fund share. Suppose a new shareholder makes a $1 investment in a $99 mutual fund. In equilibrium the fund must purchase $1 worth of securities for the fund's portfolio and pay the associated transaction costs. The shareholder who generated these costs pays only $1/100 = 1\%$ of them, and the existing shareholders pay the remaining $99/100 = 99\%$. This cost shifting is the mechanism that allows funds to offer positive, after-cost expected returns to short-term investors.

If the expected mutual fund holding period of each investor were the same, then an insurance argument would justify the fund's proportionate pricing structure. Investors in the following three models benefit from pooling. The Diamond-Dybvig (1983) model provides an insured return to identical investors who wish to share early-redemption risk via the bank's deposit contract. The Chordia (1996) model suggests that this same risk-sharing technology is one of the primary benefits of mutual funds. The closely related work of Nanda, Narayanan, and Warther (2000) models the mutual fund's provision of liquidity to investors with privately known redemption risk. Their economy responds to this heterogeneity through various pricing mechanisms that screen investors. In equilibrium, investors with comparable liquidity needs are grouped within funds while needs vary across funds.

There is a growing empirical literature that documents differences in trading behavior at the individual level. Grinblatt and Keloharju (2001) find that domestic investors, especially households, government agencies, and nonprofit investors, exhibit contrarian behavior while foreign investors exhibit momentum behavior in the Finnish stock market. Barber and Odean (2000) document that in a discount brokerage, taxable accounts have more trading than tax-deferred accounts. Lastly,

[1] 1999 Investment Company Institute and Federal Reserve Flow of Funds.

[2] *Mutual Fund Fact Book, 2001.*

[3] Assume round-trip transactions costs are 170 basis points and that the expected annualized equity return is 12%. Then, in expectation, the round-trip costs are not recovered for 54 days.

[4] This act is the federal statute that governs mutual fund companies; it supersedes state legislation.

Capon, Fitzsimons, and Prince (1996) sort mutual fund investors based on how the investor acquires information about funds and which attributes are considered when choosing among alternatives. They find strong evidence for the existence of both knowledgeable and naïve mutual fund investors.[5]

This article is the first to identify another set of differences among investors, investment horizons in mutual funds.[6] More importantly, it is the first to document that there are observable shareholder characteristics that enable the fund to predict which shareholders will be *low duration* (early redeemers) and which will be *high duration* (late redeemers).

A new literature examines the structure of the mutual fund as an investment vehicle for investors who can trade differently. Dickson, Shoven, and Sialm (2000) consider tax externalities. Their simulations show that investor flows can negatively impact the fund's after-tax return to the order of 121 basis points (bp) per year. Other papers discuss non-synchronous trading and identify inefficiencies in the industry's convention of pricing all mutual funds when the U.S. market closes using the last trade in the underlying assets even when the last trade was executed hours or days earlier. Chalmers, Edelen, and Kadlec (2001) show that the (mis)pricing can be profitably exploited in domestic equity funds. Greene and Hodges (2001) and Goetzmann, Ivković, and Rouwenhorst (2000) find a similar result in international equity funds.

This article contributes to this literature by showing that pooling shareholders with different liquidity needs in one fund generates a classic externality wherein the high-duration investors subsidize the costly behavior of the low-duration shareholders both in theory and in practice. The externality is problematic because liquidity needs are not priced even though they can be predicted.

The main result of this study is, therefore, that mutual funds are not Diamond-Dybvig-style banks that provide an equitable insurance benefit to all investors. Because the predictable differences in holding periods result in economically significant wealth transfers, the proportionate allocation of costs is structurally inefficient. A central prediction of this research is that funds must either change their pricing mechanism to eliminate the *pooling externalities* or alternative investment vehicles will compete successfully for the high-duration, low-cost investor.

The empirical work presented in this article is based on a new, proprietary database.[7] It is a six-year panel of all transactions within and across all funds in one no-load mutual fund family. A single-spell duration model is applied to accounts, showing that it is possible to predict fund holding periods. The cost of providing liquidity in a mutual fund (*liquidity costs*) is estimated under various simplifying assumptions.[8] The cost of proving liquidity to observably different investors in a pooled fund (*pooling costs*) is also measured.

The remainder of this article is organized as follows. The next section presents the theory model of pooling externalities. Section 3 describes the data. Section 4 introduces the duration model, and sections 5–7 study redemptions in the context of non- and semi-parametric duration models.

[5] Nearly forty percent of investors do not know whether their investments are in load or no-load funds and over seventy percent do not know whether their funds are domestic or international. There is another more knowledgeable group for which these percentages are zero and six, respectively.

[6] The Wyatt Company prepared, in 1990, a study of shareholder persistency in *load* and *12b-1* mutual funds. They sampled approximately 1500 new accounts in 1974 and 800 new accounts in 1984. The objective was to assist the NASD in formulating rules governing 12b-1 plan charges, so the analysis focused on the differences between 12b-1 and load shareholders. It is no surprise that the Wyatt shareholders, some of whom paid 8.5% loads, are more persistent than the shareholders in this study. This article's data covers an entire mutual fund family, considers more recent shareholders, and contains only no-load funds. The evidence presented in sections 3 and 5 by itself is a significant contribution to both professional and academic knowledge.

[7] Johnson (2001) contains summary statistics of the shareholders in this database. Demographic data, account size, trading patterns, portfolios held, and other summary statistics are discussed at length.

[8] Edelen (1999) measures the fund's aggregate cost of liquidity provision. He uses monthly fund flows and semi-annual fund holdings in a sample of 166 funds to estimate that one dollar of liquidity-motivated trading costs the fund 1.7¢.

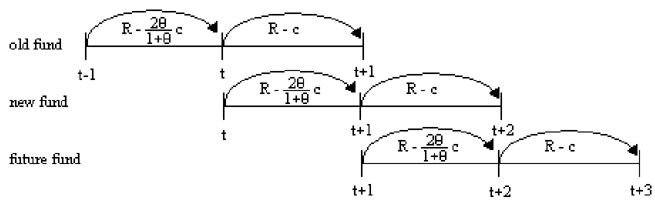

Figure 1: **Investment Choices at Time t and Fund Returns.** In an OLG model where both funds and investors live for two periods, investors choose between two funds each period. Trading costs c imposed by shareholder redemptions reduce the otherwise fixed return R of the funds. In the unique separating equilibrium, investors with high switching costs are long-term investors (proportion $1-\theta$) while those with low switching costs are short-term (θ). The first-period fund return $R - \frac{2\theta}{1+\theta}c$ is greater then the second-period fund return $R - c$. The traders' return $\left(R - \frac{2\theta}{1+\theta}c\right)^2$ is, therefore, greater than the non-traders' return $\left(R - \frac{2\theta}{1+\theta}c\right)(R-c)$.

Section 8 estimates mutual fund liquidity and pooling costs. The analysis concludes with a discussion of why mutual funds choose to pool predictably different shareholders and a summary of the main results in sections 9–10.

2 Model

This section presents a simple model of mutual fund investing where the motivation for trade is the mutual fund pricing mechanism. Because some investors have high switching costs and remain in the fund for both periods, they pay a disproportionate amount of the fund's aggregate costs and receive a lifetime return lower than the return obtained by the investors who trade. Theorem 1 documents the economy's unique separating equilibrium. Theorem 2 demonstrates that there is no pooling equilibrium. Theorem 3 proves that trading is socially wasteful. Theorem 4 shows that if the proportion of traders is high enough in the separating equilibrium, then prohibiting trade is Pareto optimal.

The model is based on the standard infinite-horizon overlapping generations (OLG) framework, where time is divided into a countable number of dates $t \in \mathcal{Z}$, the set of integers. At each t, a new generation of investors, normalized to one, is born and one new mutual fund is created. Both investors and funds live for two periods. Investors are endowed with unit wealth at *birth* (t), may switch funds at *midlife* ($t+1$), and consume only at *death* ($t+2$). Funds are endowed with a constant returns to scale investment technology that enables them to generate raw returns $R > 1$ each period. Without loss of generality, assume funds are owned by investors and charge management fee $f = 0$ each period.

When a *new investor* is born at t, he can choose to invest in either the second phase of the *old fund* born at $t-1$ or the first phase of the *new fund* born at t. If he chooses the old fund, he is forced to redeem at $t+1$ when it is liquidated. He must invest the proceeds either in the new fund's second phase or the first phase of the *future fund* born at $t+1$. On the other hand, if the investor chooses the new fund at t, he can either continue in the fund at $t+1$ or he can invest in the first phase of the future fund. There is no alternative storage technology so he must invest in a fund each period. See Figure 1.

Investors pay a dissipative *switching fee* $s \in \{s_l, s_h\}$ to exchange funds at midlife.[9] Assume

[9]Investor heterogeneity is modeled in terms of switching costs. Other devices that generate similar results include

proportion $\theta \in (0,1)$ of each generation have low switching costs, $s_l = 0$, and $1-\theta$ have high switching costs, $s_h > 0$. The fund pays a *trading cost* $c \in (0, \frac{1}{2}R)$ in proportion to realized redemptions at the end of its first and second phases. Shareholder flows reduce, therefore, the fund's after-cost performance for both redeemers and non-redeemers.

Theorem 1 shows that if switching costs are large enough for the high-cost type, then there is a separating equilibrium where the low-switching-cost (LSC) investors switch funds at midlife and the high-switching-cost (HSC) investors do not switch funds. The HSC investors recognize the LSC investors earn a higher return than they do in this full-information game, but they refrain from trade because their switching cost is larger than the trading gain.

Theorem 1 (Separating Equilibrium) *If $\theta \in (0,1)$ and $s_h \geq \frac{1-\theta}{1+\theta}c$, then investors with high switching costs invest for both periods in the new fund while investors with low switching costs invest in the new fund for one period but switch to the future fund at midlife. The first-period fund return is, therefore, $R - \frac{2\theta}{1+\theta}c$ and the second-period fund return is $R-c$. The traders' return $\left(R - \frac{2\theta}{1+\theta}c\right)^2$ is greater than the non-traders' return $\left(R - \frac{2\theta}{1+\theta}c\right)(R-c)$.*

Proof. At time t, assume the LSC investors find it optimal to switch while the HSC investors do not. In this case, the new fund acquires the entire new generation of investors (1) and all shareholders from the old fund who have low switching costs (θ). The new fund has, therefore, $1+\theta$ investors at t, 2θ of whom redeem at $t+1$ when the older θ die and the younger θ switch to the future fund. The remaining $1-\theta$ HSC investors continue in the fund until it is liquidated at $t+2$. The total trading cost at $t+1$ is $2\theta c$, and the per-investor cost is $\frac{2\theta}{1+\theta}c$. The first-period fund return is, therefore, $R - \frac{2\theta}{1+\theta}c$. At $t+2$ the total trading cost is $(1-\theta)c$, and the per-investor cost is c. The second-period fund return is, therefore, $R-c$. Given these fund returns and investing strategies, do any investors deviate? For the LSC investor, the equilibrium return is $\left(R - \frac{2\theta}{1+\theta}c\right)^2$ while the best deviating return is $(R - \frac{2\theta}{1+\theta}c)(R-c)$. Thus, they do not deviate if $\theta \in (0,1)$. For the HSC investor, the equilibrium return is $\left(R - \frac{2\theta}{1+\theta}c\right)(R-c)$ while the best deviating return is $\left(R - \frac{2\theta}{1+\theta}c - s_h\right)\left(R - \frac{2\theta}{1+\theta}c\right)$. Thus, they do not deviate if $s_h \geq \frac{1-\theta}{1+\theta}c$. ■

Theorem 2 demonstrates there is no pooling equilibrium.

Theorem 2 (Pooling Equilibrium) *There is no pooling equilibrium.*

Proof. Assume no investor switches at midlife. The return from not deviating is $R(R-c)$ while the return from deviating is $(R-s)R$. In this case, at least the LSC investors deviate because $s_l = 0$. Assume all investors switch at midlife. The return from not deviating is $(R-c-s)(R-c)$ while the return from deviating is $(R-c)(R-c)$. Thus, the HSC investors deviate because $s_h > 0$. ■

The before-trade fund return is always R. Because trades generate no social benefits and are costly, they are a deadweight loss to the economy. Theorem 3 proves this, showing that a social planner would prohibit midlife trades.

Theorem 3 (Social Welfare) *Midlife trades are a deadweight loss to the economy in the separating equilibrium.*

Proof. Assume the separating equilibrium where trades generate gains for the LSC investors and losses for the HSC investor. The aggregate benefit and cost of the midlife trades are

$$\text{aggregate benefit} \quad = \quad \theta\left(\left(R - \frac{2c\theta}{1+\theta}\right)^2 - R\,(R-c)\right)$$

search costs (some investors are better or more cheaply informed about alternative funds), trading preferences (some agents *like* to trade), and taxes (some investors are not required to pay taxes on early redemptions).

$$\text{aggregate cost} = (1-\theta)\left(R\,(R-c)-(R-c)\left(R-\frac{2c\theta}{1+\theta}\right)\right).$$

The difference between the benefits and costs is

$$\begin{aligned}\text{aggregate benefit - aggregate cost} &= \frac{c\theta\left(2c\left(1+\theta^2\right)-R(1+\theta)^2\right)}{(1+\theta)^2}\\ &< 0,\end{aligned}$$

since $\theta \in (0,1)$ and $c < \frac{1}{2}R$. ■

Were funds to price redemptions according to marginal cost, each investor would pay c for each redemption. However, the actual pricing structure in the model permits a wealth transfer in the separating equilibrium: twice redeemers pay $\frac{2\theta}{1+\theta}c + \frac{2\theta}{1+\theta}c < 2c$ while once redeemers pay $\frac{2\theta}{1+\theta}c + c > c$. Theorem 4 shows that if θ is large enough, then even the traders who exploit the pricing mechanism would prefer marginal cost pricing to the prisoner's-dilemma-like payoffs they receive in the separating equilibrium. The essential intuition is that with enough traders in the economy, their lifetime return $\left(R-\frac{2\theta}{1+\theta}c\right)^2 \approx (R-c)^2$ is less than the return they receive if midlife trades are prohibited $R(R-c)$.

Theorem 4 (Trading Tax) *Assume $\theta \in (0,1)$ and $c \in (0, \frac{1}{2}R)$. Define $V = \frac{R-2\sqrt{R(R-c)}}{4c-3R}$. If $\theta \in [V,1)$, then a midlife trading tax $p \geq c$ is Pareto optimal. If $\theta \in (0,V)$, then a midlife trading tax $p \geq c$ benefits the HSC investors but hurts the LSC investors.*

Proof. If no one trades then all investors receive the two-period return $R(R-c)$. Does anyone deviate? The LSC investors' return from deviating is $(R-p)R$, so they do not deviate. The HSC investors similarly do not deviate. Is this new, pooling equilibrium Pareto improving? The HSC investors' welfare clearly improves. The LSC investors are no worse off if $R(R-c) \geq (R-\frac{2\theta}{1+\theta}c)^2$, or $f(\theta) = \left(R(R-c)-(R-\frac{2\theta}{1+\theta}c)^2\right)(1+\theta)^2 \geq 0$ Note that $f(\theta)$ is a convex, parabolic function of θ. Its two roots are $V' = \frac{R+2\sqrt{R(R-c)}}{4c-3R}$ and $V = \frac{R-2\sqrt{R(R-c)}}{4c-3R}$, where $V' < V$. The first root, V', is negative because $c \in (0, \frac{1}{2}R)$. The second root, V, is between zero and one for the same reason. Therefore, $f(\theta)$ is negative when $\theta \in (0,V)$ and non-negative when $\theta \in [V,1)$. ■

The model has two features. First, low switching costs cause investors to redeem early. Any secular trend that reduces costs for HSC investors will be associated with shorter holding periods. The last twenty or thirty years have seen developments in the industry that have profoundly reduced switching costs. From a technological perspective, the introduction of telephone, Internet, and supermarket-style trading substantially decreased the costs of trading funds and simultaneously increased the number of funds available to investors (i.e., decreased *search costs*). From an educational perspective, the explosion of media and research firms covering the mutual fund marketplace has made searching for and switching to appropriate funds easier. Lastly, the shift in the retirement market towards defined contribution plans and IRA's has increased investors' incentives to monitor their investments which are increasingly being placed in mutual funds.

Are individual holding periods declining? Annualized redemptions in equity mutual funds increased from under 30% in 1991 to over 40% in 2001.[10] Turnover in the NYSE increased from 48% in 1991 to 78% in 1999.[11] Although it is hard, perhaps impossible, to infer individual behavior from these aggregate statistics, it is clear that at least some investors are trading more than before. See Figure 2 Panel A for monthly mutual fund redemption rates from January 1991 to May 2001 and Panel B for annual turnover in NYSE stocks from 1991 to 1999. Anecdotally, the Vanguard Group

[10] Investment Company Institute.

[11] *NYSE Fact book, 2000.*

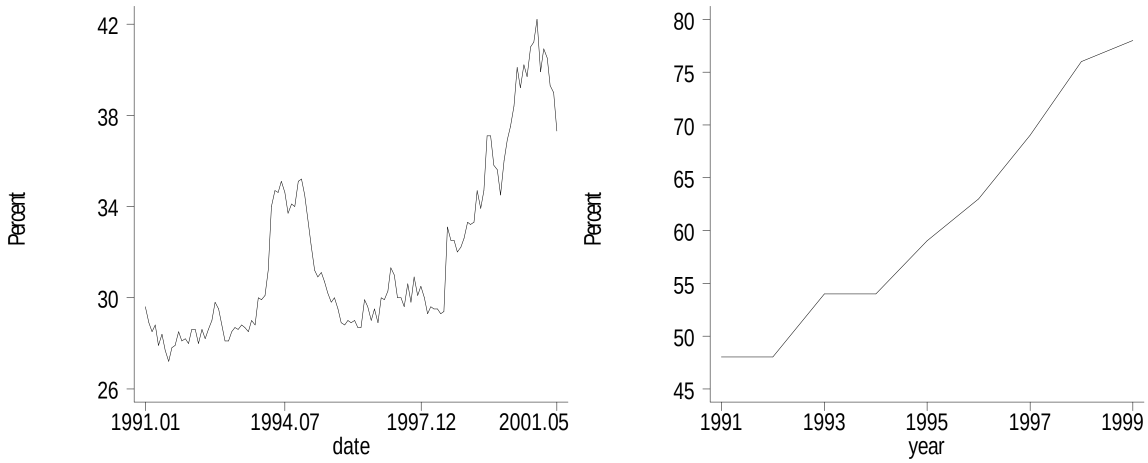

Figure 2: **Equity Mutual Fund Redemption Rate and NYSE Annual Turnover.** Panel A plots monthly mutual fund redemption rates from January 1991 to May 2001. Panel B plots annual turnover in NYSE stocks from 1991 to 1999.

reports that "Twenty years ago, the average holding period was seven, eight, even 10 years. Now three years is considered a long holding period."[12]

A second feature of the model is that the frequent traders are gaming the fund's pricing structure. Forcing them to internalize their costs would eliminate their behavior. Mutual funds are responding to the increased trading by adding redemption fees to both new and existing funds with the intent either to discourage short-term traders or to compensate the fund for their expenses.[13] Between June 1999 and July 2001, the percent of no-load equity funds that had redemption fees increased from 3.5% to 7.7%.[14] Moreover, the average fee increased from 112 bp to 119 bp, and the average horizon jumped from 7.0 months to 9.5 months.

Are redemption fees curbing early withdrawals? Chordia's (1996) empirical results show that funds with redemption fees hold less cash than funds without redemption fees. He interprets this finding to mean that redemption fees dissuade redemptions.

3 Data

The data for this research was supplied, generously, by a mutual fund family that is typical in all material respects. The family is an open-end, no-load complex with common fees and trading restrictions. It consists of approximately ten funds, including both equity and fixed-income funds (FIF's). The equity funds tend to hold neither large-cap nor international securities. Fund shares have been distributed geographically—in terms of both number and value of accounts—in a way that closely mirrors the distribution of wealth in the U.S., with the exception of a disproportionately large

[12]The Washington Post, October 10, 1999, p H01.

[13]Redemption fees date back to at least 1984 when the newly-organized Vanguard Health Care fund imposed a 1% fee on redeemed shares that had been held for less than one year. In careful language, a "redemption fee" refers to a fee on share redemptions that is paid back to the mutual fund, as opposed to fees that are paid to the investment advisor, broker, or other intermediary. In context of the model, Theorem 4 shows that a sufficiently large redemption fee would eliminate all socially wasteful trading.

[14]July 1999 and August 2001 Morningstar PrincipiaPro Plus; author's calculations.

presence in the management company's home state. The fund family wishes to remain anonymous to protect both the funds' shareholders and the investment advisor(s).

Public data on mutual fund shareholders is generally unavailable. Thus, it is difficult to say whether shareholders in this family of funds are representative of the larger population of shareholders. Johnson (2001) does not find significant differences between shareholders in this family and the "average" shareholder profiled in various Investment Company Institute (ICI) publications.[15] The transfer agent reports, however, that this family has a "more stable" shareholder base than the other families it serves.[16]

The fund family's transfer agent provided an electronic copy of its database. The data begin fall 1994 and end summer 2000. It contains three main files: shareholders (over fifty-thousand accounts), transactions (over three-quarters of a million records), and funds (around ten). The files are summarized as follows:

Shareholders Includes investors who held shares (electronically or on book) anytime during this window, whether or not transactions occurred. Accounts closed before fall 1994 are not available, so all pre-fall-1994 data come from the long-lived accounts. Demographic information includes birth date, zip code, and a field indicating affiliation with the management company, if any. Comprehensive registration information and account options are also included.

Transactions Includes for each account all transactions (such as purchases and exchanges), method of payment (such as wire, check, or exchange), and various codings for internal use. Transactions from before fall 1994 are unavailable.

Funds Includes a complete history of net asset values (NAV's) and distributions from the funds' inception.

An *account* is the uniquely registered holdings of an investor in one fund, including its complete transactional history. The account's *registration* specifies the account ownership and tax status. Most registration changes require the closure of the old account and the opening of a new account.[17] Accounts are also closed when the entire balance is *exchanged* to another fund in the family or when the entire investment is *redeemed* from the fund family.

The unit of observation throughout most of the analyses is the account and not, for example, the investor or the dollar investment. Thus the fact that particular investors might hold multiple funds, might hold the same fund multiple times, and might repeatedly reregister the investment is ignored. This biases the survival estimates downward. In practice few investors seem to have multiple registrations, though there is a handful of accounts that has been reregistered ten or more times.[18]

All FIF accounts and all accounts opened before fall 1994 are dropped from the analysis. The remaining accounts are homogeneous (only equity funds) and are complete (every transaction in every account is observed). Therefore, the database includes every investor that opened an account between fall 1994 and summer 2000 and every transaction made by those shareholders. Static account options (for example, automatic investment/withdrawal plan (AIP/AWP) participation) are effective as of the account closing date or summer 2000, though much can be inferred dynamically from the transactions file (for example, individual AIP transactions are dynamically coded as such). Some of the accounts are *censored* because they have not yet closed.

[15] The ICI is the fund industry's trade group.

[16] The transfer agent, like others, serves multiple mutual fund families.

[17] A marriage might result in an individual account being reregistered as a joint account. This would result in two observations in the data set. No attempt is made to analyze multiple fund ownership in this paper.

[18] Removing accounts that fail within one, six, or twelve months does not qualitatively change any result of interest in this study.

3.1 Fund Distribution

The industry assigns *social codes* to accounts. Examples of the scores of codes used by this family are as follows: joint tenants with rights of survivorship (JTWROS), partnership, corporate fiduciary, omnibus house, ROTH conversion, UGMA, dealer, and IRA SEP. Some of these codes exist as check-off boxes on the application; others are assigned by the transfer agent after reviewing the application and all supporting material. Protocols for assigning codes are encapsulated in a written document that is used by employees of the transfer agent making the assignment. Rigorous quality-control checks further ensure the consistency of these data across time and across employees.

All accounts are grouped into eight *distribution channels* based on their social codes: retail taxable, retail tax-free, minor, trusts, Fund/SERV, entity, omnibus house, and other. The *retail taxable* accounts are individual accounts or joint accounts. The *retail tax-free* accounts are generally IRA's, either traditional or Roth. The *minor* accounts are registered as UGMA/UTMA and are invariably managed by the minor's parent. The *trust* accounts tend to be held directly (the trustee is a family member), though many go through intermediaries such as banks. *Fund/SERV* accounts originate through the Fund/SERV system and generally lack identifying characteristics of the underlying shareholder.[19] *Entity* accounts include religious organizations, clubs, associations, educational institutions, corporations, and partnerships. Supermarkets and omnibus house accounts make up the *supermarket* category. Lastly, *other* accounts have either undefined or omitted social codes. Sections 6–7 show that the eight distribution channels are important predictors of account duration.

Within the eight distribution channels, funds and investors are connected in one of two ways, either directly or indirectly. The *direct* shareholder interacts directly with the transfer agent and completes the fund's application.[20] These shareholders are assigned the following social codes: retail taxable, retail tax-free, minor, trusts, and entity. The *indirect* shareholder trades through an intermediary who in turn trades with the transfer agent.[21] These shareholders are assigned the following social codes: supermarket, Fund/SERV, and other. Section 8 shows that liquidity costs vary across the two *account types*.

One of this article's contributions to the financial economics literature is the observation that account duration varies across distribution channels. There are two competing stories that are consistent with this fact. The first interpretation is that trading preferences are endogenous. If investors are randomly assigned to the distribution channel, then positive price effects imply that investors in the low-cost channel trade more than investors in the high-cost channel. The second interpretation is that trading preferences are exogenous. If investors know their *ex ante* level of desired trading, then those investors desiring high levels of trading will be disproportionately attracted to the low-cost channels. In either case, however, duration is predictable by the fund because it observes the channel through which the shareholder purchased fund shares.

[19] The Fund/SERV system is a technology provided by the National Securities Clearing Corporation (NSCC) that includes 571 fund families in September 2001. Broker/dealers and other professionals connected to this network are able to trade funds in a supermarket-like fashion at low financial and economic cost.

[20] Strictly speaking, the *distributor* sells shares to new shareholders. The transfer agent processes orders only from existing shareholders.

[21] The intermediary who brings investors to the fund can establish one of three bookkeeping arrangements with the transfer agent. First, an omnibus account can be created that contains all shareholders of the intermediary in the fund. In this case, the transfer agent has on its books one (large) account that is permitted to trade frequently in the fund. Large mutual fund supermarkets such as Charles Schwab or TD Waterhouse may choose this system. In the second arrangement, intermediaries open accounts for each of their individual investors using the intermediaries' name and a control number. Brokers such as e*trade and CSFB*direct* may choose this system. Under either of these schemes, the transfer agent does not know the identity of the end investor, though in the second case it is able to observe the investor's trades. In the final bookkeeping arrangement, each account is identified by individual shareholder.

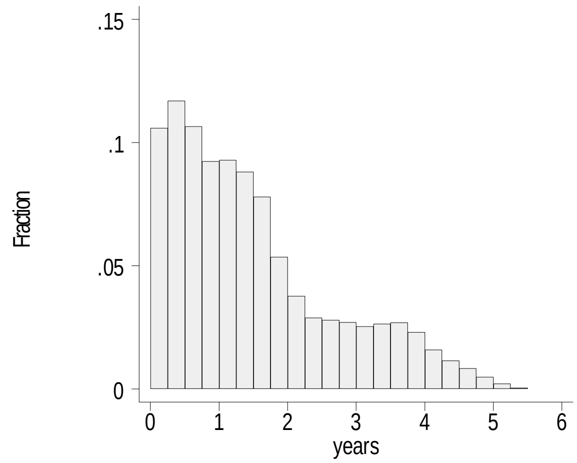

Figure 3: **Account Duration.** Presents a histogram of account duration, in years, for non-censored accounts. Long-lived accounts are disproportionately censored in the data window are therefore underrepresented in the histogram.

3.2 Summary Statistics

The remainder of this section summarizes the data, emphasizing distribution channels and account types. Table 1 provides a list of the variables that will be used in section 6 and their definitions. Table 2 provides summary statistics for these variables. These variables are not time-varying and are measured when the account was opened.[22] Variables that are time-varying will be introduced, defined, and summarized in section 7.

Figure 3 presents a histogram of account duration for closed accounts. Long-lived accounts are disproportionately censored in the six-year window and are, therefore, underrepresented in the histogram. Ignoring this fact, the histogram shows that most accounts fail early and that the percent of closures is somewhat constant between two and four years.

Unreported scatterplots of account duration by opening dates and closing dates are revealing.[23] First, they show that accounts of all durations existed at each opening and closing date. Second, it is clear that there is more clustering based on opening dates than closing dates. One interpretation of this observation is that investors buy performance but sell randomly. This interpretation is also consistent with the convex performance-flow relationship documented by Gruber (1996) and others.

A pairwise correlation matrix for some variables of interest by distribution channel is presented in Table 3. The coefficients are double starred if they are significant at the 1% level. Among other relationships, the table shows that the retail taxable accounts are correlated with automatic investment plans ($\rho = .212$), the no-check flag ($\rho = -.339$), and telephone redemption-only or redemption-and-exchange options ($\rho = .215$ and $\rho = 0.308$, respectively).

The last statistic presented is a percentage breakdown of daily dollar flows (inflows, outflows, and net) by account type (direct, indirect, and else). For example, the contribution of direct net

[22]It will be repeatedly emphasized throughout the discussion that the unit of observation is the account. It is not the investor and it is not the dollar investment. Also, it will be emphasized that all static variables are measured on the date the account was opened.

[23]Confidentiality concerns prevents their inclusion in this article.

Shareholder's age dummies	
dAGE00	true if investor's age is between 00 and 10 at account opening
dAGE10	true if investor's age is between 10 and 20 at account opening
dAGE20	true if investor's age is between 20 and 30 at account opening
dAGE30	true if investor's age is between 30 and 40 at account opening
dAGE40	true if investor's age is between 40 and 50 at account opening
dAGE50	true if investor's age is between 50 and 60 at account opening
dAGE60	true if investor's age is between 60 and 70 at account opening
dAGE70	true if investor's age is between 70 and 80 at account opening
dAGE80	true if investor's age is greater than 80 at account opening
dNOAGE	true if investor's age is is not available
Size of initial transaction dummies	
dSIZE0	true if $0.0 <$ opening day's account value ≤ 0.5 times the fund's minimum
dSIZE0.5	true if $0.5 <$ opening day's account value ≤ 1.0 times the fund's minimum
dSIZE1	true if $1.0 <$ opening day's account value ≤ 2.5 times the fund's minimum
dSIZE2.5	true if $2.5 <$ opening day's account value ≤ 5.0 times the fund's minimum
dSIZE5	true if $5.0 <$ opening day's account value ≤ 50.0 times the fund's minimum
dSIZE50	true if opening day's account value > 50.0 times the fund's minimum
Fund employee dummy	
dEMPLOYEE	true if shareholder is an employee or relative of an employee of the fund
Shareholder's state of residence dummies	
dHQ	true if account is registered in the investment advisor's home state
dSTATE1 –dSTATE6	true if account is registered in one of six large states
Account option dummies	
dAIPAWP	true if there was an automatic transaction during the first 90 days
dNOTcheck	true if the account was opened with a check (personal or certified)
dCASHdiv	true if dividends are taken out (by any means)
dTELEnone	true if neither telephone redemptions nor exchanges are allowed
dTELExchng	true if only telephone exchanges are allowed
dTELErdmptn	true if only telephone redemptions are allowed
dTELEboth	true if both telephone exchanges and redemptions are allowed
Distribution channel dummies	
dRETAILtaxable	true if social code is retail taxable
dRETAILtaxfree	true if social code is retail tax-free
dMINOR	true if social code is minor (UGMA/UTMA)
dFundSERV	true if social code is Fund/SERV
dTRUST	true if social code is indirect
dENTITY	true if social code is entity
dSUPERMRKT	true if social code is ominibus house
dOTHER	true if social code is undefined or omitted

Table 1: **Definitions of Time-constant Variables.** Defines terms used in section 6's regressions.

Covariate	Mean	Std. Dev.	Covariate	Mean	Std. Dev
censored	0.322	0.467	dAIPAWP	0.120	0.325
			dCASHdiv	0.043	0.203
dRETAILtaxable	0.579	0.494	dNOTcheck	0.215	0.411
dRETAILtaxfree	0.172	0.377			
dMINOR	0.041	0.199	dTELEnone	0.329	0.470
dTRUST	0.107	0.309	dTELExchng	0.160	0.366
dFundSERV	0.086	0.280	dTELErdmptn	0.121	0.326
dENTITY	0.004	0.065	dTELEboth	0.390	0.488
dSUPERMRKT	0.000	0.019			
dOTHER	0.011	0.106	dAGE00	0.008	0.091
			dAGE10	0.010	0.097
dEMPLOYEE	0.002	0.040	dAGE20	0.075	0.263
			dAGE30	0.179	0.383
dHQ	0.071	0.257	dAGE40	0.197	0.398
dSTATE1	0.137	0.344	dAGE50	0.157	0.364
dSTATE2	0.079	0.269	dAGE60	0.086	0.280
dSTATE3	0.055	0.227	dAGE70	0.039	0.194
dSTATE4	0.044	0.205	dAGE80	0.007	0.084
dSTATE5	0.046	0.210	dNOAGE	0.242	0.429
dSTATE6	0.044	0.205			

Table 2: **Summary Statistics of Time-constant Variables.** Presents the means and standard deviations for the static database. All variables are binary. The unit of observation is the account. There are more than fifty-thousand records. The censored accounts were active on the date the data set was extracted in summer 2000. Accounts opened before fall 1994 are excluded as are all accounts in fixed-income funds.

	dRETAIL taxable	dRETAIL taxfree	dMINOR	dTRUST	dFund SERV	dENTITY	dSUPER market	dOTHER
dSIZE0	0.038**	0.068**	0.047**	-0.081**	-0.087**	-0.012**	-0.006	-0.036**
dSIZE0.5	0.140**	-0.014**	0.016**	-0.080**	-0.130**	-0.018**	-0.011	-0.040**
dSIZE1	0.046**	-0.051**	0.005	0.021**	-0.033**	-0.004	-0.003	-0.014**
dSIZE2.5	-0.089**	0.007	-0.024**	0.055**	0.098**	0.001	-0.001	0.019**
dSIZE5	-0.196**	0.011	-0.055**	0.113**	0.208**	0.034**	0.012**	0.077**
dSIZE50	-0.086**	-0.011**	-0.018**	0.035**	0.095**	0.034**	0.058**	0.089**
dAIPAWP	0.212**	-0.135**	0.032**	-0.079**	-0.113**	-0.009	-0.007	-0.028**
dNOTcheck	-0.339**	-0.087**	-0.064**	0.125**	0.573**	0.026**	0.034**	0.111**
dCASHdiv	-0.134**	-0.096**	-0.038**	0.000	0.372**	0.001	0.026**	0.051**
dTELEnone	-0.197**	-0.060**	-0.005	-0.014**	0.436**	0.020**	0.008	0.017**
dTELExchng	-0.348**	0.643**	-0.064**	-0.053**	-0.133**	-0.013**	-0.008	-0.030**
dTELErdmptn	0.215**	-0.167**	0.012**	-0.029**	-0.113**	-0.023**	-0.007	-0.033**
dTELEboth	0.308**	-0.314**	0.045**	0.073**	-0.245**	0.006	0.003	0.028**
dMONDAY	0.039**	0.024**	0.011	-0.006	-0.098**	0.000	-0.003	-0.011
dTUESDAY	0.021**	-0.014**	0.005	0.001	-0.021**	0.005	-0.001	-0.006
dWEDNESDAY	-0.049**	-0.005	-0.007	0.005	0.085**	-0.003	0.008	0.022**
dTHURSDAY	0.011	-0.007	0.004	-0.004	-0.011	0.004	0.002	0.001
dFRIDAY	-0.038**	-0.005	-0.017**	0.007	0.079**	-0.006	-0.004	0.000

Table 3: **Correlation Matrix.** This table presents pairwise correlation coefficients for some variables of interest by distribution channel. Coefficients that are significant at the one-percent level are double starred.

Flows	Account Type Direct	Indirect	Else
Inflows	26%	71%	4%
Outflows	25	72	3
Net	28	64	7

Table 4: **Fund Flows by Account Type.** Presents the contribution to average daily fund flows (in dollars) by account type. Indirect investors purchased through an intermediary; direct investors did not.

flows to total net flows is calculated as follows:

$$\text{percent direct net flow} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T_i} \text{direct net flow}_{i,t}}{\sum_{i=1}^{N}\sum_{t=1}^{T_i} \text{total net flow}_{i,t}},$$

where N is the number of funds in the sample and T_i is the number of days observed in fund i. This statistic is tabulated in Table 4 for daily flows. Net flows consist of direct flows (28%), indirect flows (64%), and residual flows (7%). Each categories' share of gross inflows about match their share of gross outflows. The indirect flows are larger than the direct flows.[24] Overall, total assets under management increased during the sample period for two reasons. First, inflows were 17% greater than outflows. Second, the funds' returns were positive.

The data are thoroughly described in Johnson (2001). His analysis, taken together, suggests that the duration model is an appropriate framework for studying liquidity needs. In particular, he documents that the vast majority of accounts contain no trading activity between account opening and account closing (observing account opening and closing is equivalent to observing all account activity) and that most investors do not hold multiple funds within the family (observing "account" is equivalent to observing "investor").

4 Mathematics of Duration Analysis

This section reviews well-known facts about single-spell duration models.[25] The discussion is limited to concepts that will be used in the empirical analysis, and all terminology is introduced in terms of mutual funds. The time origin is the date the initial transaction is processed, and the account *fails* when it is liquidated. Between these dates, the account is *at risk*.

Cox's proportional hazards model is chosen for the main analyses and is estimated using two types of covariates. The first group is from data collected the day an account was opened; these data are *static* insofar as they do not change over the life of the account. The other data are generated after account opening; these data are *dynamic* insofar as they can change throughout the account's lifetime. The main result will be that the static covariates are economically and statistically significant in predicting the conditional probability of account failure. This result is robust to the inclusion of dynamic state variables.

[24] Unreported results show that the indirect flows are dominated by the supermarket flows. This is consistent with the belief that supermarkets decrease search/switch costs.

[25] Kiefer (1988) provides a broad introduction to the economic duration literature.

4.1 Definitions

Define a continuous random variable T^* to be the length of time a mutual fund account is open if its complete duration is observed; otherwise, let c be its observed length. Define $T = \min\{T^*, c\}$ to be the observed time the account is open. The data set was drawn on a specific date during summer 2000. Because this date is independent of all events, it is called *random censoring*. This generally simplifies the analysis. The distinction between T and T^* becomes important in the estimation part of the discussion; assume for now no accounts are censored.

There are two different concepts for the distribution of T. The first is unconditional: what is the probability that a newly opened account is closed shortly after its one-year anniversary? This is distinct from the conditional concept: what is the probability that an account that survives to its one-year anniversary fails shortly thereafter? Insofar as one distribution can be derived from the other, the mathematics are identical in either case. However, in the context of mutual fund account redemption, the later is more conducive to economic modeling.

Let $F(t) = \Pr(T < t)$ be the cumulative distribution function (CDF) of duration T, where the support is $(0, \infty)$. The corresponding probability density function (pdf) is $f(t) = F'(t)$. The *survival distribution function* is $S(t) = 1 - F(t) = \Pr(T \geq t)$. Then $\lambda(t) = f(t)/S(t)$ is the *hazard function* and $\Lambda(t) = \int_0^t \lambda(u)\, du$ is the *integrated hazard*. A simple calculation shows that

$$S(t) = \exp\{-\Lambda(t)\}. \tag{1}$$

The hazard function represents, loosely speaking, the conditional failure rate of accounts. In other words, if an account exists at t, then the product $\lambda(t)\Delta(t)$ is the approximate conditional probability that the account will fail during the short interval $[t, t + \Delta t)$.[26] A straightforward calculation provides some justification for this interpretation:

$$\lambda(t) \quad = \quad \lim_{h \to 0^+} \left(\frac{\Pr[t \leq T < t + h | T \geq t]}{h} \right).$$

Because $f(t) = \lim_{h \to 0^+} \left(\frac{\Pr[t \leq T < t+h]}{h} \right)$, some authors have called $\lambda(t)$ a *conditional density*, but this is not technically correct since the hazard does not necessarily integrate to one over the positive real line and it can exceed one. Of interest in this study is whether the hazard function increases or decreases in t. *Negative duration dependence* occurs if $d\lambda(t)/dt < 0$, and *positive duration dependence* occurs if $d\lambda(t)/dt > 0$.

This analysis assumes that every duration is a realization from the same probability distribution. If a column vector z of time-invariant *covariates* is thought to affect the distribution of the response variable through some function $\phi(z; \beta)$, then heterogeneity can be controlled for by explicitly modeling that interaction. The economics literature generally models the interaction between z and T or the interaction between z and λ. The former interaction generates the *accelerated failure time model* while the later generates the *proportional hazards model*.

4.2 Life-table Estimation

Estimation methods in the absence of an underlying model are well known. The *product-limit* method of non-parametric analysis assumes a continuous time scale and hence distinct failure times. The data set has thousands of non-unique failure times and is therefore better suited to the *life-table* method. The term *life table* comes from the actuarial sciences.

[26]Taking time to be discrete days and speaking loosely, $\lambda(51)$ is the probability that an account that survives past its 50th day is closed on its 51st day.

A strictly increasing sequence of real numbers $\{t_i\}_{i=0}^{k+1}$ is chosen that partitions the non-negative real line into $k+1$ disjoint intervals, where $t_0 = 0$ and $t_{k+1} = \infty$. Define n_i to be the number of accounts that enter $[t_{i-1}, t_i)$, q_i to be the number of closures, and c_i to be the number of censures. Assuming that the censoring in interval i occurs at its midpoint t_{m_i}, and the *effective sample size* is $n'_i = n_i - c_i/2$.

The *conditional probability of failure* in the i^{th} interval is estimated by $\hat{d}_i = q_i/n'_i$ and the survival distribution function is estimated by

$$\hat{S}(t_i) = \begin{cases} 1, & \text{if } i = 0 \\ \hat{S}(t_{i-1})[1 - \hat{d}_{i-1}], & \text{if } i > 0. \end{cases}$$

The hazard function is estimated by

$$\hat{h}(t_{m_i}) = \frac{2\hat{d}_i}{(t_i - t_{i-1})(2 - \hat{d}_i)}.$$

These non-parametric estimates can be used to get an idea of the behavior of the underlying redemption process, ignoring the influence of concomitant variables.

4.3 Proportional Hazards Model

A duration model has never before been fit to the holding time of liquid financial assets by retail investors. In the context of mutual funds, there is little economic theory to suggest what the underlying redemption process is, or even if accounts have positive or negative duration dependence. For this reason the accelerated failure time model is discarded in favor of the proportional hazards specification.

Covariates affect the *proportional hazards model* in a straightforward way by rescaling the hazard function. This is modeled by assuming that the hazard can be factored in a special way that separates time t from the covariates z:

$$\lambda(t, z; \beta, \lambda_0) = \phi(z; \beta)\lambda_0(t),$$

where the *baseline hazard* is λ_0.[27] With the distribution of both ϕ and λ_0 specified, the proportional hazards model is fully parametric. If just the distribution of ϕ is specified, the model is semi-parametric. The latter approach is followed in this paper.

The *baseline survival distribution function* is

$$S_0(t) = e^{-\int_0^t \lambda_0(u)\, du}. \tag{2}$$

Combining equations (1) and (2), the survival distribution function can be written in terms of the baseline survival distribution function and the function ϕ

$$S(t, z; \beta) = (S_0[t])^{\phi(z;\beta)}.$$

Covariates z affect the survival distribution function's rate of decline. Given distinct values of ϕ, survival distribution functions do not intersect for $t > 0$.

[27] Lancaster (1990) noted that even though econometricians apparently prefer the proportional hazards class, "no econometrician...has even given an economic-theoretic justification of why hazards should be proportional, or even approximately so."

The *hazard ratio* of accounts with covariates z_i and z_j is

$$\begin{aligned}\frac{\lambda(t, z_i; \beta, \lambda_0)}{\lambda(t, z_j; \beta, \lambda_0)} &= \frac{\phi(z_i; \beta)\lambda_0(t)}{\phi(z_j; \beta)\lambda_0(t)} \\ &= \frac{\phi(z_i; \beta)}{\phi(z_j; \beta)}.\end{aligned}$$

That this ratio depends on neither time nor the baseline hazard λ_0 is important in both the estimation and interpretation of the Cox model. A graphical interpretation of the time-invariant assumption is that graphs of hazards are multiplicative shifts of the baseline hazard's graph. If the model were to contain time-varying covariates, $\phi(z(t); \beta)$, then strictly speaking it would no longer be proportional hazards model. These models are, nevertheless, frequently called proportional hazards models.

From the definition of the integrated hazard, it is easy to see that $\Lambda(T) = \phi(z; \beta)\Lambda_0(T)$. The function ϕ is often chosen to be $\phi(z; \beta) = \exp(z'\beta)$ because the model can then be given a log-linear interpretation and there are no complicated restrictions on the coefficients β:

$$\begin{aligned}\log \Lambda(T) &= z'\beta + \log \Lambda_0(T) \\ &= z'\beta + \nu.\end{aligned}$$

Lancaster (1990) shows that the integrated hazard has an exponential distribution with parameter $\theta = 1$. Using this fact, a straightforward transformation shows that the pdf of $Y = \log \Lambda_0(T)$ must be $f_Y(y) = \exp\{y - e^y\}$. But this distribution is the well-known Type 1 Extreme Value distribution. Thus $\nu \sim \mathrm{EV}[1, 0]$, $\mathrm{E}[\nu] = -\gamma$, and $\mathrm{Var}[\nu] = \frac{\pi^2}{6}$, where $\gamma \approx 0.5772$ is Euler's constant.

Although there is a clear and intuitive relationship between z and λ in the proportional hazards model, there is no direct relationship between z and T unless the baseline hazard is parametrically specified. This makes the model hard to interpret, for the estimated coefficients are in terms of the integrated hazard: $\frac{\partial \log \Lambda(T)}{\partial z} = \beta$. Nevertheless, $\lambda = \exp(z'\beta)\lambda_0 = \lambda_0$ when $z = \mathbf{0}$. This suggests that if the covariates are carefully measured so that $z = \mathbf{0}$ represents the "average" account, then the hazard ratio of the account with covariates z_i and the average account is

$$\frac{\lambda(t, z_i; \beta)}{\lambda(t, \mathbf{0}; \beta)} = e^{z_i'\beta}.$$

Accounts for which this term is greater than one are *risky*, and accounts for which this term is less than one are *safe*. In later empirical work, the coefficient vector β is of interest only to the extent it allows for the computation of hazard ratios.

4.4 Proportional Hazards Estimation

The account observations are of the form (T_i, δ_i, z_i), $i \in \{1, \ldots, n\}$. $T_i = T^*$ is the failure time if $\delta_i = 1$ and $T_i = c$ is the censoring time if $\delta_i = 0$. The covariates associated with the ith account is the $k \times 1$ vector z_i, and the associated $k \times 1$ parameter vector is β. Let r be the number of failed accounts; thus, $n - r$ accounts are censored. The account *order statistics* are $T_{(1)} \leq T_{(2)} \leq \cdots \leq T_{(r)}$ for the r closed accounts. The associated account *rank statistics* (i) are the labels attached to the order statistics.[28] The *risk set* $R(T_{(i)})$ is the set of accounts at risk just prior to $T_{(i)}$.

There are several standard techniques for estimating parameters β and λ_0. This subsection outlines the partial likelihood approach taken by Cox (1975). The discussion in greatly simplified

[28] If the 93rd account was the second to fail, then (2) refers to the 93rd account and $T_{(2)} = T_{93}$.

by considering only the case where covariates are time constant. Failure and censoring times are assumed jointly distinct (there are no ties because the time scale is continuous), but this assumption is relaxed below.

The argument begins by recalling that in the proportional hazards specification, β is separable from λ_0. Since λ_0 is completely unspecified, it could be defined to be zero between $T_{(i-1)} < T_{(i)}$. Therefore, the fact that no failures were observed during the interval $(T_{(i-1)}, T_{(i)})$ cannot be used in the estimation of β. The partial likelihood constructed below discards this part of the data for this reason.

Given the risk set $R(T_{(i)})$ and knowledge that exactly one account fails at $T_{(i)}$, the conditional probability that account (i) fails is

$$\frac{\Pr[\text{account } (i) \text{ fails at } T_{(i)}]}{\Pr[\text{one failure at } T_{(i)}]},$$

where $i \leq r$. The numerator is simply the hazard for account (i) at $T_{(i)}$. The denominator is the sum of hazards of failure times across all accounts at risk at $T_{(i)}$. Mathematically, this is given by

$$\frac{\lambda(T_{(i)}, z_{(i)}; \beta, \lambda_0)}{\sum\limits_{l \in R(T_{(i)})} \lambda(T_{(i)}, z_l; \beta, \lambda_0)} = \frac{e^{z'_{(i)}\beta}}{\sum\limits_{l \in R(T_{(i)})} e^{z'_l\beta}}.$$

The key insight is that in the proportional hazards specification, λ_0 can be treated as a nuisance parameter and ignored.

Cox's *partial likelihood* is simply the product of these terms across all r failure times:

$$L(\beta) = \prod_{i=1}^{r} \frac{e^{z'_{(i)}\beta}}{\sum\limits_{l \in R(T_{(i)})} e^{z'_l\beta}}.$$

Even though this is not a likelihood in the normal sense since it is not proportional to the probability of observing censored and non-censored events, it has been shown that the usual likelihood estimation techniques apply. See Kalbfleich and Prentice (1980; chapter 5) for further details. Appendix C derives the true likelihood and discusses why it cannot be used.

Extending the model to include d_i ties at $T_{(i)}$ involves an assumption that censoring occurs after all failures. Handling the remaining ties of failed observations is computationally difficult. The essential intuition is that the d_i ties are a consequence of imprecisely recorded failure times (events are recorded on a discrete time scale but the failure process is truly continuous), and that all $d_i!$ permutations of these failures must contribute to the likelihood.

Let d_i be the number of accounts that fail at $T_{(i)}$, r' be the number of distinct failures, and r be the total number of failures as before (so $\sum_{i=1}^{r'} d_i = r$), and define the column vector $s_i = \sum_{j=1}^{d_i} z_{i,j}$, where $z_{i,j}$ is the vector of covariates associated with the jth failed account at $T_{(i)}$. Let $R_{d_i}(T_{(i)})$ be the set of all subsets of the risk set $R(T_{(i)})$, where the subsets are drawn without replacement and have d_i elements. It can be shown that the partial likelihood is then given by

$$L(\beta) = \prod_{i=1}^{r'} \frac{e^{s'_i\beta}}{\sum\limits_{l \in R_{d_i}(T_{(i)})} e^{s'_l\beta}}.$$

From a computational perspective, this is difficult to compute.[29] In later empirical estimation, the popular Breslow (1974) approximation is used. It is

$$L(\beta) = \prod_{i=1}^{r'} \frac{e^{s_i'\beta}}{\left(\sum_{l \in R(T_{(i)})} e^{s_l'\beta}\right)^{d_i}}.$$

The quality of this approximation decreases as the ties-to-risk-set ratio increases.

The second parameter that needs to be estimated is the baseline hazard. The common derivation is similar in form and complexity to the work just presented and will not be reproduced here. Readers interested in the details should refer to Kalbfleich and Prentice (1973).

5 Life-table Estimates

This section presents a non-parametric description of mutual fund account durations. The results are suggestive for semi-parametric modeling in later sections.

What does the mutual fund hazard function look like?[30] All else equal, a reasonable prior would be a constant hazard: investors flip a coin each period to decide whether to redeem or not. An uninformed, noise trader might have this hazard. Moreover, all investors in the same fund might use the same coin, though the probabilities might vary across funds. This belief is consistent with the observation that mutual fund families treat old and new shareholders uniformly. It is also consistent with the legal regulations incorporated in the 1940 Investment Company Act, as amended.

An alternative prior is that the hazard declines smoothly over time. Negative duration dependence is consistent with a belief that investors develop a "taste" for the fund that strengthens over time. Tax lock-in could drive this view. A contrasting story is that there are two classes of investors with different but constant hazards. Over time the proportion of high-risk investors declines through attrition so the pooled hazard appears to decrease.[31] Switching costs or searching costs could drive this view.

A third possibility is that the hazard increases with time. Positive duration dependence could be generated by life-cycle or targeted savings. Positive duration dependence is interesting econometrically because there is no ambiguity in interpretation: if the hazard increases, then it necessarily increases for at least a subset of the population.

The non-parametric survival distribution and hazard functions are shown in Figure 4, smoothed over a 91-day window. The generally convex survival distribution function in Panel A declines smoothly from 100% to 24% after approximately five and a half years, at which point the rate of decline is leveling off.[32] The product-limit mean survival time is 2.61 years.[33] Quartile estimates of survival times (in years) and the associated 95% confidence intervals are $0.84 \in (0.82, 0.85)$,

[29] In this paper's empirical work, there are 19 failures at the median survival time and 23,328 accounts in the risk set. For this specific i, the denominator has $\binom{23328}{19} = 7.969\text{e+}65$ terms.

[30] Loosely speaking, the hazard function represents the conditional failure probability of accounts. For example, the hazard addresses questions such as the following: given an account that has survived for two years, what is the (conditional) probability it will be closed in the next quarter?

[31] This is a general shortcoming in duration analysis: declining hazards are inherently ambiguous. Either duration dependence or heterogeneity could be driving the decline.

[32] There is a distinction between *trading* and *calendar* days. Financial transactions occur only on the former, but duration is more naturally measured in the later. Calculations in this paper are in terms of calendar days.

[33] This is estimated as $\hat{\mu} = \sum_{m=1}^{k} \hat{S}(t_{m-1})(t_m - t_{m-1})$, where $\hat{S}$ is the estimated survival distribution function and t_m is the ordered event time. This sum underestimates the mean because the largest observation is censored.

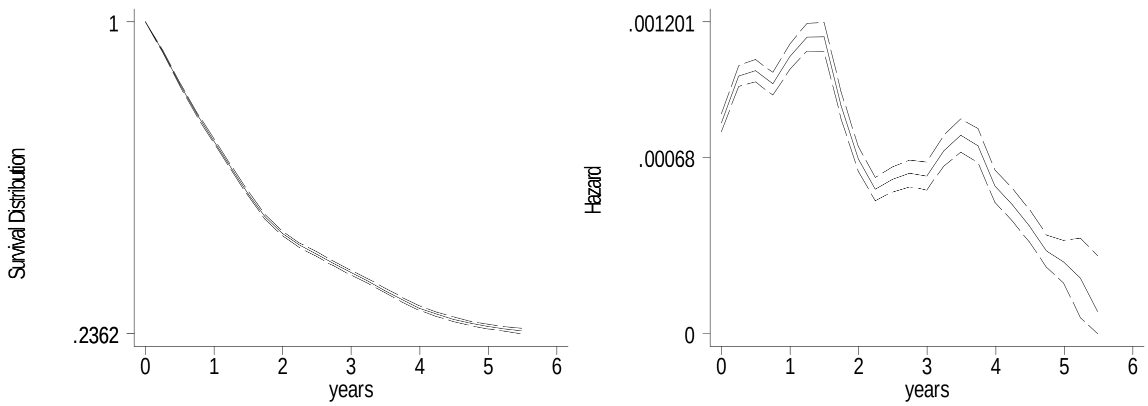

Figure 4: **Non-parametric Survival Distribution and Hazard Functions.** Plots the non-parametric survival distribution function, Panel A, and hazard function, Panel B, for all equity funds that were opened between fall 1994 and summer 2000. The usual adjustments for censoring are made. Ninety-nine percent confidence bands are included. The survival distribution function declines smoothly to 0.2362, and the mean of the hazard function is 0.000680.

$1.88 \in (1.84, 1.90)$, and $5.15 \in (4.98, 5.27)$. The hazard in Panel B declines, but not monotonically. The presence of multiple peaks is unexpected. Its mean is 0.000680.

Sample size decreases with event time. Note the declining survival distribution function. One practical implication of this is that the estimates have increasing large standard errors over event time. To rule out the possibility that the hazard's bimodality is an artifact of sampling error, 99% confidence bands are computed and added to Panels A and B. This exercise shows that the two local peaks are robust to sampling error.

Unimodal hazards appear frequently in general economic contexts, but no known study documents higher-order modality. Common finance stories do not seem equipped to explain the two extrema found in Panel B, but they are consistent with a belief that investors have fixed life-cycle horizons or that investors reevaluate the fund's performance at fixed, exogenous intervals. This evidence suggests that fully-parametric duration models may not be appropriate for the data.

6 Static Redemption Risk

This section applies the semi-parametric proportional hazards model to account lifetimes. Covariates are restricted to static data in the information set of the fund on the date the account was opened. Section 7 adds dynamic information revealed after the entry date, including subsequent fund returns and within-account trading.

The proportional hazards model is restated:

$$\lambda(t, z; \beta, \lambda_0) = e^{z'\beta} \lambda_0(t). \tag{3}$$

Because the distribution of $\lambda_0(t)$ is left unspecified, the hazard λ equals the baseline hazard λ_0 when $z = \mathbf{0}$. Thus, if the covariates are carefully measured so that $z = \mathbf{0}$ for the "average" account, then the baseline hazard can be interpreted as the hazard for the *average* or *baseline account*. Other effects are interpreted relative to the baseline. For example, if the coefficient associated with the indicator variable z_i were $\beta_i = 0.6514$, hazards of accounts possessing that characteristic are $e^{0.6514} - 1 = 91.8\%$ larger than hazards of the baseline account.[34] For this reason the estimated coefficients β_i are not of direct interest in this study. The focus instead will be on the hazard ratios e^{β_i}. If the hazard ratio is greater than one, then the account is *risky* because it is associated with higher redemption risk. Similarly, if the hazard ratio is less than one, then the account is *safe* because it is associated with lower redemption risk.

The large sample suggests that high levels of statistical significance should be required. However, it is difficult to make strong statements about economic significance. Arguments in appendix A suggest that marginal effects greater than 7.53% are probably economically significant.[35] Thus, this research's focus is on hazard ratios from indicator variables that are less than 0.9247 or larger than 1.0753 that are also statistically significant at $\alpha = 5\%$. It should be clear from context whether "significant" refers to the statistical or economic variety.

The covariates used in the regressions were defined in Table 1, and summary statistics were presented in Table 2.

6.1 Specification

The empirical specification of mutual fund account duration is motivated by the theory model: duration is a function of investor type. Those with high switching costs are low risk while those with low switching costs are high risk. The challenge is to find covariates that are direct or indirect proxies for these costs. Six types of covariates are used that collectively contribute more than one hundred dummy variables to the regressions: time, performance, fund, investor type, account options, and distribution channel.

The time dummies consist of the following: entry month (68 dummies), calendar month (12), and day of week (5). These dummies are based on the date the account was opened. The second type of covariates includes two sets of performance dummies based on relevant market indices. Historical fund returns—weekly, monthly, and quarterly—are measured in disjoint intervals at the time the account was opened.[36] These dummies equal one if and only if the fund's raw return exceeded the index. The third type of covariates consists of dummies for each of the funds. The omitted fund existed throughout the sample period. The fourth type of covariates includes the following four groups of dummies: investor age (10), size of first transaction (6), investment advisor employee status (1), and investor's state of residence (7). The fifth type of covariates consists of account options chosen by the investor at account opening: AIP/AWP participation (1), whether the initial transaction was made with a check (1), whether dividends are reinvested (1), and the selected level of telephone trading privileges (4). Lastly, a family of dummy variables indicate the distribution channel through which fund shares were purchased (8). These channels were defined in section 3.1.

The distinction between the fourth and fifth types of covariates would be important if the fund wanted to screen shareholders. Investors cannot generally change the fourth type of covariates (such as their age) in response to differential fund prices, but they can alter the fifth type of covariates (such as telephone trading privileges).[37]

The model for investor i who opens an account in fund j at time t is the log-linearized version

[34] This is the estimate for the Fund/SERV channel as shown in Table 5.

[35] A 7.53% increase in the hazard decreases the investment advisor's discounted future profit by 1%.

[36] The monthly return, for example, is the return from $t - 30$ to $t - 7$.

[37] Mutual funds are highly regulated. They could not levy higher fees against residents of a specific state. However, they could choose not to distribute fund shares in that state.

of equation 3. It is as follows:

$$\begin{aligned}\log\lambda(t, z_{i,j}; \beta, \lambda_0) \quad = \quad & (\text{time})_t\beta_1 + (\text{performance})_j\beta_2 \\ & + (\text{fund})_j\beta_3 + (\text{investor type})_i\beta_4 \\ & + (\text{account options})_i\beta_5 + (\text{distribution channel})_i\beta_6 \\ & + \log\lambda_0(t). \end{aligned} \tag{4}$$

All data used in this specification are in the fund's information set I_t at time t. There is no "look-ahead" bias. All concomitant variables are defined so that the baseline shareholder ($z = \mathbf{0}$) comes through the retail taxable channel, did not authorize telephone trading privileges, did not sign up for an automatic investment plan, reinvests dividends, and paid for the initial transaction using a check. He is not an employee of the fund and does not live in one of the six largest states or the state of the investment advisor. He discloses his age to be between 40 and 50 years old. His initial deposit was between one-half and one times the fund's minimum and was transacted on a Monday in 1994.

One of the most well-documented facts in the mutual fund literature is that aggregate mutual fund flows are highly sensitive to fund performance (Sirri and Tufano (1998), Warther (1995), and Zheng (1999)). Unreported scatterplots show that this family has attracted accounts non-uniformly over time, presumably in response to changing fund and market performance. There is also growing evidence that individual investors' trading decisions are a function of security price paths (Goetzmann and Massa (1998) and Grinblatt and Keloharju (2001)) and that portfolio allocations are related to shareholder demographics (Ameriks and Zeldes (2000)). The first four β's are, therefore, expected to be non-zero. Controlling for these effects, section 6.2 tests the hypothesis that β_5 and β_6 are zero.

The distribution channels could be related to trading frequency for two different reasons. (A similar discussion applies to the account options.) On one hand, investors are randomly assigned to distribution channels. Positive price effects imply that investors in the low-cost channel trade more than investors in the high-cost channel. Trading type is endogenously determined. On the other hand, channels serve as a screening device. Investors with preferences for high levels of trading tend to self-select into low-cost trading channels. Trading type is exogenously given. Under either interpretation, the fact that distribution channels predict duration is sufficient to show that market segmentation is possible. Nevertheless, the endogenous interpretation is taken in the discussion of the results below.

6.2 Empirical Results

This subsection discusses the estimates of equation 4. A unique ID is assigned to each account. The Breslow approximation for failure-time ties is chosen.[38] Robust standard errors are computed.[39] Unreported likelihood ratio tests show significant improvement across specifications, so only results from the final specification are discussed. The estimation results are presented in Table 5.

The age baseline (from the omitted dummy) is the group of investors between 40 and 50 years old. Accounts for which there is no age have a hazard ratio of 1.26. Apart from the twenty-year olds who are 10% riskier than the baseline, all investors under age 50 have similar redemption risk. Seventy-year olds have the largest hazard ratio at 1.66. This can be interpreted as evidence of

[38]The method of estimation was presented in section 4.4. Efron approximations do not qualitatively change the results. More sophisticated approximation techniques are computationally infeasible.

[39]Non-robust standard errors do not change the parameter estimates, and they do not qualitatively change any of the results discussed below.

log likelihood	-354414.74		-353796.42		-353423.56	
COVARIATE	HR	RSE	HR	RSE	HR	RSE
dAGE00	0.606	0.047***	0.644	0.050***	1.001	0.087
dAGE10	0.774	0.050***	0.806	0.051***	1.030	0.068
dAGE20	1.085	0.024***	1.096	0.025***	1.095	0.025***
dAGE30	0.975	0.017	0.985	0.017	0.982	0.017
dAGE50	1.175	0.021***	1.160	0.021***	1.166	0.021***
dAGE60	1.452	0.031***	1.411	0.031***	1.420	0.031***
dAGE70	1.743	0.051***	1.677	0.050***	1.662	0.049***
dAGE80	1.595	0.102***	1.553	0.100***	1.550	0.100***
dNOAGE	1.478	0.026***	1.326	0.025***	1.260	0.041***
dSIZE0	0.995	0.017	1.012	0.019	1.027	0.020
dSIZE0.5	1.100	0.015***	1.066	0.015***	1.064	0.015***
dSIZE2.5	1.178	0.022***	1.112	0.021***	1.092	0.021***
dSIZE5	1.390	0.028***	1.265	0.027***	1.237	0.026***
dSIZE50	1.948	0.153***	1.589	0.132***	1.629	0.133***
dEMPLOYEE	0.445	0.093***	0.387	0.082***	0.423	0.088***
dHQ	0.681	0.017***	0.688	0.017***	0.711	0.018***
dSTATE1	1.133	0.019***	1.124	0.019***	1.111	0.019***
dSTATE2	1.041	0.022	1.038	0.022	1.036	0.022
dSTATE3	1.163	0.027***	1.158	0.027***	1.160	0.027***
dSTATE4	0.983	0.026	0.998	0.027	1.002	0.027
dSTATE5	1.054	0.028*	1.052	0.028	1.031	0.028
dSTATE6	0.940	0.026*	0.943	0.026*	0.955	0.026
dAIPAWP			0.973	0.018	0.961	0.018*
dNOTcheck			1.622	0.032***	1.464	0.031***
dCASHdiv			1.490	0.049***	1.331	0.044***
dTELExchng			1.153	0.020***	1.330	0.027***
dTELErdmptn			1.193	0.022***	1.234	0.024***
dTELEboth			1.265	0.017***	1.339	0.020***
dRETAILtaxfree					0.858	0.018***
dMINOR					0.614	0.026***
dTRUST					1.096	0.035***
dFundSERV					1.918	0.091***
dENTITY					0.975	0.087
dSUPERMRKT					0.323	0.157*
dOTHER					1.103	0.066

Table 5: **Static Redemption Risk.** The semi-parametric Cox proportional hazards model, $\lambda(t, z; \beta, \lambda_0) = e^{z'\beta}\lambda_0(t)$, is applied to all equity accounts that were opened between fall 1994 and summer 2000. The first trade in the account is event-time zero, and the closure (censure) of the account is the (censored) failure time. The baseline hazard is left unspecified and corresponds to $z = \mathbf{0}$. The baseline account is characterized by the following: the investor is in his forties, is not an employee of the investment advisor, does not live in one of the seven flagged states, does not participate in an automatic investment plan, does not allow telephone trading, and opened the account with the fund's minimum required investment using a check through the retail taxable distribution channel. Only data in the information set of the fund the day an account is opened is used to predict the redemption probability of that account. All variables are binary and the non-reported controls include fund dummies, month dummies (68; a month in 1994 is omitted), calendar-month dummies (12; November omitted), day-of-week dummies (5; Monday omitted), and opening return dummies (week, month, and quarter). The estimated hazard ratios (HR) and robust standard errors (RSE) are displayed for three specifications. Significance at the 5%, 1%, and 0.5% levels are indicated with *, **, and ***, respectively.

changing switching costs across cohorts: those with the highest costs are preoccupied with career concerns (`d30` and `d40`) while those with the lowest cost have more free time (`d70`). A consumption story is also possible.[40]

The second group of hazard ratios come from account size dummies. The baseline account deposited between one-half and one times the fund's minimum. The smallest accounts are not significantly different from the baseline, and accounts become shorter-lived as they get larger. The largest accounts are 63% riskier than the baseline. These effects are consistent with the theory model: the smaller investor has higher switching costs because he has fewer alternatives or a smaller incentive to monitor the account. The results suggest that a more comprehensive model of mutual fund trading would include a more flexible specification of switching costs.

The funds keep track of investors who are either fund employees or relatives of fund employees. Fund employees are significantly longer-lived. However, in non-reported results, there are no significant differences in trading patterns between these accounts and the non-employee accounts. If these investors possess inside information on short-term future fund performance, they do not appear to be trading on it.

Mutual funds must register to sell their shares in every state in which they wish to do business under blue sky laws.[41] Because funds typically pay a state-specific fee that allows them to sell up to a fixed number of shares in that state, they are required to record the state of residence for every purchaser of fund shares. The seven largest states for share purchases are assigned dummy variables. The state of the management company is specially coded as `dHQ` and the other six are coded as `dSTATE1`–`dSTATE6`.[42,43]

It is surprising that two of the state dummies are statistically and economically significant, but their effects are small compared to other covariates of interest. The dummy for HQ is relatively large and statistically and economically significant, however, even after controlling for employees who presumably reside in HQ. This effect is interesting, especially in light of the home-country bias literature (see French and Poterba (1991) and Coval and Moskowitz (1999)) which documents that investors tend to own "local" assets. Here, an additional bias is found: local investors tend to be longer lived.

Some investors opt to establish automatic investment (withdrawal) plans whereby a fixed amount of money is regularly debited (credited) from (to) their bank account. There is no record of whether such a plan was established when the account was opened, only what was in effect when the account was last observed. However, since individual automatic transactions are dynamically coded as such in the transaction data, a proxy variable (`dAIPAWP`) is constructed. This dummy equals one if and only if there was an automatic transaction during the first ninety days of the account.[44] The results show that these accounts are less risky, but the effect is not economically significant.[45]

Some investors make their initial deposit with a check while others use alternative means such as

---

[40] Alternatively, finance professionals frequently advise clients to hold more bonds over the life-cycle. The evidence is broadly consistent with this advice and with some models of portfolio theory as discussed in Jagannathan and Kocherlakota (1996). Ameriks and Zeldes (2000) provide an excellent introduction to this topic.

[41] The blue sky laws are depression era regulations designed to prevent the sale of securities backed only by the "blue sky."

[42] The names of the states are not disclosed to protect anonymity.

[43] The funds have a disproportionately large share of the HQ market.

[44] Thirty days is not chosen since automatic transactions are sometimes established with a quarterly frequency. Additionally, there may be a lag in the time it takes the fund and the financial institution to make the first successful transaction.

[45] The funds keep track of only active AIP/AWP plans. A dummy for this *ex post* measure of participation has a larger effect than the *ex ante* measure presented in Table 5. This is consistent with a belief that dissatisfied investors might cancel automatic plans before deciding to liquidate the account while satisfied investors might initiate such a plan. (There are comparatively few AWP's.)

a wire or Automated Clearing House (ACH) transfer. Mailing a physical check to the fund generally takes more time and involves more uncertainty about the eventual investment date than other means. Market timers and investors with short investment horizons might find this strategy unacceptable. The evidence is consistent with this expectation: investors who do not pay with a check are 46% riskier than those who do. This difference is statistically and economically significant.

Most accounts (95.7%) reinvest dividends. Those that do not have a hazard ratio that is 33% higher than the baseline. This effect is statistically and economically significant after controlling for other variables—such as age—that might be correlated with the need for dividends.

Accounts may have telephone exchange privileges, redemption privileges, both, or none (omitted). Those accounts that select some form of telephone privileges are significantly riskier than those that do not. Either investors rationally select the telephone trading option based on their exogenous type or investors choose to trade according to their endogenously realized trading costs.

There are policy implications here. Mutual funds are increasing offering options that simplify the trading process such as telephone trading cited above and, more recently, Internet trading. This might be rational business strategy for the investment advisor if investor type is exogenous. If type is endogenous, however, reducing the investor's switching costs hurts the investment advisor in the short term by making it easier for its customers to leave. The long-term social welfare implications are not clear. With lower switching costs, will investors find funds that deliver excess returns, or will they simply migrate to funds that achieved high recent returns by luck?

The distribution channel is identified for each account.[46] The eight groups are retail taxable, retail tax-free, minors, trusts, Fund/SERV, entity, supermarket, and other. The omitted category in the regression is retail taxable, which has the largest number of accounts. As expected, the largest effect comes from the Fund/SERV channel. Accounts in this low-cost channel are 92% riskier than the baseline accounts. Minor accounts are 39% safer and trust accounts are 10% riskier than the baseline. The insignificant hazard ratios suggest that the small groups of entity and miscellaneous accounts have durations statistically indistinguishable from the baseline. The marginally significant coefficient on the omnibus accounts is ignored.[47]

The trust account dummy has the second-to-the-largest significant hazard ratio. Does this reflect lower switching costs through more active, professional management or is it a consequence of unconditionally shorter investment horizons? The retail tax-free accounts are nearly fifteen percent safer than their taxed counterparts. This is consistent with folk wisdom in the industry, but inconsistent with the tax advantages of realizing capital gains inside tax-free accounts.[48]

In the context of these channel effects, it is worthwhile to consider the fact that to the extent mutual funds choose the channels through which fund shares are distributed, they choose the type of investors in the fund. For example, funds choose whether or not to have risky investors by choosing whether or not to participate in the Fund/SERV or supermarket channels. Estimation in section 8 quantifies the cost to the fund's other shareholders when it distributes shares to observably different shareholders.

Although many of the above-mentioned individual coefficients are economically and statistically significant, it must be remembered that most accounts possess several of the measured characteristics. Suppose, for example, that a thirty-five year old opened a retail, tax-free account with a check at the fund's minimum, dividends were reinvested, no telephone privileges were authorized, and an AIP

[46] They were defined in section 3.1.

[47] Although no firm that had an omnibus relationship with the funds terminated it, there were several instances of the accounts being reregistered. Thus, the failures that exist in this category are in some sense artificial. These reregistrations not are excluded since they also exist in the other categories. In unreported regressions, dropping these accounts does not change the results.

[48] Mutual funds commonly allow lower investment limits in tax-deferred accounts despite their higher administrative costs. This suggests that investment advisors believe these accounts are somehow more profitable than the others.

were established. This account would be expected to be

$$1 - 0.982 \times 0.858 \times 1 \times 1 \times 1 \times 1 \times 0.961 = 19\%$$

safer than the baseline account. A resident of state six would be 23% safer, while one from HQ would be 42% safer than the baseline account. On the other hand, a retail-taxable account opened by a seventy-five year old depositing fifty-times the fund's minimum by wire and choosing full telephone privileges while not reinvesting dividends is 606% riskier than the average account. These combined effects are significant and relevant to both the fund and the investment advisor.

The unreported fund and time effects were qualitatively similar across specifications. However, the day-of-the-week effect did change with the controls for investor type. Monday accounts were significantly safer and Wednesday accounts were significantly riskier in the earlier specifications. In the final specification, *after controlling for investor type*, only a marginally significant Wednesday effect remains. This is consistent with Lakonishok and Maberly (1990) who find that unsophisticated individuals transact on Mondays while sophisticated institutions transact on Wednesdays.

In summary, many of the covariates analyzed in this section are economically and statistically significant predictors of account duration. The largest effects come from the fund-delivery mechanism.[49] These channels are good predictors of duration and arguably could be used by fund managers to sort investors on an *ex ante* basis.[50] The direction of these effects is consistent with a switching cost model of trading where high switching costs deters trade.

The documented pooling is not entirely based on the standard insurance argument discussed in the introduction because the covariates are known to the investment advisor when the shareholder first joined the fund. The puzzle is why funds choose to pool observably different investors.

7 Dynamic Redemption Risk

The previous section used *constant* concomitant variables, data that are available to the fund on the day an account is opened. By dynamically adding subsequent fund returns and the investor's within-account trading, this section addresses the concern that subsequent economic and fund performance might impact account retention. This robustness exercise strengthens the result by showing that the observable pooling is not simply an artifact of the earlier time-invariant assumptions.

In theory, the proportional hazards model can handle continuously-varying covariates. In practice, time is discretized. For each account i, the first *episode* begins at $t_{i,0} = 0$ and ends at $t_{i,1} = \min\{30, T_i\}$, where T_i is the account's observed duration in calendar days.[51] The second episode, if it exists, begins at $t_{i,1} = 30$ and ends at $t_{i,2} = \min\{60, T_i\}$. Thus, episodes look like $[t_{i,k}, t_{i,k+1})$, where $t_{i,k+1} - t_{i,k} = 30$ for all but possibly the final episode, and $\bigcup [t_{i,k}, t_{i,k+1}) = [0, T_i)$. Covariates for account i are assumed constant within episodes, but are allowed to jump between episodes.

Three new classes of variables are used in this specification. They are as follows: new episode-based time dummies that capture month effects (68 covariates) and calendar-month effects (12),

[49] The channel differences can be seen in another, non-parametric way by recomputing the Kaplan-Meier mean survival time for each of them. Omitting the tiny group of omnibus accounts, the mean survival times in years are as follows: 2.59 (retail taxable), 2.90 (tax-free), 3.35 (minor), 2.06 (trusts), 1.36 (Fund/SERV), 2.24 (entity), and 1.75 (other). Recall that the heavy censoring at long horizons downward biases these estimates. Refer to footnote 33.

[50] A fund family could open one growth fund to supermarket-style distribution and restrict a parallel one to direct-only investors.

[51] Thirty days is arbitrarily chosen as the base unit of time; choosing a 91 day base does not materially change the results.

Last-period-trade dummies	
dLAGbuy	true if there was a (non-automatic) purchase in the previous episode
dLAGsell	true if there was a (non-automatic) redemption in the previous episode
Ever-trade dummies	
dEVERbuy	true if there was a (non-automatic) purchase before the current episode
dEVERsell	true if there was a (non-automatic) redemption before the current episode
Fund returns	
fund_returnL1	difference between the annualized lagged fund return and the annualized average fund return
fund_returnL2	difference between the annualized twice-lagged fund return and the annualized average fund return
Account returns	
account_return	difference between the annualized account's lifetime return and the annualized average fund return

Table 6: **Definitions of Time-varying Covariates.** Defines the time-varying covariates added to the dynamic regression analysis. The trade covariates are dummies while the return covariates are continuous.

measures of within-account trading (4), and realized returns for both the fund (2) and the account (1). Unlike the static case, these time dummies change over the account life-cycle. The return variables are the only non-binary covariates used in this regression. Table 6 defines the new, time-varying variables that are added to section 6's static specification, and Table 7 provides *episode-weighted* summary statistics for these new covariates.

7.1 Empirical Results

Table 8 presents three specifications of time-varying redemption risk. The hazard ratios for the investor age, account size, employee status, and state of residence are not materially different from the static case presented in Table 5. They are not reported. The Breslow approximation for failure-time ties is again used, but non-robust standard errors are reported.[52] Unreported likelihood ratio tests show that the recent fund return variables (third specification) are more important than account lifetime returns (second specification) in predicting redemption risk. Nevertheless, the fourth specification that includes all time-varying covariates is unequivocally better than the others. Results are discussed in terms of that specification.

The first group of covariates measure non-automatic trading in the fund.[53] The recent-trade dummies (`dLAGbuy/sell`) indicate whether the account experienced a purchase or a redemption during the previous episode while the every-trade dummies (`dEVERbuy/sell`) indicate whether the account experienced a purchase or a redemption up to that point in its history.[54] The interaction between these variables must be considered when interpreting them. For example, the immediate effect of the first negative trade in an account is to increase the hazard ratio by $1.52 \times 1.51 - 1 = 130\%$. Similarly, the results show that the immediate effect of the first purchase is to make the account

[52] Robust standard errors are computationally intensive and not feasible is this larger data set. Footnote 39 suggests that they would not qualitatively change the results.
[53] Automatic trading is captured by the static dummy `dAIPAWP`.
[54] Although these trading measures do not include AIP/AWP's, they do include exchanges.

Covariate	Mean	Std. Dev.
censored	0.975	0.157
dLAGbuy	0.044	0.205
dLAGsell	0.007	0.084
dEVERbuy	0.344	0.475
dEVERsell	0.080	0.271
fund_returnL1	31.858	147.536
	6.724	median
fund_returnL2	29.026	142.123
	6.052	median
account_return	-3.358	26.437
	-5.858	median

Table 7: **Summary Statistics of Time-varying Covariates and Time-constant Channel Dummies.** Each account's lifetime is split into 30-day episodes and a final episode which may be shorter than 30 days: $[0, T_i) = \bigcup [t_{i,k}, t_{i,k+1})$. For a given account, the trading dummies start at zero but will become one later if there is an appropriate trade in the account. The one- and two-period lagged fund returns measure annualized, raw fund performance. The account return measure is the return of the fund from the date the account was opened to the start of the current episode. Both return measures have subtracted from them the average fund return during the fall 1994 to summer 2000 period. Accounts opened before fall 1994 are excluded as are all accounts in fixed-income funds. The three return variables are the only non-binary covariates used in this research. There are approximately one-and-a-half million episodes in the database.

log likelihood	-355114.31		-355108.74		-348930.17		-348748.79	
COVARIATE	HR	NRSE	HR	NRSE	HR	NRSE	HR	NRSE
dHQ	0.701	0.017***	0.702	0.017***	0.703	0.017***	0.702	0.017***
dAIPAWP	0.923	0.018***	0.923	0.018***	0.921	0.018***	0.922	0.018***
dNOTcheck	1.418	0.027***	1.419	0.027***	1.429	0.027***	1.423	0.027***
dCASHDIV	1.261	0.033***	1.262	0.033***	1.264	0.033***	1.253	0.033***
dTELExchng	1.302	0.027***	1.302	0.027***	1.295	0.027***	1.293	0.027***
dTELErdmptn	1.218	0.023***	1.218	0.023***	1.219	0.023***	1.220	0.023***
dTELEboth	1.306	0.019***	1.307	0.019***	1.307	0.019***	1.305	0.019***
dRETAILtaxfree	0.857	0.018***	0.857	0.018***	0.859	0.018***	0.861	0.018***
dMINOR	0.626	0.026***	0.626	0.026***	0.629	0.026***	0.629	0.026***
dTRUST	1.121	0.032***	1.121	0.032***	1.118	0.032***	1.118	0.032***
dFundSERV	2.236	0.089***	2.238	0.089***	2.152	0.086***	2.123	0.085***
dENTITY	0.994	0.088	0.994	0.088	0.996	0.088	0.996	0.088
dSUPERMRKT	0.235	0.097***	0.235	0.097***	0.231	0.095***	0.234	0.097***
dOTHER	1.058	0.058	1.057	0.058	1.056	0.058	1.062	0.058
dLAGbuy	0.661	0.022***	0.662	0.022***	0.666	0.022***	0.659	0.022***
dLAGsell	1.539	0.065***	1.543	0.065***	1.539	0.065***	1.517	0.064***
dEVERbuy	1.044	0.013***	1.044	0.013***	1.038	0.013***	1.035	0.013**
dEVERsell	1.500	0.028***	1.499	0.028***	1.513	0.028***	1.520	0.028***
fund_returnL1					0.992	0.000***	0.992	0.000***
fund_returnL2					0.992	0.000***	0.992	0.000***
account_return			0.999	0.000***			1.006	0.000***

Table 8: **Dynamic Redemption Risk.** The semi-parametric Cox proportional hazards model, $\lambda(t, z; \beta, \lambda_0) = e^{z'\beta}\lambda_0(t)$, is applied to all equity accounts that were opened between fall 1994 and summer 2000. The first trade in the account is event-time zero, and the closure (or censure) of the account is the (censored) failure time. The baseline hazard is left unspecified and corresponds to $z = \mathbf{0}$. The baseline account is characterized by the following: the investor is in his forties, is not an employee of the investment advisor, does not live in one of the seven flagged states, does not participate in an automatic investment plan, does not allow telephone trading, and opened the account with the fund's minimum required investment using a check through the retail taxable distribution channel. Data that is in the information set of the fund on the day the account is opened is used in the regression (every covariate from Table 5's final specification) as is *ex post* data that captures time effects, within-account trading, once- and twice-lagged fund returns, and account lifetime returns. The time dummies consist of all previous time dummies plus episode months (68; a month in 1994 is omitted) and episode calendar-months (12; November omitted). The estimated hazard ratios (HR) and non-robust standard errors (NRSE) are displayed for four specifications. Significance at the 5%, 1%, and 0.5% levels are indicated with *, **, and ***, respectively.

safer. Assuming there are no last-period trades in the account at some point during the account's lifetime (`dLAGbuy/sell` $= 0$), the relatively small but greater than one hazard ratio for the ever-trade dummy (`dEVERbuy/sell`) is initially puzzling but is consistent with the theory model: ever-bought investors have signaled their low switching costs. This effect is not economically significant, however.

The second group of covariates measures fund performance. In other contexts, studies have shown that aggregate fund flows are not particularly sensitive to the chosen measure of fund performance (Gruber (1996), Sirri and Tufano (1998), and Zheng (1999)). The present research weights accounts equally (as opposed to dollars), and Capon, Fitzsimons, and Prince (1996) document that mutual fund investors are unexpectedly unsophisticated (see footnote 5). It seems unlikely that sophisticated measures of fund return would explain account durations better than simple ones. In unreported results, no noteworthy differences are found across several metrics. A simple measure is, therefore, reported.

The demeaned, annualized, *fund return* during $[t_{i,k}, t_{i,k+1})$ for account i in fund j is

$$\text{fund return}_{j,t'_{i,k+1}} = 100 \times \left[\left(\frac{\mathrm{CR}_{j,t'_{i,k+1}}}{\mathrm{CR}_{j,t'_{i,k}}} \right)^{\frac{365}{t_{i,k+1}-t_{i,k}}} - r_j \right],$$

where $t'_{i,k}$ is the calendar time associated with the event time $t_{i,k}$, $\mathrm{CR}_{j,t'_{i,k}}$ is the cumulative return of fund j at calendar time $t'_{i,k}$, and r_j is the annualized return of fund j throughout the sample period. Subtracting r_j allows λ_0 to retains its interpretation as the hazard of the average account. In other words, the baseline account existed when the fund experienced average returns. The hazard ratio on this covariate then captures what happens when the fund performs better or worse than average.

The demeaned, annualized, *account return* at $t'_{i,k}$ is the return of fund j from the date account i was opened to the start of the episode $[t_{i,k}, t_{i,k+1})$. It is calculated as follows:

$$\text{account return}_{j,t'_{i,k}} = 100 \times \left[\left(\frac{\mathrm{CR}_{j,t'_{i,k}}}{\mathrm{CR}_{j,t'_{i,0}}} \right)^{\frac{365}{t_{i,k}-t_{i,0}}} - r_j \right],$$

where $t_{i,0} = 0$ is the start of the account's initial episode.

It is often taken for granted that investors in mutual funds receive the same return as the funds in which they invest. The returns are equal only if the investor's holding period was the same as the evaluation period of the fund. To the extent that investors can enter and exit the fund, however, their returns might diverge from those of the funds. Consider a fund that has negative returns during the first half of the year, positive returns in the second half of the year, and an annual return of zero. Suppose further that it doubles its shareholder base at mid-year. Because twice as many shareholders experienced the period of positive growth as experienced the period of negative growth, the investors as a group experienced a positive return even though the fund's return was zero. Zheng (1999) tests for Gruber's (1996) "smart money" effect and finds that funds' inflows are positively related to future fund returns. In other words, she finds that investors do have some ability to choose funds that perform well in the future.

Table 7 shows that the episode-weighted mean of the demeaned return variables are not zero. For the first lag of fund returns, the mean return is 32% and the median return is 7%. On an annualized, episode-weighted basis, the median investor outperforms the underlying fund by 7%. These investors are "smart" because they disproportionately chose to be in the fund when the fund experienced above-average returns.[55]

[55]The mean and median of the account return variable has no easy economic interpretation because these non-orthogonal returns measure the account's lifetime return.

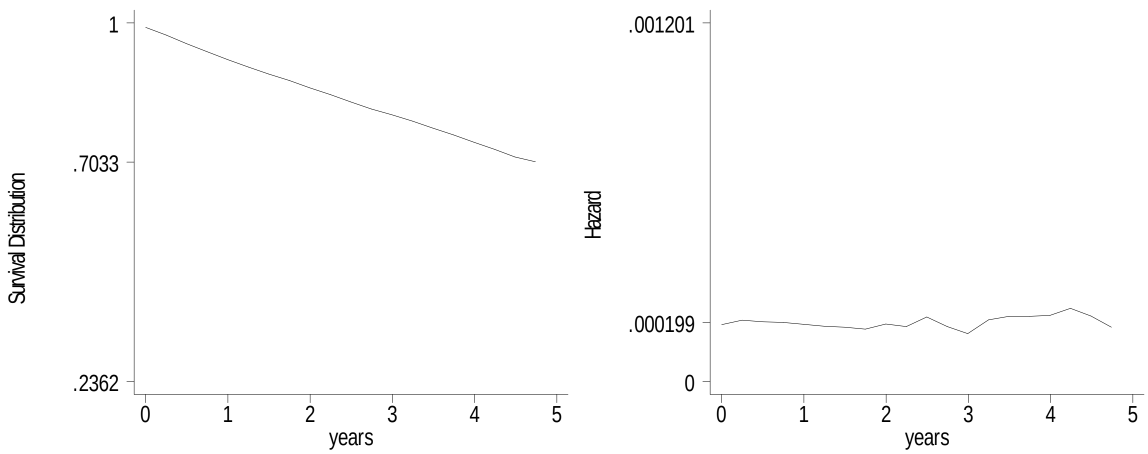

Figure 5: **Proportional Hazards Estimates of Survival Distribution and Hazard Functions.** Plots the baseline survival distribution, Panel A, and the baseline hazard functions, Panel B, for equity funds from fall 1994 to summer 2000. The vertical axes are scaled to match the non-parametric case plotted in Figure 4. The baseline survival distribution function decreases from one to 0.7033. The mean of the baseline hazard function is 0.000199.

Grinblatt and Keloharju (2001) show that returns older than one month do not generally affect the selling decision of investors in the Finnish stock market. Only one- and two-episode lags of the return variable are, therefore, included in the present article. The estimated hazard ratio for the first lag of fund returns is 0.9915. A ten percent increase in last-period annualized, average fund performance (which also increases account returns) decreases redemption risk by

$$10 \times (1 - .992 \times 1.006) = 2\%.$$

The effect of the second lag is to decrease redemption risk by the same 2%. Curiously, after controlling for recent fund returns, positive account returns increase redemption risk: a 10% increase in account returns more than two months previous raises risk by 6%. The return effects seem small relative to other covariates of interest.

Revisiting the investor and account covariates, their hazard ratios are qualitatively unchanged in this improved model, except that the automatic investment plan is now economically significant. Importantly, the channel effects are shown to be robust to changing state variables, and they remain economically and statistically significant.

7.2 Non-parametric Baselines

It is possible to recover the baseline survival distribution function S_0 and baseline hazard function λ_0. From Table 8's final specification, the baseline survival distribution and baseline hazard functions are estimated, smoothed over a 91-day window (to be consistent with section 5), and presented in Figure 5. The vertical axes are scaled to match the non-parametric case plotted in Figure 4.

The baseline survival distribution function in Panel A shows that seventy percent of the baseline accounts survive at least five years. Recall that only a quarter of all accounts survived five years in the non-parametric case. The most striking feature of the baseline hazard function in Panel B is

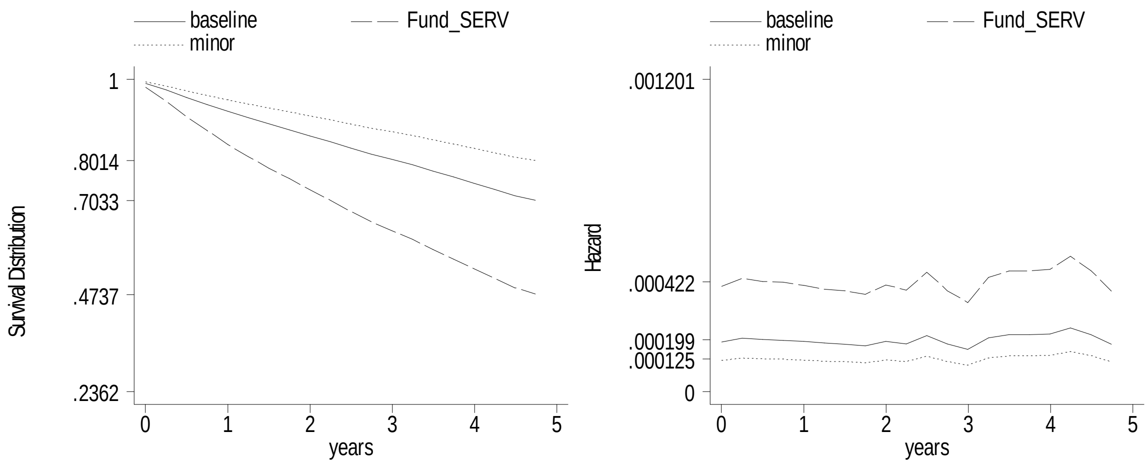

Figure 6: **Proportional Hazards Estimates of Survival Distribution and Hazard Functions for minor and Fund/SERV Accounts.** Plots the survival distribution function, Panel A, and hazard function, Panel B, for the baseline account (solid line), the Fund/SERV account (dashed line), and the minor account (dotted line) from fall 1994 to summer 2000. The vertical axes are scaled to match the non-parametric case plotted in Figure 4.

that it is relatively flat, corresponding with the nearly linear baseline survival distribution function. The mean of the baseline hazard function is 0.000199.

The peak in the baseline hazard's graph is about one and a half times the valley. Since many estimated hazard ratios of interest have effects greater than 50%, the *duration effect* of the baseline hazard is small relative to many covariates of interest. This is graphically illustrated in Figure 6. The hazard ratio for the Fund/SERV accounts is 2.123. The safest quarter of their account life-cycle is risker than the riskiest quarter of the baseline account's life-cycle. On the other hand, the hazard ratio for the minor accounts is 0.629. The riskiest quarter of this distribution channel is similarly safer than the safest quarter of the baseline account's life-cycle.

Section 4.1 illustrates that the hazard can be interpreted as a probability of conditional failure.[56] For example, the average daily baseline hazard is 0.000199. Multiplying this number by 91 gives the approximate probability (1.8%) that the baseline account fails during the next quarter.[57] This serves to reemphasize the observation about the estimated baseline survival distribution function: baseline accounts are "sticky."

In summary, this section shows that although fund returns and time effects influence redemption risk, other concomitant variables have a larger impact. It is especially noteworthy that the distribution channel effects still exist and are still strong. They are larger than most other variables of interest, even fund returns. These results suggest that funds could dramatically decrease their redemption rates were they to stop distributing fund shares through the risky channels.

The channel effects are economically large and robust to model specification. Unreported accelerated failure time and OLS regressions show equally significant channel effects.[58]

[56] Appendix A interprets the baseline hazard by constructing an artificial mutual fund and observing the realized distribution of account failures on a quarter-by-quarter basis.

[57] The minimum of the hazard is 0.000161 (1.5% of the accounts fail in that quarter) while the maximum is 0.000245 (2.2% fail).

[58] The accelerated failure time model is $\ln T_i = z'\beta + \ln T_0$. OLS estimation of this specification ignores censoring

8 Cost Estimates

Previous sections document that the funds contain shareholders with observably different liquidity needs. This section explores whether this pooling has an economic impact. In contrast to earlier work that focused on account-level data, the unit of observation here is the invested dollar. In other words, this section measures flows of dollars and not flows of accounts.

Shareholder flows impose both direct and indirect expenses on the fund. Direct costs include *implementation costs* such as brokerage commissions, bid-ask spreads, and price impact. Jones and Lipson (2001) document that these one-way costs are 85 bp for their sample of institutional investors that trade on the NYSE and AMEX. Direct costs also include the effects of $t+1$ accounting (defined below). Indirect costs include changes to the fund's portfolio made in response to fund flows, "tracking error," and will not be measured in this paper.

This section quantifies three direct costs. Section 8.1 measures the cost difference between private and pooled accounts, showing that long-term investors pay higher costs in mutual funds. An interlude in section 8.2 discusses $t + 1$ accounting. Section 8.3 measures the fund's aggregate liquidity cost. Lastly, section 8.4 measures pooling costs for the direct and indirect account types.

8.1 Private Account Cost

Investors can choose either to separate in private accounts or to pool in mutual funds. This subsection focuses on direct costs exclusively and explores which investors should pool and which should separate. In order to avoid complication, peripheral concerns such as asset divisibility and active trading strategies are ignored.

Private accounts pay transaction costs twice—first when they are created as securities are purchased and second when they are liquidated as securities are sold. Ignoring price changes and discount factors, these costs are the same for all holding periods. Using the Jones-Lipson estimate of 85 bp, the private-account cost is estimated to be 170 bp.[59]

Pooled accounts pay a proportionate amount of the fund's aggregate costs each trading day. If an indexed strategy is assumed—the fund trade only in response to shareholder flows—daily costs are simply a product of the scaled net daily flows and the implementation costs:

$$\frac{1}{N}\sum_{j=1}^{N}\sum_{t=1}^{T_j}\frac{\text{implementation cost} \times |\text{net daily flows}_{j,t}|}{\text{TNA}_{j,t}},$$

where N is the number of funds in the sample, T_j is the number of trading days for the jth fund, and $\text{TNA}_{j,t}$ is the assets in fund j at time t. The average cost of all funds in the sample is approximately 0.0044 bp per trading day or 0.0030 bp per calendar day.[60] Again ignoring price changes and discount factors, an investor's expense would be this daily cost multiplied by his holding period.

Combining these results, the *individual-level pooling cost* is simply the difference between the pooled and private account costs. Costs in the fund grow linearly with time while the private account costs are fixed. This implies that in the fund, short-term investors are better off because they get a *pooling gain* while long-term investors do worse because they pay a *pooling cost*. The *intercept* is the breakeven point. See Figure 7.

The pooling cost measure is sensitive to the choice of holding periods, so the following realistic horizons (in years) are used: 0.36, 0.84, 1.88, 2.61, and 5.15. These horizons correspond to the 10th,

and the non-normal error term.

[59] The 85 bp is the estimated institutional implementation cost. The cost for non-institutions is likely higher.

[60] Nearly all investors reinvest dividends. The fund is assumed able to reinvest those dividends at no cost in this exercise.

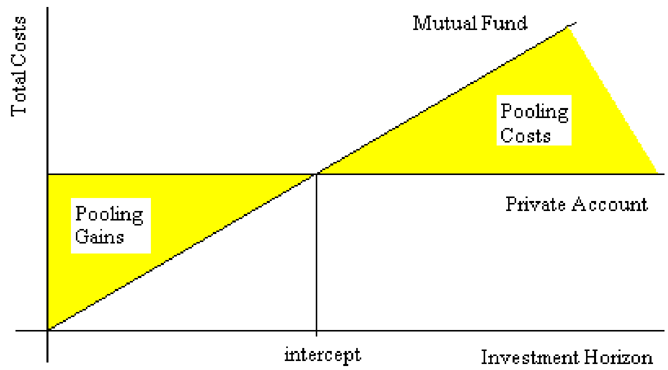

Figure 7: **Pooling Gains and Costs.** Total trading costs in the private account is time invariant. Costs in the mutual fund increase linearly with time. At the *intercept*, total costs are the same. Before that time, the mutual fund has lower costs while after that time the private account has lower costs.

25th, 50th, mean, and 75th percentiles of non-parametric account lifetimes. The annualized pooling costs are presented in Table 9.

Long-lived accounts—the 75th percentile or 5.15 years—lose 77 bp per year. On the other hand, the pooling costs for the short-lived accounts are negative and are interpreted as pooling gains. For example, at 85 bp and 0.36 years, the annualized pooling gain is +3.57%. The last item to mention is that the intercept, or the day at which the two account types yield the same expense, is 1.53 years and by construction is invariant to the implementation cost. Investors with a horizon of 1.53 years are indifferent between the two investment vehicles since both have the same costs.

8.2 $t+1$ Accounting

Unlike most financial institutions, funds close their books every business day. This poses a problem because accounts settle before the fund learns of its flows for the day. Edelen and Warner (2001) discuss the process by which the fund learns of its flows.

Once the market closes, the fund's transfer agent has to immediately strike the NAV to get it distributed through the wire services. It is calculated as follows:

$$\text{NAV}_t = \frac{\text{fund assets}_{t-1} \times \text{asset prices}_t}{\text{fund shares}_{t-1}}.$$

Although the $t-1$ shareholder flows technically occurred on date $t-1$, they will not affect the fund's NAV until t. This distinction is important if the fund's return is non-zero from $t-1$ to t. Additionally, since asset flows are also reported to the transfer agent a day late, $t-1$ manager flows will not affect the NAV until t. This distinction is important if the $t-1$ intraday asset return is non-zero. The net effect is that the full impact of shareholder flows might not be felt for two days.

To illustrate the accounting effect on the fund, consider an investor who buys shares of the fund today at \$5.00 per share. The fund manager will not learn of this purchase until tomorrow. If the fund's assets have a +2% return before that time, then the fund, in essence, sold \$5.10 worth of securities to the investor for \$5.00. In a detailed example provided in appendix B, an index fund underperforms its benchmark by 46 bp (or 23% of the index return) over a four-day period due to this effect.

Cost (basis points)		Holding periods (years)					
		10th	25th	intercept	median	mean	75th
Implementation	Calendar-day	0.36	0.84	1.53	1.88	2.61	5.15
18	0.0006	-0.77	-0.20	0.00	0.04	0.10	0.16
50	0.0018	-2.10	-0.54	0.00	0.12	0.27	0.46
85	0.0030	-3.57	-0.92	0.00	0.21	0.46	0.77
100	0.0036	-4.19	-1.08	0.00	0.24	0.54	0.90
193	0.0069	-8.04	-2.08	0.00	0.47	1.04	1.70

Table 9: **Individual-level Annualized Pooling Cost Estimate.** Presents the annualized difference between total costs in a private account and in a mutual fund. Negative entries are interpreted as *pooling gains*. The pooling cost measure is sensitive to the choice of holding periods, so the following horizons (in years) are used: 0.36, 0.84, 1.88, 2.61, and 5.15. These horizons correspond to the 10th, 25th, 50th, mean, and 75th percentiles of non-parametric account lifetimes. The intercept (1.53 years) is the point at which both accounts have the same costs. Implementation costs run from 18 bp to 193 bp. Fund-level costs are generated assuming the fund trades the net shareholder flows daily.

The implication of this analysis is simply that if shareholders trade in the same direction as asset returns, then the accounting system benefits the individual at the expense of the fund. On the other hand, if shareholders trade in the opposite direction, the effect is ambiguous. For example, suppose a shareholder purchases shares the day before a large drop in the value of the fund's portfolio. The fact that the fund gets to purchase securities for the portfolio at the new, lower price more than compensates it for the associated transaction costs. If the drop in asset prices were small (say, three basis points), however, then the price savings do not compensate the fund for the trading costs.

This mispricing is distinct from the non-synchronous trading effect documented by Greene and Hodges (2001), Goetzmann, Ivković, and Rouwenhorst (2000), and Chalmers, Edelen, and Kadlec (2001). They show that if fund shares trades non-synchronously (for example, fund assets stop trading at 1:00 p.m. while the fund allows trading in its shares until 4:00 p.m. when they are priced), traders can earn a higher return than the underlying fund with less risk. The $t+1$ accounting effect, in contrast, is driven by after-the-market-closes price changes. To the extent the fund assets have a positive expected return, the accounting effect is a systematic mispricing of fund shares.

8.3 Liquidity Cost

The mutual fund stands ready each day to buy or sell an unlimited number of its shares. This liquidity provision is expected to be costly and to impact negatively fund returns. A comprehensive estimate of this cost is beyond the scope of the present paper.[61] Instead, a simple approximation is made by adding back daily estimated trading costs to the fund's return. In other words, given the actual fund returns in the database ("post-flow fund returns"), a new return series ("pre-flow fund returns") is generated by adding to it 57 bp of the daily net shareholder flows.[62] The difference between these series is the fund's cost of trading shareholder flows. Formally, the *liquidity cost* in

[61] One technical methodology can be found in Edelen (1999). He regresses semi-annual manager flows on monthly shareholder flows in a sample of 166 funds. The present family is too small to mimic his approach.

[62] The Jones-Lipson 85 bp estimate of trading costs is based on NYSE and AMEX trades of the average institutional investor. However, this simulation assumes the fund manager indexes: she mechanically trades net flows daily and holds no cash. The Jones-Lipson trading cost estimate for index managers, 57 bp, is chosen as the more conservative estimate.

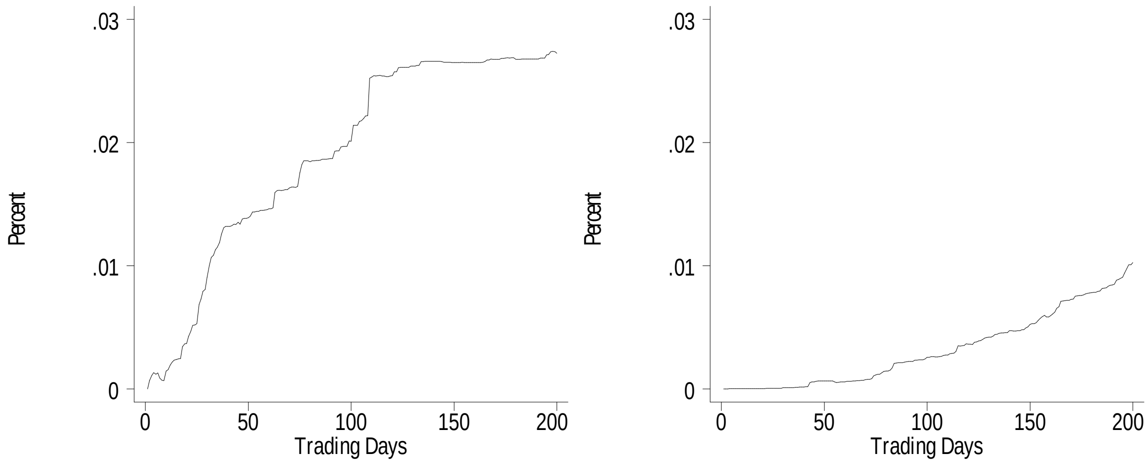

Figure 8: **Liquidity Cost Over 200 Trading Days.** Liquidity cost is the geometric difference between the pre-flow fund return and the post-flow fund return. Panel A plots the daily liquidity cost in a simulated mutual fund that had high flows relative to fund assets. Panel B plots the costs in a low flow-to-assets fund. Both plots cover 200 trading days.

fund j at time t is the geometric difference between the cumulative pre-flow fund return and the cumulative post-flow fund return:

$$\mathrm{LC}_{j,t} = \log\left(\frac{\text{cumulative pre-flow fund return}_{j,t}}{\text{cumulative post-flow return}_{j,t}}\right),$$

where the cumulative returns are normalized to one when the fund enters the sample.

Three variables will influence this measure of fund-level liquidity cost. The first consideration is the choice of the implementation cost. The second issue is flow volume *relative to fund assets*. A $1 million trade costs the fund $5,700. This expense is more significant on a per-invested-dollar basis in a $5 million fund than in a $500 million fund. The final concern is the timing of the fund flows *relative to next-day fund return* due to the $t + 1$ accounting effect. A $1 million purchase the day before a +10% (−10%) spike in asset prices costs the fund $100,000.00 ($-\$100,000.00$).

The annualized liquidity cost is computed fund-by-fund. The equally weighted, arithmetic average is 90 bp. Figure 8 shows representative examples of the liquidity cost for a specific high flow-to-asset fund and a specific low flow-to-asset fund over 200 trading days. Both funds were collecting assets daily, but Panel A's fund had a much higher growth rate and hence higher costs.[63] The high flow-to-assets funds' average is 120 bp while the low flow-to-assets funds' average is 46 bp.

One weakness of using the Jones-Lipson implementation cost measure in this context is that if a large trade raises (lowers) asset prices, then there is a residual benefit (cost) to the fund insofar as the fund already owned shares in that asset. Capturing this effect—which might be large in some markets—is beyond the scope of the present paper.

[63] *Early adopter* costs are found in many industries. For example, early cellular phone subscribers paid to build the infrastructure that is now in place for current and future users.

	85 bp		57 bp		0 bp	
Account Type	raw	scaled	raw	scaled	raw	scaled
Direct Only	65	96	45	65	2	-0.4
Indirect Only	173	220	124	157	25	33
Entire Fund	127	127	90	90	15	15

Table 10: **Liquidity Costs by Account Type.** Presents the average liquidity cost (in basis points) by account type assuming the fund's implementation costs are 85 bp, 57 bp, or 0 bp. Costs are computed using raw and scaled net flows, where the scaling is in terms of total fund-level flows.

8.4 Pooling Cost

It was shown above that the fund underperforms its benchmark by 90 bp per year due to shareholder flows. This subsection estimates what the liquidity cost would have been in a fund limited to specific types of shareholders. To keep the analysis simple, the shareholders are split into only two *ex ante* types: direct and indirect.[64] This is a test of the theory model's wealth transfer: do short-term shareholders impose a portion of their liquidity costs on the long-term investors?

Repeating the analysis from the previous subsection, the annualized costs for the separated direct and indirect shareholders are 45 bp and 124 bp, respectively. This is consistent with the model: direct shareholders are better for the fund than the indirect type (i.e., flows from the direct shareholders are less costly to trade).

Since the indirect shareholders generate a disproportionate amount of expenses, it is clear that direct shareholders could have gotten a higher return in a fund restricted to their type. This return can be estimated as the difference between aggregate liquidity cost and the direct-only liquidity cost:

$$90 \text{ bp} - 45 \text{ bp} = 45 \text{ bp}.$$

In other words, a direct-only fund could earn annually 45 bp more than a pooled fund. The pooling cost for the indirect shareholder is -34 bp ($= 90\text{bp} - 124\text{bp}$). In other words, and in accordance with the model, indirect shareholders prefer the pooled fund because the fund's pricing mechanism allows them to shift a portion of their costs onto others. The results are presented in Table 10 for the Jones-Lipson implementation costs of 57 bp (average implementation cost for index institutional investors trading on the NYSE or AMEX) and 85 bp (average implementation cost for all institutional investors trading on the NYSE or AMEX).

This cost differential might be driven by higher flows (and hence higher aggregate implementation costs) by the indirect type. Table 4 in section 3.2 showed that the direct investors contribute 26% of the daily net fund flows while the indirect investors contribute 64%. To test for this, investor flows by type are scaled up to fund flows:

$$\text{scaled investor flows}_t = \frac{\text{investor flows}_t}{\text{investor average}} \times \text{daily fund average}.$$

The scaling does increase the liquidity cost for both types, but the ratio of direct costs to indirect costs is still about 40%.

If implementation costs are additionally assumed to be zero, the annualized pooling costs are estimated to be 15 bp (total), 0 bp (direct), and 33 bp (indirect). This is also reported in Table 10.

[64] These are the investors who purchased, respectively, directly from the fund or indirectly through a broker or other financial intermediary. See section 3.1.

This measure of pooling cost (scaled flows without implementation costs) is interesting because it captures the pure $t+1$ accounting effect at the fund level. It provides suggestive evidence that investors have some ability to predict next-day fund returns. The results are also consistent with the related non-synchronous mispricing story of Chalmers, Edelen, and Kadlec (2001). Irrespective of the cause of the positive $t+1$ accounting loss to the fund, the fact that it varies across account type is problematic.

In summary, trading shareholder flow generates costs for funds that are born by all shareholders. Investors with long horizons (in excess of 1.53 years) pay lower lifetime costs in private accounts. Symmetrically, investors with short horizons (fewer than 1.53 years) pay lower expenses in the mutual fund since the fund's pricing mechanism enables them to impose a portion of their expenses onto others in the fund.

This section also estimated the fund's aggregate liquidity cost to be an annualized 90 bp. Because these costs are disproportionately generated by account type, the low-cost, direct investor implicitly pays a pooling cost while the high-cost, indirect investor receives a pooling benefit. The direct type's pooling cost was estimated at 45 bp annually, but unreported results show that it is easy to further segment the shareholders (say, by distribution channel) into *ex ante* groups that generate larger pooling costs.[65] To the extent that investors have private information that is not revealed by this article's sorting mechanisms, the cost differential would be greater still.

Pooling generates negative externalities. A comprehensive examination that included other effects such as tax concerns would undoubtedly yield a much larger estimate of the total pooling cost. This brings the discussion back to where it started. Why do funds choose to pool observably different shareholders? Alternatively, why do low-cost shareholders choose to pool with high-cost investors?

9 Discussion

Some funds charge shareholders a *load* when shares are purchased. This fee is generally paid at the time of share purchase, though contingent deferred sales charges (CDSC's) that are billed at the time of redemption are also popular. Either way, the fee comes out of the investment and, in effect, introduces a bid-ask spread in the pricing of fund shares. Funds also have the option of implementing a *redemption fee*. Unlike front-end loads and CDSC's, this fee is paid to the fund's portfolio—and not to the investment advisor or other intermediaries—and directly benefits the fund's remaining shareholders.

Because loads are generally family specific and not fund specific, loads will not necessarily reduce the fund-level volatility of shareholder flows. For example, a new shareholder will pay a load on his initial investment in an equity fund but not on a subsequent exchange to another equity fund in the same family, even if his original investment appreciated. This point is often overlooked: to the extent that trading occurs within families instead of across families, loads by themselves do not curb redemptions.[66] Additionally, loads do not compensate the non-redeeming shareholders for trading costs as redemption fees do. Loads do not, therefore, provide a complete solution to the externalities documented in this article.

The trading-cost effect can be eliminated by charging purchase and redemption fees equal to the transaction's implementation cost. The $t+1$ accounting effect can be minimized by either giving

[65]For example, the annualized pooling cost for the retail taxable, retail tax-free, and minor channels are 54 bp, 69 bp, and 83 bp, respectively.

[66]According to the *Mutual Fund Fact Book, 2001*, one-third to one-half of redemptions are reinvested within the family.

shareholders next-day pricing or improving information flow between the investment advisor and the transfer agent.

The pooling externality is important because despite the fact that it is so easy to fix, few funds have done so (see section 2). The remainder of this section presents a potpourri of possible reasons why more funds have not chosen to address these concerns. Predictions for the future direction of the industry are also made.

9.1 Moderating Factors

The investment advisor compensation is usually a fixed percentage of assets under management. If there are economies of scale in the industry, then the advisor's profit margin rises with fund assets. Thus, the advisor of a direct-only fund has strong financial incentives to open the fund to indirect shareholders because another dollar of fund assets has essentially no management costs but does generate additional income for the investment advisor.[67]

It is possible that investment activity across distribution channels is not perfectly correlated. For example, tax law changes might make 401(k) investments attractive even when taxable investors have withdrawn from the market. If so, then standard diversification models suggest that the investment advisor would find it advantageous to participate in multiple channels.

The compartmentalization of fund services builds a virtual Chinese Wall between the investment advisor, the transfer agent, and the other parts of the organization. On this more pragmatic plane, it is possible that funds are not aware of their ability to predict duration because the data necessary to do so is housed off site by the transfer agent.

There are costs in the fund that have not been explicitly considered. Take as an example the administrative or bookkeeping fee. The Vanguard Group says that shareholder servicing costs are approximately $40 or $45 per account per year.[68] A fund's expense ratio would have to include 100 bp on a $4,000 account just to recover the administrative fee. This research shows that long-lived accounts carry small balances. Even if long-term accounts subsidize short-term accounts' trading costs, the wealth transfer likely runs the other way for administrative costs. Netted across all costs, the wealth transfer might be negligible.

In the limit, a high duration fund with no net flows begins to look like a closed-end fund. This paper's pooling externality might interact with a monitoring externality. In Calomiris and Kahn (1991), demandable debt is a key monitoring mechanism in the banking industry. If the open-end structure is a monitoring mechanism in the mutual fund industry, then the agents doing the monitoring are the active traders. Their attentiveness might be the force that better aligns incentives.[69]

This article's theory model shows that the first-period investment by the long-term investor is $1-\theta$. If the optimal size of the fund is greater than $1-\theta$, short-term investors could provide needed capital. The fund industry commonly cites this line of reasoning when defending the use of 12b-1 fees.[70]

[67] As alluded to earlier, adding indirect shareholders to the fund decreases its performance. The investment advisor is indirectly hurt to the extent direct shareholders trade off the performance difference. This effect has not been measured in this article.

[68] The Wall Street Journal, August 8, 2001, p C1.

[69] Hedge funds have non-symmetric management fees: the manager may earn 20% of fund profits and suffer no down-side penalty. Mutual funds are required to levy symmetric fees. In practice, fees are not performance based. Gruber (1996) discusses the fact that performance is not priced in no-load mutual funds.

[70] Some funds levy a fee (for example, 25 bp) on assets under management. The investment advisor uses the proceeds from this fee to promote advertise the fund and attract new shareholders. It may also be paid as a "trail" to the broker that placed the trade.

This research has focused on duration differences across investor types. More insight might be gained by examining within-channel dispersion. Consider an investor who knew his type to be long. He would prefer a fund with a short-term redemption fee that keeps high-cost investors out. However, sufficient uncertainty about his type would make a risk-averse investor unwilling to bear the redemption risk. Joining a fund that has no redemption fees could be considered the same as buying a fund with the fee while also purchasing an early-redemption insurance policy. In other words, purchasing the no-fee fund is equivalent to buying *redemption insurance*.

The fund industry persistently emphasizes the long-term nature of fund investments. This may shift investor's perception about their relative longevity in funds. Contrary to their actual long-term type, unsophisticated investors may feel they are low-duration investors and object to any policy aimed against short-term shareholders.

9.2 Future Direction

Mutual funds have many things going against them. First, a long history of academic research suggests that mutual funds do not outperform passive indices (Jensen (1968)). Second, the industry has enormous costs, ranging from advisory fees to distribution fees to shareholder servicing fees. Third, there are economically significant pooling externalities in mutual funds.

The most commonly cited benefits of mutual fund ownership are professional management, diversification, and low transaction costs. Actively-managed mutual funds face stiff competition from competitors that may more efficiently deliver these three services, especially to the long-term investor. Within the industry, actively-managed funds have lost considerable market share to index funds. Index funds now command more than nine percent of total assets in domestic and international equity funds.[71] Competitors from outside the industry include exchange-traded funds (ETF's) and folios.[72] Importantly, ETF's and folios eliminate the mutual fund pooling externalities documented in this paper. A central prediction of this research is that funds will either change their pricing mechanism to eliminate the *pooling externalities* or alternative investment vehicles will successfully compete for the high-duration, low-cost investor.

10 Conclusion

Mutual fund investors are pooled in a common financial product where one investor's actions can affect the return obtained by others in the fund. The article begins with a model of mutual fund trading. In this stylized, full-information economy, some investors have high switching costs and remain in the same fund both periods. They pay a disproportionate amount of the fund's aggregate costs and receive a lifetime return lower than the return obtained by the investors who trade.

The first goal of the analysis is to test whether *ex post* differences in account duration are predictable at the time of the shareholder's initial investment. Regressions show that the baseline hazard is flat over the account life-cycle and the conditional probability of account closure is 1.8% per quarter for the "average" shareholder. Surprisingly, the fund can predict the investor's holding period based on data the transfer agent ordinarily collects. For example, older investors, larger accounts, and Fund/SERV accounts are riskier while automatic investment plan accounts, local investors, and tax-advantaged accounts are safer. Investors find recent fund returns salient (they are less sensitive

[71] August 2001 Morningstar PrincipiaPro Plus; author's calculations.

[72] *ETF's* are essentially closed-end index funds that trade on exchanges. Because they are redeemable on demand for the underlying securities, their price closely tracks the underlying assets. *Folios* are baskets of stocks chosen by the investor at specialized brokerage houses. The baskets are traded in any desired quantity which allows investors to maintain balanced portfolios while making, for example, small monthly contributions.

to lifetime account returns), but the economic impact on redemption risk is surprisingly small. The puzzle raised by this research is why funds choose to distribute shares to observably different investors.

The second goal of the analysis is to estimate the liquidity and pooling costs. Conservative simulations estimate the fund's aggregate cost of liquidity provision to be 90 bp. This cost varies by account type. Splitting the shareholders into two *ex ante* groups, simulations show that the high-duration shareholders receive 45 bp more each year in a fund restricted to their type while the low-duration receive 34 bp less each year in a fund restricted to their type. Splitting the shareholders into more than two *ex ante* groups substantially increases these estimates. To the extent that investors know their type more precisely than the fund does, the actual pooling costs are greater still. Lastly, the $t+1$ accounting effect shows that funds systematically underprice their shares and that low-duration investors successfully exploit the mispricing. Taken together, the analysis shows that pooling heterogeneous shareholders is economically costly.

The main result of this study is that mutual funds are not Diamond-Dybvig-style banks that provide an equitable insurance benefit to all investors. Because the predictable differences in holding periods result in economically significant wealth transfers, the proportionate allocation of costs is structurally inefficient. A central prediction of this research is that funds must either change their pricing mechanism to eliminate the pooling externalities or alternative investment vehicles will successfully compete for the high-duration, low-cost investor.

This research suggests a number of promising directions for further inquiry. First, is trading type endogenous or exogenous? Second, how large are switching costs? Third, what impact do shareholder flows have on the fund's portfolio? Answers to these questions would enable the fund industry and regulators to price liquidity equitably. Meanwhile, a temporary solution to the transaction-cost externality is to have the fund charge every investor both purchase and redemption fees. Note that the fixed-length redemption fees currently used by some mutual funds do not resolve the externality, though they are an improvement to the status quo ante.

A Economic Significance

This appendix uses the estimated baseline hazard from section 7 to suggest which duration effects may be economically significant from the perspective of the investment advisor. A by-product of this investigation is a clearer understanding of how the baseline hazard affects the lifetime of accounts.

Suppose a fund's expense ratio is 100 bp and that it begins with 5,000 accounts, each valued at $1,000.[73] Let the investment advisor's return on revenue be 23% and discount factor be 20%.[74] Assume that the fund has only "average" shareholders (as defined in section 6.1) and that accounts fail only at the end of the quarters.

The investment advisor's quarterly profit is calculated by multiplying the fund's quarterly TNA by the quarterly expense ratio and the investment advisor's return on revenue.[75] For the first quarter, this is given by

$$5000 \times \$1000 \times 0.01 \times \frac{91}{365} \times 0.23 = \$2867.$$

Discounted one quarter, the present value is $2,731.

The estimated baseline hazard function from section 7 can be used to determine the number of accounts that fail at the end of each period. The approximate probability of conditional account failure in a given quarter is simply that quarter's average daily baseline hazard multiplied by ninety-one. For example, 5000 accounts begin the first quarter, and that period's average daily hazard is 0.0001906. The estimated number of failures during that period is $0.0001906 \times 91 \times 5000 = 87$.

After the first 87 fail, there are 4,913 accounts that begin the second quarter. Repeating this procedure over the 19 quarterly estimates of λ_0 and arbitrarily assuming that all accounts fail at the end of the 20th quarter (the baseline hazard is only estimated out through quarter 19), the mutual fund manager's profit is $31,938. Table 11 presents the quarter-by-quarter calculations.

If the previous exercise were repeated with shareholders who are 7.53% riskier, then the investment advisor's profit would drop to $31,615. Two items are worth noting. The first is that this represents a 1.01% decrease in profits. In other words, an x% change in the hazard changes firm profits in the opposite direction but by much less than x%. Second, on a per-shareholder basis, the increased hazard results in a "loss" of $322/5000 = 6¢ per shareholder.

The idealized value-maximizing investment advisor would not leave a penny on the table. In practice, companies probably leave many pennies lying around. The investment advisor discussed in this paper says that a 1.01% change in profits is significant. For this reason hazard ratios more than 7.53% different from one are called economically significant. This is probably a conservative threshold, however.

[73] According to the *Mutual Fund Fact Book, 2000*, the mode of mutual fund minimum investments was between $500 and $1000 in 1999.

[74] The investment advisor for the T. Rowe Price family of funds is T. Rowe Price Associates. In 1999 its net income was $239 mm and its revenues were $1,036 million. Thus, their return on revenue was 23.1%. The T. Rowe Price funds had an asset-weighted expense ratio of 76.4 bp in July 1999. The product $23.1\% \times 76.4$ bp gives an indication of how the average dollar in a T. Rowe Price mutual fund benefits the shareholders of T. Rowe Price Associates. Due to the economies of scale thought to exist in mutual funds, profit from the marginal dollar might be better approximated by $100\% \times 76.4$ bp.

[75] Fees are separately billed at the end of the quarter. In practice, managers bill daily and collect monthly from fund assets.

quarter	failure probability	start	failure	PV profit
0	1.73%	5000	87	$2,731
1	1.87%	4913	92	$2,556
2	1.82%	4821	88	$2,389
3	1.80%	4733	85	$2,234
4	1.75%	4648	81	$2,090
5	1.69%	4567	77	$1,956
6	1.66%	4490	75	$1,831
7	1.60%	4415	71	$1,715
8	1.76%	4344	76	$1,608
9	1.67%	4268	71	$1,504
10	1.97%	4197	83	$1,409
11	1.67%	4114	69	$1,316
12	1.47%	4045	59	$1,232
13	1.88%	3986	75	$1,157
14	1.99%	3911	78	$1,081
15	1.99%	3833	76	$1,009
16	2.01%	3757	76	$942
17	2.23%	3681	82	$879
18	2.00%	3599	72	$819
19	1.65%	3527	58	$764
20	1.65%	3469	3469	$716
				$31,938

Table 11: **Quarterly Profit from Assets under Management.** Quarter-by-quarter discounted profit from mutual fund accounts that fail according to the baseline hazard estimates presented in Figure 5. Accounts are valued at $1,000 each period. The fund's return is zero. The investment advisor charges a 100 bp management fee which is collected at the end of each period directly from the shareholders, at which point accounts may be liquidated. The investment advisor's return on revenue is 23% and discount factor is 20%.

B Striking Fund NAV

Once the market closes, the fund's transfer agent has to immediately strike the NAV to get it distributed through the wire services. It is calculated as follows:

$$\text{NAV}_t = \frac{\text{fund assets}_{t-1} \times \text{asset prices}_t}{\text{fund shares}_{t-1}}.$$

Although the $t-1$ shareholder flows technically occurred on date $t-1$, they will not affect the fund's NAV until t. This distinction is important if the fund's return is non-zero from $t-1$ to t. Additionally, since asset flows are also reported to the transfer agent a day late, $t-1$ manager flows will not affect the NAV until t. This distinction is important if the $t-1$ *intraday* asset return is non-zero. The net effect is that the full impact of shareholder flows might not be felt for two days.

The $t+1$ *accounting effect* on the fund's return is illustrated using flows in a simulated fund over four days, Tuesday to Friday. Suppose that the fund invests in one asset which is priced at \$5.00 until Thursday's opening bell when the price jumps to (and remains at) \$5.10. Suppose further that the fund experiences a \$300 purchase before the market closes on Wednesday and that the manager completely trades that flow when he first learns of it on Thursday. Even though the return of the underlying asset is 2% from Tuesday to Friday, the fund's return is only 1.54% due to the intermediate flow. In other words, the long-term shareholders lose 46 bp or 23% of the asset return. Note that this calculation ignores peripheral concerns such as trading costs that would further magnify the loss.

The fund has 1,000 shares outstanding on Tuesday and holds 200 shares of the asset. The fund's Tuesday NAV is calculated as follows:

$$\frac{200 \times \$5.00}{1,000} = \$1.0000.$$

The first transaction occurs when a shareholder invests \$300 on Wednesday, but neither the transfer agent nor the manager learns of it until after the market closes and the NAV is struck. The Wednesday NAV is therefore calculated as if there had been no flow:

$$\frac{200 \times \$5.00}{1,000} = \$1.0000.$$

Given this NAV, the shareholder purchase yields 300 shares of the fund on Wednesday, the fund actually holds \$300 in cash, and there are 1,300 shares outstanding in the fund. On Thursday, the manager trades the shareholder's \$300 and buys 58.82 shares of the asset for the fund.[76] The transfer agent does not learn of this asset purchase until Friday—on Thursday it thinks the fund holds cash when the market closes.[77] The Thursday NAV is calculated as follows:

$$\frac{\$300 + 200 \times \$5.10}{1,300} = \$1.0154.$$

There are no transactions of any kind on Friday, but the transfer agent does update the number of asset shares and amount of cash owned by the fund. The final NAV is calculated as follows:

$$\frac{258.82 \times \$5.10}{1,300} = \$1.0154.$$

[76] The same \$300 could have purchased 60 shares of the asset for the fund on Wednesday.

[77] Since the asset's intraday price is constant at \$5.10, this distinction is not important in this example.

Balance Sheet		Tuesday	Wednesday	Thursday	Friday
Prices and Flows	Asset Price	\$5.00	\$5.00	\$5.10	\$5.10
	Shareholder Flows	\$0.00	\$300.00	\$0.00	\$0.00
	Manager Flows	\$0.00	\$0.00	\$300.00	\$0.00
Holdings (Actual)	Fund Shares	1,000.00	1,000.00	1,300.00	1,300.00
	Cash	\$0.00	\$0.00	\$0.00	\$0.00
	Asset Shares	200.00	200.00	200.00	258.82
	New Asset Shares	0.00	0.00	58.82	0.00
Holdings (Esimated)	Fund Shares	1,000.00	1,000.00	1,300.00	1,300.00
	Cash	\$0.00	\$0.00	\$300.00	\$0.00
	Asset Shares	200.00	200.00	200.00	258.82
	NAV	\$1.0000	\$1.0000	\$1.0154	\$1.0154

Table 12: **Impact of Flows on NAV.** The balance sheet, real and assumed, of the fund over a four-day period. The top three lines present the end-of-day asset price and the day's actual shareholder and manager flows. The next four lines account for the actual holdings of the fund just after the market closes. The next three present the assumed (by the transfer agent) holdings of the fund at the end of each day. These data are used by the transfer agent to compute the daily NAV, which is reported in the table's last line.

The fund underperforms the asset by 46 bp despite the fact that the fund was fully invested at all times. This differential is a direct result of the fact that the asset price changed before the manager could react to the shareholder flow. Had the asset return been negative, the fund would have outperformed the asset.

This example is summarized in Table 12. The top three lines show the asset price and the flows. The next four lines show the *actual* holdings of the fund when the market closes but before the NAV is struck. The next three lines show the *estimated* fund holdings used by the transfer agent to strike the NAV, which is reported in the bottom line.

The only piece of information correctly used every day is the end-of-day asset price. The extent to which the other, incorrect information affects the fund's long-term return is governed by the relative size of the flows (30% in this example) and the magnitude of the asset return (2% in the example). It should be clear that even smaller flows and smaller returns could still generate large return differentials over longer periods of time.

C Proportional Hazards Likelihood

The full likelihood of the proportional hazards model can be derived by considering the observations $(T_i, 1, z_i)$ and $(T_i, 0, z_i)$. The probability of the first observation (a failed account) is the pdf $f(T_i, z_i; \beta, \lambda_0)$. For the second observation (a censored account), all that is known is that the survival time was at least T_i. Thus, this observation contributes $S(T_i, z_i; \beta, \lambda_0)$ to the likelihood. The complete contribution is

$$[f(T_i, z_i; \beta, \lambda_0)]^{\delta_i} [S(T_i, z_i; \beta, \lambda_0)]^{1-\delta_i},$$

where $\delta_i = 1$ if and only if the account failed. Assuming independent observations, the full likelihood is

$$
\begin{aligned}
L(\beta) &= \prod_{i=1}^{n} [f(T_i, z_i; \beta, \lambda_0)]^{\delta_i} [S(T_i, z_i; \beta, \lambda_0)]^{1-\delta_i} \\
&= \prod_{i=1}^{n} [\lambda(T_i, z_i; \beta, \lambda_0)]^{\delta_i} S(T_i, z_i; \beta, \lambda_0) \\
&= \prod_{i=1}^{n} [e^{z'\beta} \lambda_0(T_i)]^{\delta_i} S_0(T_i)^{e^{z'\beta}},
\end{aligned}
$$

using the result that $\lambda = f/S$ and $S(t, z; \beta) = [S_0(t)]^{\phi(z,\beta)}$. As long as λ_0 is left unspecified—the ability to do so being the primary justification for choosing the Cox model—the full likelihood cannot be maximized.

D References

- Ameriks, John and Stephen P. Zeldes, 2000, How do household portfolio shares vary with age?, *Working paper*, Graduate School of Business, Columbia University.
- Barber, Brad M. and Terrance Odean, 2000, Trading is hazardous to your wealth: the common stock investment performance of individual investors, *Journal of Finance*, 55, 773–806.
- Breslow, N. E., 1974, Covariance analysis of censored survival data, *Biometrics*, 30, 89–99.
- Calomiris, Charles W. and Charles M. Kahn, 1991, The role of demandable debt in structuring optimal banking arrangements, *American Economic Review*, 81, 497–513.
- Capon, N., G. Fitzsimons and R. Prince, 1996, An individual level analysis of the mutual fund investment decision, *Journal of Financial Services Research*, 10, 59-82.
- Chalmers, John M.R., Roger M. Edelen, and Gregory B. Kadlec, 2001, On the perils of security pricing by financial intermediaries: the wildcard option in transacting mutual fund shares, *Journal of Finance*, 61, –.
- Chordia, Tarun, 1996, The structure of mutual fund charges, *Journal of Financial Economics*, 41, 3–39.
- Coval, Joshua D. and Tobias J. Moskowitz, 1999, Home bias at home: local equity preference in domestic portfolios, *Journal of Finance*, 54, 2045–2073.
- Cox, D. R., 1975, Partial likelihood, *Biometrika*, 62, 269–276.
- Diamond, Douglas W. and Philip H. Dybvig, 1983, Bank runs, deposit insurance, and liquidity, *Journal of Political Economy*, 91, 401–419.
- Dickson, Joel and John B. Shoven and Clemens Sialm, 2000, Tax externalities of equity mutual funds, *National Tax Journal*, 53, 607–628.
- Edelen, Roger, 1999, Investor flows and the assessed performance of open-end mutual funds, *Journal of Financial Economics*, 53, 439–466.
- Edelen, Roger and Jerold Warner, 2001, Aggregate price effects of institutional trading: a study of mutual fund flow and market returns, *Journal of Financial Economics*, 59, 195–220.
- French, Kenneth R. and James M. Poterba, 1991, Investor diversification and international equity markets, *American Economic Review*, 81, 222–226.
- Goetzmann, William N., Zoran Ivković, and K. Geert Rouwenhorst, 2001, Day trading international mutual funds: evidence and policy solutions, *Journal of Financial and Quantitative Analysis*, 36, 287–309.
- Goetzmann, William N. and Massimo Massa, 2001, Index funds and stock market growth, *Journal of Business*, forthcoming, –.
- Greene, Jason T. and Charles W. Hodges, 2000, The dilution impact of daily flows on open-end mutual funds, *Journal of Financial Economics*, forthcoming, .

- Grinblatt, Mark and Matti Keloharju, 2001, What makes investors trade?, *Journal of Finance*, 56, 589–616.
- Gruber, Martin, 1996, Another puzzle: the growth in actively managed mutual funds, *Journal of Finance*, 51, 783–810.
- Jagannathan, Ravi and Narayana R. Kocherlakota, 1996, Why should older people invest less in stocks than younger people?, *Federal Reserve Bank of Minneapolis Quarterly Review*, 20 no. 3, 11–23.
- Jensen, Michael, 1968, The performance of mutual funds in the period 1945–1964, *Journal of Finance*, 23, 389–416.
- Johnson, Woodrow, 2001, On mutual fund investors, *Working paper*, Graduate School of Business, Columbia University.
- Jones, Charles M. and Marc L. Lipson, 2001, Sixteenths: direct evidence on institutional execution costs, *Journal of Financial Economics*, 59, 253–278.
- Kalbfleisch, J. D. and R. L. Prentice, 1973, Marginal likelihoods based on Cox's regression and life model, *Biometrika*, 60, 267–278.
- Kalbfleisch, J. D. and R. L. Prentice, 1980, *The Statistical Analysis of Failure Time Data* (John Wiley & Sons, New York, NY).
- Kiefer, Nicholas M., 1988, Economic duration data and hazard functions, *Journal of Economic Literature*, 26, 646–679.
- Lakonishok, Josef and Edwin Maberly, 1990, The weekend effect: trading patterns of individual and institutional investors, *Journal of Finance*, 45, 231–243.
- Lancaster, Tony, 1990, *The Econometric Analysis of Transition Data* (Cambridge University Press, New York, NY).
- Nanda, Vikram, M. P. Narayanan, and Vincent A. Warther, 2000, Liquidity, investment ability, and mutual fund structure, *Journal of Financial Economics*, 57, 417–442.
- Sirri, Erik and Peter Tufano, 1998, Costly search and mutual fund flow, *Journal of Finance*, 53, 1589–1622.
- Warther, Vincent A., 1995, Aggregate mutual fund flows and security returns, *Journal of Financial Economics*, 39, 209–235.
- Zheng, Lu, 1999, Is Money Smart? A study of mutual fund investors' fund selection ability, *Journal of Finance*, 54, 901–933.