Predictable Investment Horizons and Wealth Transfers among Mutual Fund Shareholders
Author(s): Woodrow T. Johnson
Source: *The Journal of Finance*, Vol. 59, No. 5 (Oct., 2004), pp. 1979-2012
Published by: Blackwell Publishing for the American Finance Association
Stable URL: http://www.jstor.org/stable/3694814
Accessed: 23/11/2009 12:51

THE JOURNAL OF FINANCE • VOL. LIX, NO. 5 • OCTOBER 2004

Predictable Investment Horizons and Wealth Transfers among Mutual Fund Shareholders

WOODROW T. JOHNSON*

ABSTRACT

This study analyzes the distribution of investment horizons in a large, proprietary panel of all shareholders in one no-load mutual fund family. A proportional hazards model shows that there are observable shareholder characteristics that enable the fund to predict reliably on the day each account is opened whether the account will be short term or long term. Simulations show that the liquidity costs imposed on the fund by the expected short-term shareholders are significantly greater than those imposed by the expected long-term shareholders. Combining these results, the analysis argues that mutual funds do not provide equitable liquidity-risk insurance.

INVESTORS OSTENSIBLY RECEIVE NO- OR LOW-COST LIQUIDITY from open-end mutual funds. However, their trades in fund shares may force the fund to make costly transactions in its portfolio. Because the fund pays these expenses, shareholders who trade implicitly impose the financial burden of their liquidity demands onto others in the fund. If shareholders differentially generate fund-level costs, there will be a wealth transfer from the low-cost shareholders to the high-cost shareholders. The purpose of this article is to explore shareholders' demanded liquidity and its pricing in open-end mutual funds. The first part of the inquiry asks whether shareholders' liquidity demands can be predicted. The second part examines whether shareholders' predicted demand for liquidity is proportional to the liquidity costs they bear in the fund.

The empirical analysis is based on a proprietary database that includes a panel of all shareholder transactions within and across all funds in one no-load mutual fund family between 1994 and 2000. The shareholders' liquidity transactions are one of four types: open an account, buy additional shares, sell some

*Johnson is from the University of Oregon. This article is based on the third chapter of my dissertation at Columbia University. I thank my dissertation committee for helpful guidance and constructive criticism: Charles Calomiris (my advisor), Stephen Cameron, Dwight Dennison, Lawrence Glosten, and Charles Jones. I have also benefitted from interactions with John Chalmers, Larry Dann, Richard Green (the editor), Charles Himmelberg, Laurie Hodrick, Susan Ji, Val Larsen, Megan Partch, Lenore Robbins, Joseph Stiglitz, Xiaoyan Zhang, an anonymous referee, and seminar participants at many schools during winter 2002. Special thanks go to the anonymous mutual fund family that provided proprietary data for this study and to its employees who helped in its interpretation. Financial support from the Center for the Marketing of Financial Services at Columbia University is acknowledged. This study was presented at the 2004 American Finance Association meetings in San Diego after being accepted for publication. An earlier version of this study was entitled "Pooling Externalities in Mutual Funds."

shares, or sell all shares (i.e., close the account). Johnson (2002) documents that most of these shareholders make no transaction between account opening and closing. Moreover, he shows that those who do trade tend to make transactions that are small relative to the account-opening purchase. This suggests that their liquidity needs are primarily revealed by their investment horizons. Hence, predicting liquidity needs of existing fund shareholders reduces to predicting account closures.[1]

The empirical investigation proceeds in two stages. First, it applies a duration model to the data.[2] Results show that the fund can predict shareholder investment horizons using data it ordinarily collects when accounts are opened. This finding is robust to data that are collected after account opening, including subsequent fund returns. The second stage of the analysis splits the shareholders into two ex ante groups and measures the liquidity demands of each group. Results show that their annual liquidity needs are significantly different. In other words, the shareholders' predicted demand for liquidity is not proportional to the liquidity costs they bear in the fund.

Combining the two main results, this study argues that the long-term shareholders pay for more liquidity than they demand. Similarly, the short-term shareholders pay for less liquidity than they demand. The resulting wealth transfer is noteworthy because it is predictable to the extent expected investment horizons are observable.

A narrow interpretation of this article is that it shows that mutual funds misprice their shares because shareholders do not pay for the costs their trades impose on the fund. This result contributes to the literature that examines the structure of the mutual fund as an investment vehicle for heterogeneous shareholders. Dickson, Shoven, and Sialm (2000) demonstrate that shareholder flows negatively affect mutual funds' after-tax returns. Other articles discuss nonsynchronous trading and identify inefficiencies in the industry's convention of pricing mutual funds using the underlying assets' last trades even when those last trades were executed hours or days earlier. Chalmers, Edelen, and Kadlec (2001) show that this mispricing can be profitably exploited in domestic equity funds. Goetzmann, Ivković, and Rouwenhorst (2001) and Greene and Hodges (2002) document a similar result in international equity funds. Taken together, this is growing evidence of economically significant pooling externalities in mutual funds.

[1] This may not be an appropriate measure of liquidity needs in families where investors frequently exchange assets between funds without closing their accounts. Nevertheless, evidence presented in Goetzmann and Massa (2002), Ameriks and Zeldes (2001), and elsewhere shows that mutual fund investors trade very little. This suggests that more detailed measures of liquidity needs may not be necessary in most families.

[2] The earliest known public study of account durations is the 1990 Wyatt Company analysis of shareholder persistency in load and 12b-1 mutual funds. They sampled approximately 1,500 new accounts in 1974 and 800 new accounts in 1984 and documented the *realized* account holding periods. In contrast, this study is the first known public analysis of *expected* account durations. Moreover, its data cover an entire mutual fund family, include more recent shareholders, and contain only no-load funds.

A broad reading of this article suggests that it has far-reaching implications if the main results generalize to other financial intermediaries that are thought to add value by insuring against some risk through the pooling of homogeneous agents. One example of insurance-through-pooling intermediation is found in the seminal investment model developed by Diamond and Dybvig (1983).[3] Investors are risk averse and do not know whether they will redeem early or late. Investors bear liquidity risk if they invest directly: early redeemers receive a low return and late redeemers receive a high return. The financial intermediary (called a bank) insures against this risk through an ex post wealth transfer from the long-term investors to the short-term investors. Thus, early redeemers have higher ex post utility if they choose the financial intermediary, and late redeemers have higher ex post utility if they choose to invest directly. Nevertheless, on an ex ante basis, all shareholders prefer the financial intermediary because investment horizons—and hence wealth transfers—cannot be predicted. If other financial intermediaries' investors have predictable investment horizons that cause wealth transfers, theorists will need to revisit the question of how financial intermediaries add value. In any case, the following analysis suggests that mutual funds are not Diamond-Dybvig liquidity providers, contradicting the Chordia (1996) and Nanda, Narayanan, and Warther (2000) models of mutual fund investment.

A second example of how lessons from this research may extend to other financial contracts that provide insurance is found in the mortgage-backed securities literature.[4] These securities insure against prepayment risk—the risk that mortgagors will exercise their prepayment options—by transferring it to all investors in the mortgage pool. Valuation studies suggest that these pools contain mortgagors with heterogeneous prepayment risk.[5] While it is well understood that observable holding period heterogeneity affects the value of the pool in the secondary market, an overlooked corollary suggested by this study is that it could affect mortgage prices in the primary market. If lenders could screen prepayment risk (or borrowers could credibly signal their risk), they could charge lower interest rates to low-risk borrowers and higher interest rates to high-risk borrowers.

The remainder of this discussion is organized as follows. Section I summarizes the model of mutual fund trade and the resulting wealth transfer. Section II describes the data. Sections III and IV study investment horizons in the context of single-spell, semi-parametric duration models. Section V estimates three measures of costs generated by shareholders' trades. The analysis concludes with a discussion of the wealth transfer in Section VI and a summary of the main results and major implications in Section VII.

[3] A long literature follows in the tradition of Diamond and Dybvig (1983). Recent banks-as-liquidity-insurance-providers articles include Diamond (1997), Allen and Gale (1998), and Kashyap, Rajan, and Stein (2002). Nevertheless, there are competing banking models—such as Calomiris and Kahn (1991) and Flannery (1994)—that do not rely on liquidity-risk insurance.

[4] Rick Green kindly suggested this example.

[5] This has been documented at both the aggregate level (Stanton (1995)) and micro level (Deng, Quigley, and Van Order (2000)).

I. Model

An overlapping generations (OLG) model motivates this study. The model's essential intuition is outlined in this section; the formal model appears in the Appendix.

The economy contains mutual funds that are endowed with the same constant-returns-to-scale investment technology that enables them to generate identical returns each period. The funds do not charge shareholders any kind of fee to trade fund shares. Nevertheless, funds must pay to trade their underlying assets whenever their shareholders sell fund shares. Each fund's return is reduced, therefore, whenever any of its shareholders sells shares.

There are two types of shareholders, each investing for two periods. The first type must pay a dissipative switching fee to change funds after one period. These investors will choose to remain in the same fund both periods if the fee is large enough. The second type does not pay a switching fee to change funds. These investors will choose to change funds after one period if another fund offers a marginally higher return in the next period.

The economic basis for the switching-fee heterogeneity can be motivated several ways. Shareholders may have access to different trading technologies that effectively change their price of trading. For example, some shareholders are permitted to trade over the telephone while others are restricted to making all trades in writing. Some shareholders may be more aware of alternative investments. For example, shareholders originating in mutual fund supermarkets have electronic tools that simplify the fund search process. Sophisticated shareholders—for example, the investment advisor's own employees—can easily evaluate the fund's performance. Tax-advantaged accounts have lower after-tax trading costs than taxable accounts have during periods of gains. Other shareholders may simply like to trade frequently.

The model has a unique separating equilibrium (see Theorem 1) in which the zero-switching-cost investors change funds after one period and the high-switching-cost investors do not. The essential intuition behind the equilibrium is that shareholder-level redemptions decrease fund returns, driving a wedge between the return of funds that have high redemption rates and those that have low redemption rates. This wedge will make the zero-switching-cost shareholders opportunistically switch to funds with low redemption rates. The high-switching-cost investors recognize that their return is lower than the traders' return, but they refrain from trade because their switching cost is at least as large as the trading gain.

Sections III–V test two implications of the model. The first is that switching-cost differences are associated with investment horizon differences. The second is that investment horizon differences cause a wealth transfer from long-term to short-term shareholders.

II. Data

The data for this research were supplied, generously, by an anonymous mutual fund family. The family is an open-end, no-load complex with fees and trading restrictions that are standard in the industry. It consists of approximately 10 funds, including both equity and fixed-income funds. The equity funds tend not to hold large-cap securities and are thus different from the typical fund (see Falkenstein (1996)). 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 United States, with the exception of a disproportionately large presence in the investment advisor's home state. Section VI.C discusses whether the main results of this study generalize to other mutual fund families.

The fund family's transfer agent provided an electronic copy of its database for the period between fall 1994 and summer 2000. The database contains three main files: shareholders (well over 50,000), transactions (just under one million), and funds (around 10). The shareholder file includes, for each account opened during the data window, comprehensive registration information (such as investor's age and state of residence) and account options (such as participation in the dividend reinvestment or automatic investment programs). The transaction file includes, for each fund, all shareholder transactions (such as purchases and exchanges), method of payment (such as wire, check, or exchange), and various codes for internal use. The fund file includes, for each fund, a complete history of net asset values and distributions.

The unit of observation throughout the analysis is the account and not, for example, the shareholder or the invested dollar. Thus, the fact that a particular shareholder might hold multiple funds, might hold the same fund multiple times (e.g., in both taxable and tax-advantaged accounts), and might repeatedly reregister the investment (e.g., a newly married shareholder might reregister an individual account as a joint account) is ignored; each account is treated as if it were independent. In practice, few shareholders seem to have multiple registrations, although there is a handful of accounts that has been reregistered 10 or more times. "Investor" and "shareholder" are hereafter used to mean "account."

All fixed-income accounts are excluded from the analysis because they may be traded differently from the equity accounts. All accounts opened before fall 1994 are dropped because data about transactions and account closures that occurred during that period are unavailable. Account options such as telephone redemption privileges are effective as of the earlier of the account closing date or summer 2000; the analysis assumes that the observed options were set at account opening.

A. Fund Distribution

The industry assigns social codes to accounts. A few examples of the scores of social codes used by this family are joint tenants with rights of survivorship, partnership, corporate fiduciary, omnibus house, Roth conversion, uniform gifts to minors, dealer, and trust. Some of these codes exist as check-off boxes on the application while others are assigned by the transfer agent after it reviews the application and all supporting materials. Protocols for assigning social codes are encapsulated in a written document that is used by employees of the transfer agent making the assignment. Rigorous internal 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 for this study: Retail Taxable, Retail Tax-free, Minor, Trust, Fund/SERV, Entity, Supermarket, and Other. The Retail Taxable accounts are taxable individual or joint accounts. The Retail Tax-free accounts are generally IRAs, either traditional or Roth. The Minor accounts are registered as uniform gifts or transfers to minors and are invariably managed by the minor's parent. The Trust accounts tend to be managed either by a family member or a bank. The Fund/SERV accounts originate through the Fund/SERV system and frequently lack identifying characteristics of the underlying shareholder.[6] The Entity accounts include religious organizations, clubs, associations, educational institutions, corporations, and partnerships. Supermarkets and omnibus house accounts make up the Supermarket category. The Other accounts have either undefined or omitted social codes. Sections III and IV show that the eight distribution channels are important predictors of investment horizons.

Within the eight distribution channels, funds and shareholders are connected in one of two ways, either directly or indirectly. Direct shareholders interact with the transfer agent and complete the fund's application. These shareholders originate in the Retail Taxable, Retail Tax-free, Minor, Trust, and Entity distribution channels. Indirect shareholders trade through intermediaries that in turn trade with the transfer agent.[7] These shareholders originate in the Supermarket, Fund/SERV, and Other distribution channels. Section V shows that the costs imposed by shareholder trading vary across the two account types.

B. Summary Statistics at Account Opening

Table I describes the data collected by the fund when each account is opened, and Table II provides summary statistics; data generated after account opening—for example, fund returns—are summarized in Section IV. One-fourth of accounts do not indicate the shareholder's age; the age distribution of the

[6] The Fund/SERV system is a technology provided by the National Securities Clearing Corporation that included 571 fund families in September 2001. Investors connected to this network are able to trade funds in a supermarket-like fashion at low financial and economic cost. A list of Fund/SERV participants is available at `http://www.nscc.com/directory/fundserv.pdf`.

[7] The intermediary that 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 HARRIS*direct* may choose this system. Under either of these schemes, the transfer agent does not know the identity of the ultimate investor, although in the second case it is able to observe the investor's trades. In the final bookkeeping arrangement, each account is identified by its shareholder.

Table I

Definitions of Variables Collected at Account Opening

This table defines the variables collected by the fund at account opening. Continuous data are recoded as dummy variables because they are easier to interpret in the duration model.

Age dummies	
Age 0	Shareholder's age is less than 10
Age 10	Shareholder's age is between 10 and 20
Age 20	Shareholder's age is between 20 and 30
Age 30	Shareholder's age is between 30 and 40
Age 40	Shareholder's age is between 40 and 50
Age 50	Shareholder's age is between 50 and 60
Age 60	Shareholder's age is between 60 and 70
Age 70	Shareholder's age is between 70 and 80
Age 80	Shareholder's age is greater than 80
No Age	Shareholder's age is either not available or undefined
Initial transaction dummies	
Size 0	Opening trade is less than 0.5 times the fund's minimum
Size 0.5	Opening trade is between 0.5 and 1.0 times the fund's minimum
Size 1	Opening trade is between 1.0 and 2.5 times the fund's minimum
Size 2.5	Opening trade is between 2.5 and 5.0 times the fund's minimum
Size 5	Opening trade is between 5.0 and 50.0 times the fund's minimum
Size 50	Opening trade exceeds 50.0 times the fund's minimum
Employee dummy	
Employee	Shareholder is an employee or relative of an employee of the fund
State of residence dummies	
Headquarters	Account is registered in the investment advisor's home state
State 1–State 6	Account is registered in one of six large states
Account option dummies	
AIP	There was an automatic transaction during the first ninety days
Not check	The account was not opened with a check
Cash dividends	Dividends are not reinvested
Telephone none	Neither telephone redemptions nor exchanges are allowed
Telephone exchange	Only telephone exchanges are allowed
Telephone redemption	Only telephone redemptions are allowed
Telephone both	Both telephone exchanges and redemptions are allowed
Distribution channel dummies	
Retail Taxable	Social code is taxable individual or joint household
Retail Tax-free	Social code is tax-free individual; the account is generally an IRA
Minor	Social code is uniform gifts or transfers to minors
Fund/SERV	Social code is Fund/SERV
Trust	Social code is trust
Entity	Social code is club, corporation, partnerships, or similar entity
Supermarket	Social code is supermarket or ominibus house
Other	Social code is undefined or omitted

remaining accounts is centered at 40 and is roughly normal. Three-fourths of accounts are opened with less than 2.5 times the fund's minimum initial investment. Fewer than 1% of accounts are held by fund employees. The investment advisor's home state ranks third on account registrations yet only contains half of the accounts registered in State 1. About one-third of accounts have no

Table II

Summary Statistics of Variables Collected at Account Opening

This table presents the means and standard deviations of the variables collected by the fund at account opening for all equity accounts opened between fall 1994 and summer 2000 in one anonymous mutual fund family. There are over fifty thousand accounts, the unit of observation. The censored accounts were active on the date the data set was extracted in summer 2000. All variables are dichotomous.

Variable	Mean	Std. Dev.	Variable	Mean	Std. Dev.
Age 0	0.008	0.091	Headquarters	0.071	0.257
Age 10	0.010	0.097	State 1	0.137	0.344
Age 20	0.075	0.263	State 2	0.079	0.269
Age 30	0.179	0.383	State 3	0.055	0.227
Age 40	0.197	0.398	State 4	0.044	0.205
Age 50	0.157	0.364	State 5	0.046	0.210
Age 60	0.086	0.280	State 6	0.044	0.205
Age 70	0.039	0.194			
Age 80	0.007	0.084	Telephone None	0.329	0.470
No Age	0.242	0.429	Telephone Exchange	0.160	0.366
			Telephone Redemption	0.121	0.326
Size 0	0.160	0.367	Telepone Both	0.390	0.488
Size 0.5	0.303	0.459			
Size 1	0.291	0.454	Retail Taxable	0.579	0.494
Size 2.5	0.125	0.330	Retail Tax-free	0.172	0.377
Size 5	0.115	0.318	Minor	0.041	0.199
Size 50	0.007	0.085	Trust	0.107	0.309
			Fund/SERV	0.086	0.280
Employee	0.002	0.040	Entity	0.004	0.065
			Supermarket	0.000	0.019
AIP	0.120	0.325	Other	0.011	0.106
Cash Dividends	0.043	0.203			
Not Check	0.215	0.411	Censored	0.322	0.467

telephone privileges and one-third have full privileges. Retail accounts comprise about three-fourths of all accounts, and Fund/SERV accounts make up just under 10% of the total pool. About one-third of accounts are censored, meaning they had not yet closed when the data were drawn in summer 2000.

A pairwise correlation matrix for the principal variables of interest by distribution channel is presented in Table III. The table shows that the Retail Taxable accounts are correlated with automatic investment plans ($\rho = 0.212$), the no-check flag ($\rho = -0.339$), and telephone redemption-only or redemption-and-exchange options ($\rho = 0.215$ and $\rho = 0.308$, respectively). The Retail Tax-free accounts are correlated with automatic investment plans ($\rho = -0.135$) and the telephone exchange-only option ($\rho = 0.643$). The Fund/SERV accounts are larger than the retail accounts, take cash dividends ($\rho = 0.372$), have no telephone privileges ($\rho = 0.436$), and do not initially invest using a check ($\rho = 0.573$).

Public data on mutual fund shareholders are generally unavailable. Although it is difficult to say whether shareholders in this family of funds are representative of the larger population of shareholders, the distribution of investor ages,

Table III
Pairwise Correlation Coefficients of Variables Collected at Account Opening

This table presents pairwise correlation coefficients for the principal variables of interest by distribution channel for the database derived from all equity accounts opened between fall 1994 and summer 2000 in one anonymous mutual fund family. Coefficients that are significant at the % level are triple starred.

	Retail Taxable	Retail Tax-free	Minor	Trust	Fund/ SERV	Entity	Super-market	Other
Size 0	0.038***	0.068***	0.047***	−0.081***	−0.087***	−0.012***	−0.006	−0.036***
Size 0.5	0.140***	−0.014***	0.016***	−0.080***	−0.130***	−0.018***	−0.011	−0.040***
Size 1	0.046***	−0.051***	0.005	0.021***	−0.033***	−0.004	−0.003	−0.014***
Size 2.5	−0.089***	0.007	−0.024***	0.055***	0.098***	0.001	−0.001	0.019***
Size 5	−0.196***	0.011	−0.055***	0.113***	0.208***	0.034***	0.012***	0.077***
Size 50	−0.086***	−0.011***	−0.018***	0.035***	0.095***	0.034***	0.058***	0.089***
AIP	0.212***	−0.135***	0.032***	−0.079***	−0.113***	−0.009	−0.007	−0.028***
Not Check	−0.339***	−0.087***	−0.064***	0.125***	0.573***	0.026***	0.034***	0.111***
Cash Dividend	−0.134***	−0.096***	−0.038***	0.000	0.372***	0.001	0.026***	0.051***
Telephone None	−0.197***	−0.060***	−0.005	−0.014***	0.436***	0.020***	0.008	0.017***
Telephone Exchange	−0.348***	0.643***	−0.064***	−0.053***	−0.133***	−0.013***	−0.008	−0.030***
Telephone Redemption	0.215***	−0.167***	0.012***	−0.029***	−0.113***	−0.023***	−0.007	−0.033***
Telephone Both	0.308***	−0.314***	0.045***	0.073***	−0.245***	0.006	0.003	0.028***

account balances, and trading frequencies closely matches statistics reported by the Investment Company Institute (the national association of the U.S. investment company industry). However, the funds' transfer agent anecdotally reports that shareholders in this family are "more stable" (longer term) than those in other families it services.

III. Closure Risk Estimated at Account Opening

This section applies a single-spell, semi-parametric proportional hazards model to estimate closure risk. Regressors—usually called covariates in the duration literature—are restricted to data collected by the fund on the day each account opens. Section IV adds data collected after account opening, including subsequent trades made by shareholders and fund returns.

The primary function of interest in the duration model is the hazard function $\lambda(t)$. It represents, loosely speaking, the conditional account-closure rate at event time t. Letting the random variable T be the length of time a mutual fund account is open, the hazard function can be expressed as follows:

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

Although the hazard function can be parametrically specified, there is little economic theory to suggest an appropriate specification for the mutual fund account-closure process. Erring on the side of caution, therefore, the proportional hazards model developed by Cox (1975) is selected for the analysis. In this semi-parametric model, the hazard function is specified as follows:

$$\lambda(t, z; \beta, \lambda_0) = e^{z'\beta} \lambda_0(t), \tag{1}$$

where t is measured in calendar days and z is a vector of regressors that are hypothesized to affect the distribution of fund holding periods. The empirical analysis estimates β while allowing λ_0 to vary freely.

An important feature of the Cox model is that the hazard function λ equals the baseline hazard λ_0 when $z = 0$. Thus, if the regressors are carefully defined so that $z = 0$ for the "average" account, the baseline hazard can be interpreted as the hazard of the average or baseline account. All other effects are interpreted relative to this baseline using the hazard ratio $e^{\beta_i} = (e^{\beta_i}\lambda_0)/(e^0\lambda_0)$. If the hazard ratio is greater than one, the account is called risky because it is associated with higher closure risk than that of the average account. Similarly, if the hazard ratio is less than one, the account is called safe because it is associated with lower closure risk than that of the average account. For example, if the coefficient associated with the indicator variable z_i were $\beta_i = 0.6514$, accounts possessing that characteristic are risky: they are $e^{0.6514} - 1 = 91.8\%$ more likely than the baseline account to close. The estimated coefficients β_i are not of direct interest and are not reported.

It is difficult to make strong statements about economic significance in the proportional hazards model. Unreported results show that marginal effects greater than 7.53% in absolute value alter the investment advisor's discounted future profit by 1%. Therefore, the analysis focuses on hazard ratios from indicator variables that are outside the interval (0.9247, 1.0753) and that are statistically significant at $\alpha = 5\%$.

A. Specification

The empirical specification of mutual fund closure risk is motivated by the switching-cost model: fund holding period is a function of shareholder type and fund performance. Six classes of variables are hypothesized to proxy for the shareholder's switching costs and other economic factors: time, performance, fund, shareholder characteristics, account options, and distribution channel. These six groups collectively contribute more than 100 variables collected at account opening to the regression. Continuous variables are recast as dummy variables because they are easier to interpret in the Cox model.

The time dummies capture opening month (69 dummies), calendar month (12), and day of the week (5) effects. The performance dummies indicate whether the fund's raw return exceeded its benchmark index or an alternative index in the past week, month, or quarter (6). The third group consists of dummies for each of the funds. The fourth type of regressors summarizes shareholder characteristics: shareholder age (10), size of first transaction (6), investment advisor employee status (1), and state of residence (7). The fifth type of regressors describes options chosen by the shareholder at account opening: automatic investment or withdrawal plan 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). The final type of dummy variables indicates the distribution channel through which fund shares are purchased (8).

The closure risk model for shareholder i who opens an account in fund j at time t is the log-linear version of equation (1):

$$\begin{aligned}\log \lambda(t, z_{i,j}; \beta, \lambda_0) = {} & (\text{time})_t \beta_1 + (\text{performance})_j \beta_2 \\ & + (\text{fund})_j \beta_3 + (\text{shareholder characteristics})_i \beta_4 \\ & + (\text{account options})_i \beta_5 + (\text{distribution channel})_i \beta_6 \\ & + \log \lambda_0(t). \end{aligned} \tag{2}$$

Three empirical specifications of this equation are presented in the next subsection: both β_5 and β_6 are constrained to equal the zero vector; only β_6 is constrained to equal the zero vector; neither β_5 nor β_6 are constrained.

All data used in these specifications are collected by the fund when the account is opened. Variables are measured so that the baseline shareholder ($z = 0$) is not an employee of the fund, does not live in one of the six largest states or the state of the investment advisor, is between 40 and 50 years old, and paid for the initial transaction using a check. The initial deposit was between one-half and one times the fund's minimum and was transacted on a Monday in

Table IV

Closure Risk Estimated at Account Opening

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 in one anonymous mutual fund family. The first, opening trade in the account is event-time zero. The closure of the account is observed if it occurs before summer 2000; otherwise, it is censored. The baseline hazard λ_0 is left unspecified and corresponds to $z = 0$. The baseline account is characterized by a shareholder who is between 40 and 50 years old, 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 collected by the fund when the account is opened are used to estimate the account's closure risk. All variables are dichotomous. The unreported controls include fund dummies, month dummies (69; December 1994 is omitted), calendar-month dummies (12; November is omitted), day-of-week dummies (5; Monday is omitted), and six dummies that capture whether the fund exceeded its benchmark index or an alternative index at account opening over the previous week, month, or quarter. The estimated hazard ratios (HR) and robust standard errors (RSE) are displayed for three specifications. Statistical significance at 10%, 5%, and 1% is denoted with one, two, and three asterisks, respectively. The variables are defined in Table I.

Log-Likelihood	−354,414.74		−353,796.42		−353,423.56	
Variable	HR	RSE	HR	RSE	HR	RSE
Age 0	0.606	0.047***	0.644	0.050***	1.001	0.087
Age 10	0.774	0.050***	0.806	0.051***	1.030	0.068
Age 20	1.085	0.024***	1.096	0.025***	1.095	0.025***
Age 30	0.975	0.017	0.985	0.017	0.982	0.017
Age 50	1.175	0.021***	1.160	0.021***	1.166	0.021***
Age 60	1.452	0.031***	1.411	0.031***	1.420	0.031***
Age 70	1.743	0.051***	1.677	0.050***	1.662	0.049***
Age 80	1.595	0.102***	1.553	0.100***	1.550	0.100***
No Age	1.478	0.026***	1.326	0.025***	1.260	0.041***
Size 0	0.995	0.017	1.012	0.019	1.027	0.020
Size 1	1.100	0.015***	1.066	0.015***	1.064	0.015***
Size 2.5	1.178	0.022***	1.112	0.021***	1.092	0.021***
Size 5	1.390	0.028***	1.265	0.027***	1.237	0.026***
Size 50	1.948	0.153***	1.589	0.132***	1.629	0.133***
Employee	0.445	0.093***	0.387	0.082***	0.423	0.088***

(continued)

December 1994. The account comes through the Retail Taxable channel, reinvests dividends, does not have telephone trading privileges, and has neither an automatic investment nor disinvestment plan.

The parameters β_1, β_2, and β_3 are each expected to differ from the zero vector because of the evidence that fund flows are sensitive to fund returns at both the aggregate level (Sirri and Tufano (1998), Warther (1995), and Zheng (1999)) and investor level (Goetzmann and Massa (2002) and Grinblatt and Keloharju (2001)). Additionally, β_4 is expected to differ from the zero vector because demographics have been shown to affect investors' trades (Ameriks and Zeldes (2001)).

Table IV—*Continued*

Log-Likelihood	−354,414.74		−353,796.42		−353,423.56	
Variable	HR	RSE	HR	RSE	HR	RSE
Headquarters	0.681	0.017***	0.688	0.017***	0.711	0.018***
State 1	1.133	0.019***	1.124	0.019***	1.111	0.019***
State 2	1.041	0.022*	1.038	0.022*	1.036	0.022*
State 3	1.163	0.027***	1.158	0.027***	1.160	0.027***
State 4	0.983	0.026	0.998	0.027	1.002	0.027
State 5	1.054	0.028**	1.052	0.028*	1.031	0.028
State 6	0.940	0.026**	0.943	0.026**	0.955	0.026*
AIP			0.973	0.018	0.961	0.018**
Not check			1.622	0.032***	1.464	0.031***
Cash dividends			1.490	0.049***	1.331	0.044***
Telephone exchange			1.153	0.020***	1.330	0.027***
Telephone redemption			1.193	0.022***	1.234	0.024***
Telephone both			1.265	0.017***	1.339	0.020***
Retail Tax-free					0.858	0.018***
Minor					0.614	0.026***
Trust					1.096	0.035***
Fund/SERV					1.918	0.091***
Entity					0.975	0.087
Supermarket					0.323	0.157**
Other					1.103	0.066

Controlling for these four classes of variables, this section tests the hypothesis that β_5 and β_6 are equal to the zero vector. Distribution channels (β_6) could be related to investment horizons for two different reasons. (A similar discussion applies to account options (β_5).) First, there may be treatment effects: shareholders that happen to be in a low-cost trading channel choose high levels of trade simply because trades are cheap. Second, there may be selection effects: shareholders with a priori preferences for high levels of trade will self-select into a low-cost trading channel. Under either interpretation, the fact that distribution channels predict duration is sufficient to show that market segmentation is possible, ceteris paribus. In order to simplify the discussion of the following empirical results, the selection effect interpretation is taken below.

B. Results

Table IV presents the estimation results from three specifications of equation (2) using the popular Breslow approximation for failure-time ties and robust standard errors. Only results from the third specification are discussed.

The age baseline (from the omitted dummy) is the group of shareholders between 40 and 50 years old. Accounts that indicate no age have a hazard ratio of 1.26: their conditional closure risk is 26% larger than the risk for shareholders in their 40s. Apart from the 20-year-old shareholders who are 10% riskier than the baseline, all shareholders under age 50 have similar closure risk. Seventy-year-old shareholders have the largest hazard ratio at 1.66. This can be interpreted as evidence of changing switching costs across cohorts: those with the highest costs may be preoccupied with their careers and families (Age 30 and Age 40) while those with the lowest cost may have more free time (Age 70). A life-cycle or consumption story is also possible. For example, finance professionals frequently advise clients to hold more bonds toward the end of the life cycle as the clients shift from saving assets to consuming assets. The evidence is broadly consistent with this advice and with some related models of portfolio theory as discussed in Jagannathan and Kocherlakota (1996). Ameriks and Zeldes (2001) provide an excellent introduction to this topic.

The second group of hazard ratios comes from account size dummies. The baseline account deposited between one-half and one times the fund's minimum. Results show that the smallest accounts are not significantly different from the baseline, but larger accounts are uniformly shorter term. The largest accounts are 63% riskier than the baseline. These effects are consistent with the model: the smaller shareholder has higher switching costs because he has fewer alternatives, a smaller incentive to monitor the account, or has less trading experience.

The transfer agent keeps track of shareholders who are either fund employees or relatives of fund employees. The regressions show that fund employees' accounts are significantly longer term.

Mutual funds are required to record the state of residence for every purchaser of fund shares under blue sky laws. The seven largest states for share purchases are assigned dummy variables. The investment advisor's home state is specially coded as Headquarters and the other six are coded as State 1–State 6. It is surprising that dummies for State 1 and State 3 are statistically and economically significant, but their effects are small compared with other regressors of interest. The hazard ratio for Headquarters is surprisingly small, however, even after controlling for employees who presumably reside in this state: accounts opened in this state are 29% safer. The propensity for local shareholders to have long investment horizons is interesting in light of the geography literature that documents investors' preference for local assets (e.g., French and Poterba (1991) and Coval and Moskowitz (1999)).

Some shareholders opt to establish automatic investment (withdrawal) plans through which a fixed amount of money is regularly debited from (credited 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 last appears in the data set. However, since individual automatic transactions are dynamically coded as such in the transaction data, a proxy variable AIP is constructed. This dummy equals one if and only if there is an automatic transaction during the first 90 days of the account. (Ninety days is chosen because automatic transactions are sometimes established with a quarterly frequency.) The results show that these accounts are less risky, but the effect is not economically significant.

Some shareholders make their initial deposit with a check while others use alternative means such as a bank wire or an Automated Clearing House transfer. Mailing a physical check to the fund generally takes more time and involves more uncertainty about the eventual transaction date than other means. Thus, market timers and investors with short investment horizons might find this approach unacceptable. The evidence is consistent with this expectation: shareholders who do not pay with a check are 46% riskier than those who do. This difference is statistically and economically significant.

Most accounts (96%) reinvest dividends. Those that do not reinvest dividends 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). Accounts that select some form of telephone privileges are significantly riskier than those that do not.

The account option results have practical implications for the investment advisor. Mutual funds are increasingly offering options that simplify the trading process, including telephone and Internet trading technologies. Lowering shareholders' trading costs might be rational business strategy if shareholder type were exogenously determined (i.e., if there were a selection effect). However, this might not be wise if trading preferences were endogenously determined based on the options provided by the fund (i.e., if there were a treatment effect): low costs will encourage short investment horizons. In either case, the long-term social welfare implications of lower switching costs are not clear. On the one hand, shareholders may move assets to funds that deliver excess returns, efficiently reallocating capital (see Gruber (1996) and Zheng (1999)). On the other hand, they may transfer assets to funds that achieved high recent returns by chance, imposing socially wasteful transaction costs on the economy (see Theorem 3 in the Appendix).

The final group of regressors consists of the eight distribution channel dummies. Retail Taxable is the omitted channel in the regression because it has the largest number of accounts. Entity and Other accounts have durations that are statistically indistinguishable from the baseline account. Fund/SERV accounts are 92% riskier than the baseline; these shareholders have economically and statistically shorter investment horizons. Trusts are 10% riskier than the baseline. The Retail Tax-free accounts are nearly 15% safer than the Retail Taxable accounts. This is consistent with folk wisdom in the industry, but inconsistent with the tax advantages of realizing capital gains inside tax-free accounts. Minor accounts are 39% safer than the baseline.

The marginally significant coefficient on Supermarket is ignored. Although no firm that had an omnibus relationship with the family terminated it, there were several instances of account reregistration. Thus, the closures that exist in this category are in some sense artificial. These reregistrations are not excluded from the database because similar reregistrations exist in the other channels. Dropping these accounts does not materially change the results.

It is important to understand that the fund chooses the type of its shareholders when it chooses its distribution channels. For example, the fund chooses whether or not to have risky shareholders by choosing whether or not to participate in the Fund/SERV and Supermarket channels. Simulations in

Section V quantify the costs of pooling predictably heterogeneous shareholders.

Although many of the above-mentioned individual coefficients are economically and statistically significant, the univariate effects do not capture the full range of redemption risk. Suppose, for example, that a 35-year-old shareholder opens a Retail Tax-free account with a check at the fund's minimum, reinvests dividends, does not authorize telephone privileges, and establishes an automatic investment plan. This account is 19% safer than the baseline account:

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

Were this shareholder also a resident of State 6 (Headquarters), the account would be 23% (42%) safer than the baseline account. In contrast, a Retail Taxable account opened by a 75 year old depositing 50 times the fund's minimum by wire and choosing full telephone privileges while not reinvesting dividends is 606% riskier than the baseline account. These multivariate effects appear to be large enough to concern regulators, the investment advisor, and the fund's other shareholders.

The unreported fund and time effects are qualitatively similar across specifications. However, the day-of-the-week effect changes with the controls for shareholder type. Monday accounts are significantly safer and Wednesday accounts are significantly riskier in the first two reported specifications. In the final reported specification—after controlling for shareholder 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 regressors analyzed in this section are economically and statistically significant predictors of investment horizons. The directions of these effects are consistent with the switching-cost model. Some of the largest effects come from the fund-delivery mechanism, suggesting that the fund could dramatically decrease its redemption rate were it to stop distributing shares to predictably risky investors.

IV. Closure Risk Estimated after Account Opening

The previous section uses data that are collected by the fund when accounts are opened to show that the fund can predict shareholder investment horizons. This section extends the analysis by incorporating information about the shareholder's trades, account performance, and fund performance after account opening. This robustness exercise strengthens the previous results.

A. Restructured Database

The database is restructured by splitting each account into 30-day observations. Regressors are assumed constant within each 30-day interval, but they

Table V

Definitions of Variables Collected after Account Opening

This table defines the variables collected by the fund after account opening. The trade variables are dichotomous, and the return variables are continuous.

Recent-trade dummies	
Recent Buy	There was a nonautomatic purchase in the previous 30-day interval
Recent Sell	There was a nonautomatic redemption in the previous 30-day interval
Ever-trade dummies	
Ever Buy	There was a nonautomatic purchase before the current 30-day interval
Ever Sell	There was a nonautomatic redemption before the current 30-day interval
Fund returns	
Fund Return Lag 1	Difference between the annualized lagged fund return and the annualized average fund return
Fund Return Lag 2	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

are allowed to change between intervals. Thirty days is arbitrarily chosen as a base unit of time; choosing a 91-day interval does not materially change the results.

Table V defines two new classes of variables, and Table VI provides summary statistics. The first group includes four dummy variables that measure nonautomatic trading in the fund (automatic trading is captured by the AIP dummy defined in Table I). Recent Buy and Recent Sell indicate whether the account executed a purchase or a redemption during the previous 30-day observation. Ever Buy and Ever Sell indicate whether the account executed a purchase or a redemption at any time before the start of the current 30-day observation.

Table VI

Summary Statistics of Variables Collected after Account Opening

This table presents the means and standard deviations of the after-opening trading dummies for the database that consists of all equity accounts opened between fall 1994 and summer 2000 in one anonymous mutual fund family. Each account is split into 30-day observations. For a given account, the trading dummies start at zero but become one later if there is an appropriate trade in the account. There are approximately 1.5 million observations (i.e., account months) in the database.

Variable	Mean	Std. Dev.
Censored	0.975	0.157
Recent Buy	0.044	0.205
Recent Sell	0.007	0.084
Ever Buy	0.344	0.475
Ever Sell	0.080	0.271

The second category of new regressors measures fund performance using two different yardsticks: fund return and account return.[8] For each 30-day observation, the fund's holding period return is calculated and annualized. The fund's annualized return between fall 1994 and summer 2000 is then subtracted from these returns so λ_0 retains its average-account interpretation. Thus, the hazard ratio on this forward-looking measure of returns captures how closure risk changes when the fund performs better or worse than average.

The account return captures the shareholder's cumulative return experience in the fund. It is calculated by subtracting the fund's annualized return between fall 1994 and summer 2000 from the fund's annualized holding period return between account opening and the start of the current 30-day observation. The hazard ratio on this variable captures what happens when the account performs better or worse than the fund. Note that fund and account returns are not disjoint.

The new regressions also include all time dummies based on each account's opening date from the previous section and two new sets of time dummies based on the after-account-opening observations. The first new group captures effects for each of the 69 months in the sample period; the second group captures effects for each of the 12 months in the calendar year.

B. Results

Table VII presents four specifications of time-varying closure risk. The popular Breslow approximation for failure-time ties is again used, but nonrobust standard errors are reported because robust standard errors are not computationally feasible in this larger data set. The results are discussed in terms of the fourth specification that includes all time-varying regressors. The hazard ratios for shareholder age, account size, employee status, and state of residence are not reported because they are similar to those presented in Table IV.

The first group of new variables are important predictors of closure risk. Investors who have ever sold shares are 52.0% more likely than investors who have never sold shares to close their accounts. Investors who recently redeemed are 51.7% more likely than investors who did not recently redeem to close their accounts. However, the effect of the first redemption in the account (Recent Sell and Ever Sell are both equal to one) is to increase the closure risk by $1.517 \times 1.520 - 1 = 131\%$.

Closure risk is reduced by both recent purchases (34%) and the first purchase in the account ($1 - 0.659 \times 1.035 = 32\%$). However, closure risk is higher in accounts that have made a purchase but not recently (Ever Buy is one and

[8] 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 weighting by account value), and Capon, Fitzsimons, and Prince (1996) document that mutual fund shareholders are unexpectedly unsophisticated. It seems unlikely that sophisticated measures of fund return would explain account durations better than simple ones. Because no noteworthy differences are found across several specifications, a simple measure is reported.

Table VII

Closure Risk Estimated after Account Opening

The semi-parametric Cox proportional hazards model, $\lambda(t, z;\ \beta, \lambda_0) = e^{z'\beta}\lambda_0(t)$, is estimated for all equity accounts that were opened between fall 1994 and summer 2000 in one anonymous mutual fund family. The first, opening trade in the account is event-time zero. The closure of the account is observed if it occurs before summer 2000; otherwise, it is censored. The baseline hazard λ_0 is left unspecified and corresponds to $z = 0$. The baseline account is characterized by a shareholder who is between 40 and 50 years old, 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. The regression includes data that are collected by the fund on the day the account is opened (i.e., every regressor from the third specification of Table IV) and data that are collected after account opening (including fund returns, account returns, shareholder trades, and dummies that capture more robust time effects). The estimated hazard ratios (HR) and non-robust standard errors (NRSE) are displayed for four specifications. Statistical significance at 10%, 5%, and 1% is denoted with one, two, and three asterisks, respectively. The variables are defined in Tables I and V.

Log Likelihood	−355,114.31		−355,108.74		−348,930.17		−348,748.79	
Variable	HR	NRSE	HR	NRSE	HR	NRSE	HR	NRSE
Headquarters	0.701	0.017***	0.702	0.017***	0.703	0.017***	0.702	0.017***
AIP	0.923	0.018***	0.923	0.018***	0.921	0.018***	0.922	0.018***
Not Check	1.418	0.027***	1.419	0.027***	1.429	0.027***	1.423	0.027***
Cash Dividends	1.261	0.033***	1.262	0.033***	1.264	0.033***	1.253	0.033***
Telephone Exchange	1.302	0.027***	1.302	0.027***	1.295	0.027***	1.293	0.027***
Telephone Redemption	1.218	0.023***	1.218	0.023***	1.219	0.023***	1.220	0.023***
Telephone Both	1.306	0.019***	1.307	0.019***	1.307	0.019***	1.305	0.019***
Retail Tax-free	0.857	0.018***	0.857	0.018***	0.859	0.018***	0.861	0.018***
Minor	0.626	0.026***	0.626	0.026***	0.629	0.026***	0.629	0.026***
Trust	1.121	0.032***	1.121	0.032***	1.118	0.032***	1.118	0.032***
Fund/SERV	2.236	0.089***	2.238	0.089***	2.152	0.086***	2.123	0.085***
Entity	0.994	0.088	0.994	0.088	0.996	0.088	0.996	0.088
Supermarket	0.235	0.097***	0.235	0.097***	0.231	0.095***	0.234	0.097***
Other	1.058	0.058	1.057	0.058	1.056	0.058	1.062	0.058
Recent Buy	0.661	0.022***	0.662	0.022***	0.666	0.022***	0.659	0.022***
Recent Sell	1.539	0.065***	1.543	0.065***	1.539	0.065***	1.517	0.064***
Ever Buy	1.044	0.013***	1.044	0.013***	1.038	0.013***	1.035	0.013***
Ever Sell	1.500	0.028***	1.499	0.028***	1.513	0.028***	1.520	0.028***
Fund Return Lag 1					0.992	0.000***	0.992	0.000***
Fund Return Lag 2					0.992	0.000***	0.992	0.000***
Account Return			0.999	0.000***			1.006	0.000***

Recent Buy is zero): they are 3.5% more likely than the baseline account to close. This effect is not economically significant, but it is consistent with the switching-cost model: having revealed their low switching costs, ever-bought shareholders are more likely to trade.

The second group of new variables is less important than the first. The estimated hazard ratio for the first lag of fund returns is 0.992. A 10% increase in last-period annualized, average fund performance (which also increases account returns) decreases closure risk by $10 \times (1 - 0.992 \times 1.006) = 2\%$. The effect of the second lag is to decrease closure risk by about the same 2%. Curiously, after controlling for recent fund returns, positive account returns increase closure risk: a 10% increase in account returns more than 2 months previous raises risk by 6%.

The hazard ratios of the other regressors of interest are qualitatively unchanged in this model, except that the automatic investment plan is now economically significant. Importantly, the channel effects remain economically and statistically significant.

In summary, this section shows that although trades made after account opening, fund returns, and account returns affect closure risk, other data available at the account's opening generally have larger effects. In particular, distribution channel effects are larger than the effects of most other variables of interest, including fund returns. Because regressors that predict duration are collected by the investment advisor when the shareholder first joins the fund, it is clear that mutual funds do not pool homogeneous shareholders to insure against liquidity risk as suggested by Diamond and Dybvig (1983), Chordia (1996), Nanda, Narayanan, and Warther (2000), and others. One puzzle raised by this study is why funds choose to pool observably different shareholders.

V. Cost Estimates

The previous two sections document that shareholders in one mutual fund family have predictably different liquidity needs. This section measures whether the shareholders' predicted demand for liquidity is proportional to the liquidity costs they bear in the fund.

The fund incurs liquidity costs as investors trade its shares. Five components of these costs are outlined as follows. First, shareholder flows impose transaction costs—brokerage commissions, bid-ask spreads, and price impact—on the fund if the portfolio manager trades the flows. Jones and Lipson (2001) document that one-way transaction costs are 85 bp for their sample of institutional investors that trade on the NYSE and the AMEX.

Second, shareholder flows may force the fund to trade at inopportune times. Edelen (1999) argues that the fund will earn positive risk-adjusted returns when the portfolio manager trades on private information. He predicts, however, that the fund will incur losses when the portfolio manager trades to satisfy shareholders' liquidity needs.

Third, shareholder flows will make the fund deviate from its preferred portfolio weights. Purchases will result in excess cash in the fund's portfolio; redemptions have the opposite effect. Even after the flow is fully traded, there may be tracking error if the flow was not large enough to be traded across each of the fund's assets. The other side of the coin is that shareholder flows allow a misaligned portfolio to be rebalanced more cheaply than if there were no flows. In the case of net inflows, for example, underweighted assets could be purchased without the sale of overweighted assets.

Fourth, shareholder flows may alter the type of securities placed in the fund's portfolio. Chordia (1996) provides evidence that mutual funds hold more liquid assets when there is more uncertainty about shareholder redemptions. For example, closed-end funds' portfolios may contain more illiquid assets than open-end funds' portfolios contain because closed-end funds do not need to sell assets to satisfy shareholder redemptions.

Last, shareholder flows may change the fund's distributions. This will affect the return obtained by taxable shareholders. On the one hand, their tax burden will increase if net outflows force the portfolio manager to sell appreciated assets because the fund is required to pass through gains to its shareholders proportionately. On the other hand, their tax burden will decrease if net inflows dilute the fund's realized and unrealized capital gains through the creation of new shares. Dickson et al. (2000) report that this tax externality may have cost shareholders more than 10 bp per month between 1983 and 1998.

A comprehensive analysis of these five costs is beyond the scope of this study. Moreover, data required to measure precisely these costs are unavailable. For example, the tax basis of the fund's portfolio must be known before the tax effect can be measured accurately. For these reasons, the liquidity cost estimates presented in this section are entirely based on Jones-Lipson transaction costs and Edelen uninformed trading costs. This necessary simplification suggests that the reported estimates will understate the total costs shareholder flows impose on the fund.

A. Results

The fund's technology enables it to generate a particular series of returns. As shareholders trade fund shares, however, the fund will make extra trades and incur extra costs. Thus, the post-shareholder-flow fund returns will be inferior to the pre-shareholder-flow fund returns. The difference between these returns is the fund's cost of liquidity provision. Formally, fund j's liquidity cost at time t is the geometric difference between the cumulative pre-flow fund return and the cumulative post-flow fund return:

$$\text{liquidity cost}_{j,t} = \log\left(\frac{\text{cumulative pre-flow fund return}_{j,t}}{\text{cumulative post-flow fund return}_{j,t}}\right), \tag{3}$$

where the cumulative returns are normalized to one when the fund enters the sample. The liquidity cost is annualized and averaged across all funds in the family. The difference between indirect and direct shareholders' average liquidity cost is the liquidity difference:

$$\text{liquidity difference} = \text{liquidity cost}_{\text{indirect}} - \text{liquidity cost}_{\text{direct}}.$$

This statistic will be zero if there are no wealth transfers between the indirect and direct shareholders. The difference between the fund's liquidity cost and that of one shareholder type is the pooling cost:

$$\text{pooling cost}_{\text{type}} = \text{liquidity cost}_{\text{fund}} - \text{liquidity cost}_{\text{type}},$$

where "type" is either indirect or direct. This statistic estimates how much a shareholder type loses (or gains, if negative) in the pooled fund.

Note that the liquidity cost statistic is affected by the ratio of shareholder flows to fund assets. For example, \$10,000 in transaction costs is more significant on a per-invested-dollar basis in a \$5 million fund than in a \$500 million fund. Thus, funds with high percentage flows (e.g., rapidly growing funds) incur high liquidity costs. Nevertheless, the liquidity difference statistic is not affected by fund growth if the indirect and direct growth rates are equal: the liquidity difference statistic is zero if both indirect and direct assets grow at the same rate. The key observation necessary to see this is that liquidity costs associated with a given percentage inflow will be the same for both indirect and direct shareholders because liquidity costs are stated as percentages.

The liquidity cost model in equation (3) has four inputs. The first is fund returns. The post-flow fund returns are the realized returns of the funds in the database. The pre-flow fund returns are simulated by adding the estimated shareholder-imposed transaction costs to the post-flow fund returns.

The second input is the fund's transaction costs. They are affected by, among other things, the fund's underlying assets and management style. To conserve space, however, a menu of only the following four management-style transaction cost estimates reported by Jones and Lipson (2001) is provided: 152 bp (momentum manager), 85 bp ("average" manager), 57 bp (index manager), and 50 bp (value manager). No attempt is made to adjust for the fact that if a large trade raises (lowers) an asset's price, there is a residual benefit (cost) to the fund insofar as the fund already owned shares of that asset.

The third input is the process by which the fund responds to shareholder flows. One model regresses fund trades on shareholder flows:

$$\text{fund purchases} = \alpha_1 + \beta_1(\text{shareholder purchases}) + \epsilon_1,$$

$$\text{fund redemptions} = \alpha_2 + \beta_2(\text{shareholder redemptions}) + \epsilon_2,$$

where β_1 and β_2 capture the shareholders' incremental effect on the fund. This model cannot be estimated because the fund's trades are not available. However, Edelen (1999) estimates it for a sample of mutual funds and reports that 70% of gross shareholder flows are traded by the fund (i.e., $\hat{\beta}_1 \approx \hat{\beta}_2 \approx 0.70$). The remaining 30% of gross shareholder flows are either netted across some accumulation period or are incorporated into trades the fund would have otherwise made. These results are used to estimate a gross-flow model.

Edelen (1999) presents a second model of the fund's response to shareholder flows: the fund nets shareholder flows over an accumulation period before trading them. This approach corresponds more closely with the intuition that net (not gross) shareholder flows cause the fund to trade. However, this methodology is sensitive to assumptions about the fund's cash management strategy. For example, trading costs will depend on the number of days the fund accumulates flows before it trades them. Additionally, the fund's return during the accumulation period will depend on how the accumulated cash is held. For example, noninterest bearing cash will behave differently from S&P 500 futures contracts. These cash management strategies may be especially relevant for funds that have high portfolio turnover and funds that do not track popular indices. For these reasons, Edelen's net-flow model may be inferior to his gross-flow model. Nevertheless, a net-flow model, too, is estimated.

The final input in the liquidity cost model is shareholder flows. To simplify the analysis, account-level transactions are aggregated into daily gross flows from three types of shareholders: direct (accounts opened through the transfer agent), indirect (accounts opened through an intermediary), and all (the sum of indirect and direct). Results from Sections III and IV suggest that direct shareholders' liquidity cost will be less than indirect shareholders' liquidity cost.[9]

The gross-flow model is presented in Panel A of Table VIII. Pre-flow fund returns are estimated by adding daily transaction costs for 70% of gross shareholder flows to the post-flow fund returns. Results show that liquidity provision costs the fund 112 bp annually if trading costs are assumed to be 85 bp.[10] However, liquidity costs are very different for the two subgroups. The indirect shareholders' liquidity cost is 126 bp while the direct shareholders' liquidity cost is 62 bp. The resulting 65 bp liquidity difference is large in a relative sense—it is slightly larger than the direct shareholders' liquidity cost. It is also large in an absolute sense—it is nearly three-quarters of the 91 bp average asset-awaited expense ratio of domestic equity mutual funds in 2000.[11] These results suggest that shareholders' predicted demand for liquidity is not proportional to the liquidity costs they bear in the fund.

Since the liquidity cost imposed on the fund by the indirect shareholders is higher than that imposed by the direct shareholders, it appears that the fund's return would have been higher had the fund excluded the costly indirect shareholders. The direct shareholders' estimated pooling cost is 51 bp, suggesting that an otherwise identical direct-only fund would outperform the pooled fund by 51 bp annually.

The other rows of Panel A show the liquidity costs, pooling costs, and liquidity differences for the other three management styles identified by Jones and

[9] If indirect and direct shareholders really trade shares differently, the fund may adapt its response to shareholder flows. In the gross-flow model, for example, this suggests that $\beta_{\text{direct}} \neq \beta_{\text{indirect}}$. Data limitations prevent an exploration of this issue; the estimated costs assume that the fund's response does not vary by shareholder type.

[10] Although liquidity cost levels are not the focus of this section, it is worth noting that Panel A's estimates are consistent with other research. In particular, estimates in Edelen (1999) suggest that liquidity provision costs the fund 90 bp = 0.0174(0.24 + 0.28) annually. This level of liquidity costs obtains when transaction costs are 68 bp.

[11] January 2001 Morningstar PrincipiaPro; author's calculations.

Table VIII

Liquidity Difference

This table presents, in basis points, the average liquidity cost (LC), pooling cost (PC), and liquidity difference (LD) by shareholder type. LC is the geometric difference between the simulated pre-flow fund return and the actual post-flow fund return. It is annualized and averaged across all funds in the family. PC is the difference between the fund's liquidity cost and that of either indirect or direct shareholders. LD is the difference between LCs for the indirect and direct shareholders. Panel A assumes the fund trades 70% of gross shareholder flows. The fund's transaction costs are 152 bp (momentum manager), 85 bp (average manager), 57 bp (index manager), or 50 bp (value manager). Panel B assumes the fund trades 100% of net shareholder flows over accumulation periods that range from 1 to 20 trading days. Trading costs are 85 bp.

	Shareholder Type					
	All	Indirect		Direct		
	LC	LC	PC	LC	PC	LD
A. Gross Shareholder Flows						
Trading Costs (in Basis Points)						
152	195	216	−22	108	86	108
85	112	126	−14	62	51	65
57	76	86	−10	42	35	45
50	67	76	−9	37	31	39
B. Net Shareholder Flows						
Accumulation Period (in Trading Days)						
1	110	138	−28	63	47	75
2	105	131	−26	60	45	71
5	99	121	−22	58	41	63
10	99	115	−16	58	41	57
15	100	109	−9	60	40	49
20	103	106	−3	61	42	45

Lipson (2001). For example, the results show that liquidity differences between indirect and direct shareholders in momentum, index, and value managed funds are 108, 45, and 39 bp, respectively.

The net-flow model is presented in Panel B of Table VIII. The pre-flow fund returns are calculated by adding transaction costs for 100% of shareholder net flows to the post-flow fund returns over accumulation periods ranging from 1 to 20 trading days.[12] The model assumes that trading costs are 85 bp. The results are qualitatively similar to those in Panel A.

[12] It is unlikely the fund accumulates flows over periods longer than one calendar month. First, the transfer agent sends the fund daily or twice-daily updates on gross shareholder flows. This suggests that the fund considers daily whether or not to trade shareholder flow. Second, the ratio of monthly net flows to monthly fund assets averages 7%. Tracking error would be significant if the fund had longer accumulation periods.

B. Robustness Checks

The liquidity cost model has four inputs. Table VIII shows that the results are driven by neither assumptions of trading costs (Jones and Lipson (2001)) nor assumptions about the fund's response to shareholder flows (Edelen (1999)). It is worthwhile to consider briefly the model's sensitivities to the other two inputs because they are unique to the mutual fund family that provided the data.

The first family-specific input is daily fund returns. The gross-flow model is reestimated with three sequences of random returns. The first sequence is normally distributed with the same mean and standard deviation as the family's returns. The second sequence is normally distributed with mean and standard deviation equal to that of the value-weighted stock market between 1994 and 2000, the sample period. The third sequence is normally distributed with mean and standard deviation equal to that of the value-weighted stock market between 1962 and 2000. In each case, the unreported results are similar to those presented in Table VIII. This suggests that the family's particular sequence of returns is not driving the results.

The other family-specific input is shareholder flows. One concern is that the positive liquidity difference is primarily driven by fund growth differences and not investment horizon heterogeneity: a higher growth rate of indirect assets would mechanically impose higher aggregate transaction costs even if investment horizons were identical. (Recall that the liquidity difference statistic controls for common growth rates.) If the differential growth rate were predictable, the liquidity difference statistic would remain meaningful in the sense that it captures a predictable wealth transfer. Even so, an adjustment for differential growth seems warranted. The first robustness check reestimates the gross-flow model with scaled shareholder flows:

$$\text{scaled type inflows}_t = \text{raw type inflows}_t \times \left| \frac{\text{average fund inflows}}{\text{average type inflows}} \right|,$$

$$\text{scaled type outflows}_t = \text{raw type outflows}_t \times \left| \frac{\text{average fund outflows}}{\text{average type outflows}} \right|,$$

where "type" is either indirect or direct. The scaled liquidity difference for momentum, average, index, and value managers is 48, 29, 21, and 18 bp, respectively. This shows that even when the indirect and direct shareholders impose exactly the same level of costly aggregate flows on the fund, there are still differences in the fund's corresponding liquidity costs that are directly attributable to investment horizon differences.

The second adjustment for growth differences focuses on outflows from existing shareholders while ignoring inflows from new shareholders.[13] This is operationalized in the gross-flow model by assuming the fund trades gross shareholder redemptions at 85 bp and gross purchases at 0 bp. In this model,

[13] The referee kindly suggested this approach.

the liquidity difference for momentum, average, index, and value managers is 55, 31, 21, and 19 bp, respectively. This shows that the results are not driven by inflows.

Out-of-sample tests for the robustness of the main results to shareholder flows are not possible because shareholder data from other mutual fund families are not available. Future research is needed to determine whether this family's mixture of shareholders is representative of mixtures found in other open-end mutual fund families. The item of interest is whether predictable liquidity differences (not simply liquidity costs) are economically large.

In summary, this section explores liquidity costs generated and borne by fund shareholders, focusing on two easy-to-identify groups of shareholders (namely, indirect and direct) that have different predicted liquidity needs. Even though these shareholders bear the same daily costs in the pooled fund, their annual liquidity needs differ in the basic, scaled, and redemption-only models by 65, 29, and 31 bp, respectively. In other words, these shareholders' predicted demand for liquidity is not proportional to the costs they bear in the fund. In the basic model, for example, direct shareholders annually pay 51 bp more for liquidity in the pooled fund than they would pay in an otherwise identical fund that excluded indirect shareholders.

The reported liquidity and pooling cost estimates are expected to be lower bounds of the true costs for three reasons. First, the estimation process excludes liquidity cost components such as tracking error and taxes. Second, the estimates are based on two very coarse types of shareholders—indirect and direct—that are categorized by just one ex ante characteristic. It is easy to segment further the shareholders into groups that have larger costs. Third, the estimates ignore shareholders' private information about their liquidity needs.

VI. Discussion

The first part of this study documents that shareholders in one mutual fund family have predictably different liquidity needs. The second part shows that the shareholders' predicted demand for liquidity is not proportional to the costs they pay in the fund. This raises three questions about the predictable wealth transfer: can it be reduced or eliminated, is it offset by other costs, and does it exist in other mutual fund families? These are discussed in turn.

A. Reducing the Wealth Transfer

The wealth transfer could be reduced if fees were charged that either discouraged costly, short-term investments or compensated the fund for its transaction costs. One common fee is the load. It is generally imposed by the fund at the time of share purchase (front-end load), but it is occasionally imposed at the time of redemption (back-end load). Either way, the load introduces, in effect, a bid-ask spread in the pricing of fund shares that makes trading more costly for shareholders. Nevertheless, loads may not have a large effect on fund-level investment horizons because they are generally family (and not fund) specific. For example, a new shareholder will pay a load on the initial investment in an equity fund but not on a subsequent exchange to another equity fund in the same family. Moreover, loads are paid to the broker and do not compensate the fund for its transaction costs.

Some funds charge purchase or redemption fees that are paid directly to the fund's portfolio—not to the broker—to offset the fund's trading expenses. Purchase fees are imposed on all share purchases; redemption fees are only levied against shares sold within a specified time after purchase, typically fewer than three months. These fees should reduce the transaction cost externality because they may discourage very short holding periods and they compensate the fund for at least some of the shareholder-imposed transaction costs. Even when combined, however, purchase and redemption fees will not eliminate the wealth transfer because shareholders whose investment horizon exceeds the redemption-fee window do not pay the redemption fee. The percent of no-load equity funds that impose redemption fees increased from 3.5% in 1999 to 7.7% in 2001.[14]

In theory, the transaction cost externality would be eliminated were every shareholder charged, irrespective of expected or realized holding periods, a properly calibrated transaction fee for all purchases and redemptions that is paid directly to the fund's portfolio. However, transaction fees would not address the other four liquidity costs (such as taxes) highlighted in Section V.

B. Offsetting Costs

It may be naïve to focus on just one cost of investing in mutual funds when it might be offset by other factors. For example, there may be economies of scale in the industry that reduce expense ratios as funds grow. If the expense ratio reduction were large enough, short-term shareholders would provide a net benefit to long-term shareholders when they join the fund.

Another unexplored cost comes from servicing shareholders. The Vanguard Group says that shareholder servicing costs are approximately $40 per account per year.[15] A fund's expense ratio would have to include 100 bp on a $4,000 account to recover just these expenses. Even if long-term accounts subsidize short-term accounts' trading costs, the wealth transfer likely runs the other way for shareholder servicing costs because long-term accounts tend to carry small balances.

C. Generalizability

The predictable wealth transfers documented within one mutual fund family are more interesting if they are indicative of economically significant wealth

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

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

transfers in other mutual fund families. The remainder of this section discusses, therefore, whether shareholders in other funds have predictably heterogeneous liquidity needs and whether other funds have different liquidity costs.

The transfer agent that provided data for this study anecdotally reports that this family's shareholders are more stable (longer term) than those in other families it services. Beyond this, it is impossible to say whether shareholders in other mutual funds have liquidity needs that are different from those documented in this study. Nevertheless, it is reasonable to conjecture that other funds contain shareholders with heterogeneous expected investment horizons because fund shares are frequently sold through multiple distribution channels. For example, some funds are distributed through both small 401(k) retirement plans that do not provide many alternative investments and large mutual fund supermarkets that do. Even if a particular fund were distributed through a single channel—such as a proprietary pension fund for a large employer—investors may have unequal access to account options such as Internet trading privileges that affect or are affected by their switching costs. For example, white-collar managers with computers on their desks and Internet trading privileges trade shares more cheaply than blue-collar workers on the factory floor without computer access trade. Although this sort of within-fund heterogeneity is expected to lead to measurable ex ante investment horizon differences within other mutual funds, it is unclear whether the average difference will be smaller or larger than those documented in this study.

The reported liquidity cost estimates may be representative of the average costs found in other funds because they are grounded in general results from Edelen (1999) and Jones and Lipson (2001). On a fund-by-fund basis, however, it is clear that shareholder-imposed costs vary. A few examples illustrate which funds might have high liquidity costs. First, shareholder flows increase costs if they are volatile (as might be the case in mutual fund supermarkets) or if they are large relative to the size of the fund (as might be the case in a small fund that has gotten positive press). Second, the fund accentuates the negative effect of shareholder flow if it accumulates flows over short periods (as might be the case in a fund that is sensitive to deviations from its preferred portfolio weights). Last, the fund's costs will increase if the fund aggressively trades assets (as might be the case in a momentum fund) or if the fund trades relatively illiquid assets (as might be the case in a small-cap fund).

VII. Conclusion

A new friction in trading mutual fund shares is modeled: switching costs. Shareholders with low switching costs sell fund shares early and impose high costs on the fund. Similarly, shareholders with high switching costs sell fund shares late and impose low costs on the fund. The net effect is a wealth transfer from the long-term shareholders to the short-term shareholders. Two implications of the model are tested in this study: switching-cost differences are associated with investment horizon differences and investment horizon differences cause a wealth transfer among shareholders.

The first half of the analysis shows that the fund can predict shareholders' investment horizons using data it collects in the ordinary course of business that proxy for shareholders' switching costs. For example, older investors, larger accounts, and Fund/SERV accounts have shorter investment horizons while automatic investment plan, tax-advantaged, and employee accounts have longer investment horizons. Accounts with recent purchases are less likely than others to close; those with recent redemptions are more likely than others to close. Accounts with distant purchases or redemptions are uniformly more likely than others to close, but the distant-purchase effect is economically small. The evidence also shows that even though various measures of fund returns are statistically significant predictors of investment horizons, the economic effects are surprisingly small.

The duration analysis also shows that residents of the investment advisor's home state have predictably longer investment horizons than nonresident shareholders have. This large effect is interesting in light of the geography literature that documents investors' preference for local assets (e.g., French and Poterba (1991) and Coval and Moskowitz (1999)).

The second half of the analysis splits the shareholders into two groups based on the channel through which they purchased fund shares and measures the liquidity demands of each group. Estimates show that even though the indirect and direct shareholders bear the same liquidity costs in the pooled fund, their liquidity needs differ significantly. In other words, the shareholders' predicted demand for liquidity is not proportional to the costs they bear in the fund, suggesting that the fund systematically misprices its shares. For example, expected long-term shareholders pay more for liquidity in the pooled fund than they would pay if the fund excluded expected short-term shareholders.

The levels of the reported cost estimates are large relative to explicit mutual fund expenses such as management fees and 12b-1 fees. Moreover, the simulations likely underestimate the true pooling externalities for two reasons. First, this study only looks at one dimension of the problem, ignoring, for example, the effects of taxes and tracking error. Second, the cost estimates are based on only two ex ante types of shareholders, neglecting not only the shareholders' private information about their liquidity needs, but also many other observable variables that predict account closures.

An important result of this article is that mutual funds are not Diamond-Dybvig liquidity providers that offer an equitable insurance benefit to all shareholders. Because predictable investment horizon differences result in economically significant wealth transfers, the fund's proportionate allocation of costs allows expected short-term shareholders to impose some of their costs on the long-term shareholder. Future research could pursue this unexpected wealth transfer from the supply side: Why do mutual funds pool predictably heterogeneous shareholders? Alternatively, the demand side could be explored: Why do long-term investors commit resources to mutual funds that engage in such pooling? Although these questions are narrowly focused on mutual funds, parallel questions may apply to banks, mortgage-backed securities, and other financial structures. Answers to these questions may add important insights into the perennial question of how mutual funds and related financial intermediaries add value.

Appendix

This appendix presents a model of mutual fund investing where the motivation for trade across funds 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 trading 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, prohibiting trade is Pareto optimal.

The model is based on the standard infinite-horizon 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 $t+1$. If the investor chooses instead 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 A1.

Investors pay a dissipative switching fee $s \in \{s_l, s_h\}$ to exchange funds at midlife.[16] Assume 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 trading costs $c \in (0, \frac{1}{2}R)$ 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 nonredeemers.

Theorem 1 shows that if switching costs are large enough for the high-cost type, there is a separating equilibrium where the low-switching-cost (LSC) investors switch funds at midlife and the high-switching-cost (HSC) investors do

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

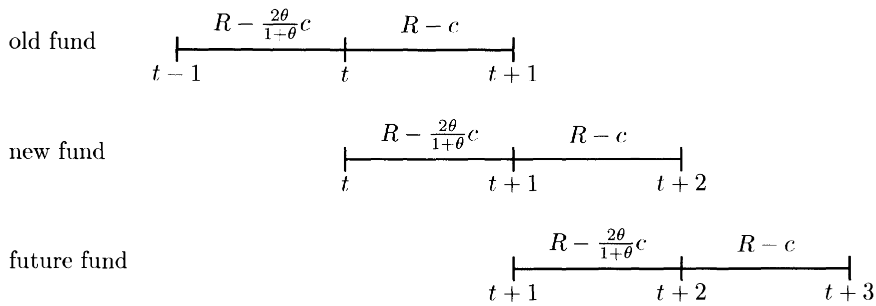

Figure A1. Investment alternatives at time *t*. In an infinite-horizon OLG model where both funds and investors live for two periods, investors choose between two funds each period. Trading costs c imposed by shareholder redemptions reduce the otherwise fixed return R of the funds. In the unique separating 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 than the second-period fund return $R-c$. The traders' return $(R - \frac{2\theta}{1+\theta}c)^2$ is, therefore, greater than the nontraders' return $(R - \frac{2\theta}{1+\theta}c)(R-c)$.

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 at least as large as the trading gain.

THEOREM 1 (Separating Equilibrium): *If $\theta \in (0,1)$ and $s_h \geq \frac{1-\theta}{1+\theta}c$, 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 only one period, switching 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 $(R - \frac{2\theta}{1+\theta}c)^2$ is greater than the non-traders' return $(R - \frac{2\theta}{(1+\theta)}c)(R-c)$.*

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 $(R - \frac{2\theta}{1+\theta}c)^2$ 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 $(R - \frac{2\theta}{1+\theta}c)(R-c)$ while the best deviating return is $(R - \frac{2\theta}{1+\theta}c - s_h)(R - \frac{2\theta}{1+\theta}c)$. Thus, they do not deviate if $s_h \geq \frac{1-\theta}{1+\theta}c$. Q.E.D.

Theorem 2 demonstrates there is no pooling equilibrium in this economy.

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 $s_h > 0$. Q.E.D.

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 as follows:

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

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

The difference between the benefits and costs is

$$\text{aggregate benefit} - \text{aggregate cost} = \frac{c\theta(2c(1+\theta^2) - R(1+\theta)^2)}{(1+\theta)^2} < 0,$$

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

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 were large enough, 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 $(R - \frac{2\theta}{1+\theta}c)^2 \approx (R-c)^2$ 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 $\theta \in [V,1)$, a midlife trading tax $p \geq c$ is Pareto optimal. If*

$\theta \in (0, V)$, a midlife trading tax $p \geq c$ benefits the HSC investors but hurts the LSC investors.

Proof: If no one trades, all investors receive the two-period return $R(R - c)$. The LSC investors' return from deviating is $(R - p)R$, so they do not deviate. The HSC investors' return from deviating is even lower so they, too, do not deviate. All that remains to be shown is that this new pooling equilibrium is 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, equivalently, $f(\theta) = (R(R - c) - (R - \frac{2\theta}{1+\theta}c)^2)(1 + \theta)^2 \geq 0$. 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 nonnegative when $\theta \in [V, 1)$. Q.E.D.

REFERENCES

Allen, Franklin, and Douglas Gale, 1998, Optimal financial crises, *Journal of Finance* 53, 1245–1284.

Ameriks, John, and Stephen P. Zeldes, 2001, How do household portfolio shares vary with age?, Working paper, Columbia University.

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, Noel, Gavan Fitzsimons, and Russ 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 financial intermediaries setting security prices: The mutual fund wild card option, *Journal of Finance* 56, 2209–2236.

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.

Deng, Yongheng, John M. Quigley, and Robert Van Order, 2000, Mortgage terminations, heterogeneity, and the excercise of mortgage options, *Econometrica* 68, 275–307.

Diamond, Douglas W., 1997, Liquidity, banks, and markets, *Journal of Political Economy* 105, 928–956.

Diamond, Douglas W., and Philip H. Dybvig, 1983, Bank runs, deposit insurance, and liquidity, *Journal of Political Economy* 91, 401–419.

Dickson, Joel M., John B. Shoven, and Clemens Sialm, 2000, Tax externalities of equity mutual funds, *National Tax Journal* 53, 607–628.

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

Falkenstein, Eric G., 1996, Preferences for stock characteristics as revealed by mutual fund portfolio holdings, *Journal of Finance* 51, 111–135.

Flannery, Mark J., 1994, Debt maturity and the deadweight cost of leverage: Optimally financing banking firms, *American Economic Review* 84, 320–331.

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, 2002, Daily momentum and contrarian behavior of index fund investors, *Journal of Financial and Quantitative Analysis* 37, 375–389.

Greene, Jason T., and Charles W. Hodges, 2002, The dilution impact of daily fund flows on open-end mutual funds, *Journal of Financial Economics* 65, 131–158.

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* 20:3, 11–23.

Johnson, Woodrow T., 2002, Shareholder heterogeneity in mutual funds, unpublished doctoral thesis, Columbia University.

Jones, Charles M., and Marc L. Lipson, 2001, Sixteenths: Direct evidence on institutional execution costs, *Journal of Financial Economics* 59, 253–278.

Kashyap, Anil K., Raghuram Rajan, and Jeremy C. Stein, 2002, Banks as liquidity providers: An explanation for the coexistence of lending and deposit-taking, *Journal of Finance* 57, 33–73.

Lakonishok, Josef, and Edwin Maberly, 1990, The weekend effect: Trading patterns of individual and institutional investors, *Journal of Finance* 45, 231–243.

Nanda, Vikram, M. P. Narayanan, and Vincent A. Warther, 2000, Liquidity, investment ability, and mutual fund structure, *Journal of Financial Economics* 57, 417–443.

Sirri, Erik R., and Peter Tufano, 1998, Costly search and mutual fund flows, *Journal of Finance* 53, 1589–1622.

Stanton, Richard, 1995, Rational prepayment and the valuation of mortgage-backed securities, *Review of Financial Studies* 8, 677–708.

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.