A Theory of Mutual Funds: Optimal Fund Objectives and Industry Organization

by

Harry Mamaysky†

and

Matthew Spiegel‡

April 17, 2001

Comments Welcome

†Yale School of Management, P.O. Box 208200, New Haven CT 06520-8200. Phone: 203-436-0649, Fax: 203-436-0630, email: **harry.mamaysky@yale.edu**, web page: **http://som.yale.edu/~hm68**.

‡Yale School of Management, P.O. Box 208200, New Haven CT 06520-8200. Phone: 203-432-6017, Fax: 203-432-8931 email: **matthew.spiegel@yale.edu**, web page: **http//som.yale.edu/~spiegel**.

A Theory of Mutual Funds:
Optimal Fund Objectives and Industry Organization

Abstract

This paper presents a model in which investors cannot remain in the market to trade at all times. As a result they have an incentive to set up trading firms (financial market intermediaries) to take over their portfolio while they engage in other activities. Previous research has assumed that such firms act like individuals endowed with a utility function. Here, as in reality, they are firms that simply take orders from their investors. From this setting emerges a theory of mutual funds. Many of the model's results adhere to past empirical findings, while others are tested and verified here. Among the paper's conclusions: (1) Mutual funds outnumber traded securities. (2) Mutual funds have a number of objective functions which lead to observably different trading styles. (3) Adding mutual funds to the economy makes stock prices more volatile. (4) Adding mutual funds does not impact the risk premium. (5) There are incentives for funds that trade on market fundamentals to "spread out" in strategy space. This produces divergent trading styles that investors who are not currently active in the market can use to better hedge out the risks they face. Several empirical tests are shown to support the primary hypotheses.

JEL Classification: G20, G12

Mutual funds, investment houses, and other financial market intermediaries (FMIs) have not appeared out of whole cloth. Rather investors set them up to meet some particular demand. Recently, a number of empirical papers have noted that the institutions produced by this process exhibit a wide range of behaviors, many of which are difficult to reconcile with the existing theory. For example, Gruber (1996) notes that while most models predict that investors will create only a small number of passively managed funds in reality there exist thousands and most of these are actively managed. Moreover, as Sharpe (1992) and Brown and Goetzmann (1997) demonstrate these funds encompass a fairly wide array of trading strategies and these strategies appear to change over time. This paper provides a model that helps to explain these empirical patterns. At the same time it produces a number of new hypotheses, several of which are tested here and found to hold within the data.

This paper takes a first principles approach to the issue of financial market intermediation. It does so by using as primitives the population, their preferences, and trading technologies. Members of the population are then free to set up FMIs as they see fit and give them whatever instructions they like. From this setting emerges a distinct theory of such institutions and their behavior.[1]

The plethora of managed funds noted by Gruber (1996) arise naturally within the model developed here. Essentially, a multiplicity of funds and other financial firms arise to cater to a population of individuals with divergent equilibrium trading strategies. While each investor

---

[1] Note that the analysis in this paper differs from the traditional principal-agent paradigm (see Ross (1973)). Rather than focus on how to induce an agent whose actions are unobservable to act appropriately on behalf of a principal, we analyze what actions the principal would ideally want the agent to take. Our assumption is that the actions of the agent are contractible, and hence the agent may be induced to carry out exactly the principal's orders. The question of what those orders should be is itself far from trivial in our setup.

might like to see his or her optimal trading strategy carried out this is economically infeasible unless FMIs can be produced and staffed at no cost. A second best solution is to construct a small number of intermediaries each of which trades along what can be thought of as unique strategy basis. Investors then carry out their preferred trading strategies by selectively buying shares in these different fund types.[2]

Note the distinction here between trading strategies and securities: funds offer agents particular trading strategies, rather than simply committing to hold some fixed portfolio of securities. Since the number of desired trading strategies may be quite large (even in a two security world), the number of funds may not be closely related to the number of traded securities.

In most microstructure models FMIs are modeled as utility maximizing individuals. These firms/people are then assumed to be either risk neutral if the paper follows along the lines of Kyle (1985), or risk averse if it follows along the lines of Ho and Stoll (1983), or Campbell and Kyle (1993). But, of course, financial firms are corporations that simply carry out their investors instructions, and this is how they are modeled here. As a result, the equilibrium comparative statics and thus this paper's empirical predictions differ from those produced by prior models. For example, if trading firms are risk averse individuals then additional firms reduce price risk and the equilibrium discount rate. In our setting, neither occurs. The creation of additional firms increases price risk and but has *no* impact on the expected discount rate.

---

[2] While people generally think of mutual funds as stand ins that trade on behalf of investors, other institutions do the same. When somebody purchases a share of Merrill Lynch they thereby allow Merrill Lynch to trade on their behalf. Thus, paper's use of the word fund is typically meant to encompass other FMIs as well.

Extending the model to allow for multiple risky assets produces a theory of fund families. Each family may be thought of as having a research department which specializes in collecting a particular piece of information about the state of the economy. The family then offers to trade portfolios of securities based on the realization of the information which they collect. A fund family will offer to trade portfolios which are useful to the subset of the population which is concerned with the economic information to which that fund family has access. The result of this analysis is a mutual fund theorem that can be used to explain both why fund families exist, and why investors frequently invest in several funds both within and across such families.[3]

The paper also seeks to test some of the new empirical predictions generated from the model. As noted above investors seek funds that help complete the strategy space. Thus, fund families should attempt to spread out in strategy space as quickly as possible. The data confirm that this is what they do. For example, when a fund family with only a few funds introduces a new fund it will typically use a strategy that places it in a different Morningstar category than its older siblings.[4] Another test of the model comes from an examination of the asset allocation decisions made by funds. According to the model investors should value funds that help them to time their entry into and out of particular parts of the market. In fact, there exist over 1,700 such funds as evidenced by their having moved of at least 20% of the portfolio into and out of stocks during their life.

---

[3] Mathematically, it turns out that the dimensionality of an economy with multiple securities is not the number of factors which drive security payoffs, but rather the number of factors which drive endowment shocks. It is this dimensionality that drives investors to demand a wide array of fund families each of which has a number of individual fund offerings.

[4] For very large fund families this becomes impossible as Morningstar may have fewer categories than the family has funds.

Since the model predicts that investors value a range of trading strategies this should show up within the data as time varying betas with particular patterns within and across fund families. Within a fund family the model predicts that these time varying betas should be highly correlated, as the funds share a common signal. Across families the correlations should be lower. The data support both predictions. However, more subtly the model also predicts that the first fund introduced by a family should follow a relatively unique strategy. Thereafter the correlations should increase as a family fills in the strategy space and is thus forced to introduce products closer to those already offered elsewhere. The analysis finds exactly this pattern. Funds introduced earlier in a family's life will exhibit lower correlations than those introduced later. This pattern is monotonic across pairings. That is pairing the two oldest funds across families produces (on average) a lower correlation than pairing the two second oldest funds, or the oldest from one family with the second oldest from another.

The paper is organized as follows. Section 1 presents the model and the following two subsections lay out its structure. Subsection 1.3 examines the problem of setting up a trading firm and the equilibrium price. Subsection 1.4 looks at stock price volatility. Section 2 examines entry by new funds, the type of information they will collect, and the trading rules they will use. Section 3 extends the analysis to the case of an economy with multiple risky securities. Section 4 presents empirical evidence supporting the model. Section 5 relates the current paper to the existing literature, and Section 6 contains the paper's conclusions.

1 Single Risky Security Model

1.1 Underlying Securities

The model examines a three period economy. Within this setting there exists one risky and one risk free asset. The risky asset pays off D per share in period three. The market's prior is that D is normally distributed with a mean of zero, and a variance of σ_D^2. For simplicity assume that D is independent of all other random variables. In period three the risk free asset pays one dollar per dollar invested. This simply normalizes the risk free rate to zero.

1.2 Agents

In period one a continuum of traders arrive with a mass equal to one, an aggregate endowment of X_1^0 shares of the risky asset, and B_1^0 units of the risk free asset. To facilitate the paper's exposition these traders will be referred to as period one traders. In period two, period one trader i then receives an endowment shock of the risky asset equal to $n(i)$. Each shock $n(i)$ is normally distributed with mean zero and variance σ_n^2. One can think of this shock as representing human capital. In period two the trader's employment efforts cost him $n(i)P_2$ in capital but then allow him to obtain $n(i)D$ in period three.[5] The model assumes that individuals are arrayed in characteristic space and that the distance between people determines the correlation between their shocks. Thus, two traders at nearly the same location will have nearly perfectly correlated endowment shocks while those far away will receive nearly uncorrelated shocks. These shocks are not publically observable. Per the standard assumption, both these and all other investors have negative exponential utility functions over terminal (period three) wealth

---

[5] Charging the trader P_2 to obtain the shares of human capital is done only to reduce the model's algebraic complexity. It has no qualitative impact upon the paper's results.

with risk aversion parameter γ.

While the individuals arriving in period one wish to invest in the market they also wish to engage in other activities (i.e. they have a life). The model assumes that in order to carry out these other activities (e.g. full time employment, raise a family, etc.) traders cannot spend all of their time monitoring the market and issuing buy and sell orders. Thus, while they can trade at will in period one they must trade via designated agents in period two. These period two agents take the form of corporations whose purpose is simply to trade in period two on their owner's behalf. These trading firms can be given whatever objective function their creators wish. Equivalently, one can think of their existing a population of entrepreneurs that have the ability to form trading firms and seek period one traders to invest in their companies.

In period two a second set of investors arrive (the "period two traders"), this time with an aggregate endowment of X_2^0 shares of the risky asset and B_2^0 units of the risk free asset. Their share endowment has a mean of $\overline{X}_2^0$ and variance $\sigma_{X_2^0}^2$. Since these investors have just arrived in the market they trade directly on their own behalf. The model assumes there are a continuum of such traders with mass equal to one.

1.3 Trade and the Formation of Trading Firms

Forming a trading firm to buy and sell securities can only be done by expending some fixed cost which implies that only a finite number of them will be created. For now fix the number of such firms at J+2. Since these firms are nothing more than corporations set up by investors they are not endowed with any nonpublic information. However, each firm j can acquire one and only one signal e_j. For notational reasons it is convenient to group the signals into two types. One is an uninformative signal or one about publically observable information, denoted by $e_j(u)$. It is assumed that such signals are free. This simply allows people to set up firms that will either purchase or sell shares on the basis of purely public information. The second signal type $e_j(\ell)$ tells the firm what shock $n(\ell)$ was received by the investor at location ℓ. Such signals may be costly.

Summary 1: A Comparison with the Assumptions In a Standard Model of Financial Market Intermediation

Assumption	Standard Model	First Principle's Model
Security Payoffs	Risky asset with normally distributed payoff. Risk free asset.	Same
Investor Entry and Exit Dates	Agents enter for one period, trade, and leave.	Same
Investor Utility Functions	Negative exponential.	Same
Financial Market Intermediary (FMI) Entry and Exit Dates	FMIs can trade in both periods one and two.	FMIs are formed in period one and trade in period two.
FMI Objective Function	Assumed risk neutral or negative exponential utility function.	Determined by the firm's investors.

1.3.1 Problem of the Period Two Traders

In period two market clearing occurs when the aggregate demand of the trading firms (Y) plus the aggregate demand of the period two traders (X_2) equals X_2^0. Let P_2 represent the period two market clearing price. Then period two trader i will seek to maximize his expected utility over final wealth $X_2(i)D + B_2(i)$ subject to the budget constraint

$$[X_2(i) - X_2^0(i)]P_2 + B_2(i) - B_2^0(i) = 0. \qquad \textbf{(1)}$$

From the standard arguments this yields the following equilibrium demand $X_2(i)$,

$$X_2(i) = -\frac{P_2}{\gamma\sigma_D^2}. \qquad \textbf{(2)}$$

1.3.2 Problem of the Period One Traders

At the start of the model the period one traders arrive with their initial endowment of shares which they can trade among themselves. At this point they can also set up trading firms (FMIs) to carry out their wishes in period two. Instructions of the form “maximize my utility” are not possible. Naturally, the firms cannot observe each trader’s life in sufficient detail to know exactly what trades should or should not be executed. Instead firms carry out orders that are effectively given in advance. For example, a firm may announce that it is a “growth fund” and will thus buy shares when it appears a company is growing and sell when it is not. Investors then tell the manager how many shares to purchase on their behalf, in response to whatever signal is acquired. They do this by buying a certain number of shares of the trading firm in advance.

Based on the above setting, in period one each period one trader must decide on what position to take in both the corporation and each FMI. What complicates matters is that the FMIs will then trade in period two which then impacts the investor’s portfolio in a way that the investor cannot directly control. Thus, investors face the problem of selecting a portfolio today, that will change randomly and in proportion to their purchases next period. Fortunately, there is a mathematical shortcut through this problem.

Within the model one can view the above process as taking place via a sequence of events in which first each FMI announces what signal $e_j(\ell)$ it will purchase. Second, investor i then issues orders of the form, “in the event the signal $e_j(\ell)$ is observed, purchase $f_j(e_j(\ell),i)$ shares on my behalf.” Within this formulation f_j can be linear, but in principle can take on other forms as well. The important point is that f_j is a demand conditional on the observed signal $e_j(\ell)$. Note that *a priori* it is not obvious that the solution to this problem will solve the investor’s actual problem. The problem of issuing orders of the form $f_j(e_j(\ell),i)$ provides the investor with a richer strategy space than is actually available. For example, investors can determine how many shares of Merrill Lynch or Vanguard's Asset Allocation Fund they will buy, but cannot otherwise issue orders for particular purchases. This limits the actual set of strategies to linear rules. It thus turns out that the solution to the problem of selecting $f_j(e_j(\ell),i)$ corresponds to the solution to the restricted problem because a linear rule will turn out to be the optimal form for $f_j(e_j(\ell),i)$ to take.

Based upon the above notation and the discussion in Section 1.2 trader i's final wealth equals

$$X_1(i)D + \sum_j \left[f_j(e_j(\ell),i)D + \hat{B}_j(i) \right] + B_1(i) + n(i)[D - P_2] \quad \textbf{(3)}$$

where $\hat{B}_j$ represents fund j's holdings of the risk free asset. There are now two budget constraints that must also be satisfied. The first is the standard one for the trader

$$[X_1(i) - X_1^0(i)]P_1 + B_1(i) + B_1^0 = 0 \quad \textbf{(4)}$$

and the second a budget constraint for each fund

$$f_j P_2 + \hat{B}_j = 0. \quad \textbf{(5)}$$

Plugging **(4)** and **(5)** into equation **(3)** produces

$$X_1(i)[D - P_1] + \left\{ n(i) + \sum_j f_j(e_j(\ell),i) \right\} [D - P_2] + X_1^0(i)P_1 + B_1^0(i) \quad \textbf{(6)}$$

for the trader's terminal wealth. Since the e_j terms represent normally distributed signals about the endowment shocks of traders at particular locations in characteristic space, one can write the endowment shock for trader *i* as

$$n(i) = \sum_j \alpha_j(i) e_j + \eta(i). \qquad \textbf{(7)}$$

The $\alpha_j(i)$ terms equal the loading on signal *j* used to calculate the expected value of *n*(*i*) given the correlation across signals and *n*(*i*). Due to the assumption of joint normality this leaves a residual endowment shock $\eta(i)$ that is itself normally distributed with mean zero, variance σ_η^2, and independent of the signals. After substituting **(7)** into **(6)**, standard solution techniques allow one to obtain the solution for the f_j.

Lemma 1*: Investor i sets* f_j *to solve*

$$X_1(i) + \sum_j \left(\alpha_j e_j + f_j\right) + \frac{1}{\gamma}\sigma_D^{-2} P_2 = 0 \qquad \textbf{(8)}$$

where two of the funds use uninformative signals with values of α *equal to zero.*

Proof: See Appendix for this and all other proofs.

Equation **(8)** yields each investor's instructions to the trading firm. Since these firms are effectively mutual funds that base their trades on the signal e_j that they receive the paper will often refer to them as such. Thus, an investor *i* orders funds one through *J* to purchase $\alpha_j e_j$ shares, given that the fund *j* has received signal $e_j(\ell)$. In addition the investor seeks two additional funds. One which can be called a "constant" fund that simply sells off the investor's purchases in period one and a second called a "price fund" that sells $P_2 / \gamma\sigma_D^2$ shares. These two funds correspond to a pair of signals with α values equal to zero. Since these two funds trade only on the basis of publically available information, the paper assumes from here on in that they always exist. This is simply a way of stating that it is sufficiently inexpensive to set up such funds that investors will always choose to do so.

The above discussion leads to the following *J*+2 fund separation theorem.

Theorem 1. *In period 2, with J signals available the population of investors that are absent from the market can have their demands satisfied via J+2 funds. Funds one through J offer to trade on a signal e_j, fund J+1 offers to buy or sell a fixed number of shares, and fund J+2 offers to trade based on the period two price of the stock.*

There are three important points to note. First, trading via the *J*+2 funds is a second best solution for the period one traders. Each trader would rather trade based upon his own endowment shock. However, given that the trader has to engage in other activities a second best solution for *all* traders can be found via the formation of *J*+2 funds. Second, the funds do not behave like individuals: risk neutral or otherwise. The first *J*+1 funds trade without regard to equilibrium prices. No individual trader, trading on his own account would act like this. However, funds are not individuals. They are simply robotic entities that carry out their prescribed instructions. Investors allow for this by using fund *J*+2 (the price fund) to adjust their final positions in response to the price. Third, even though there is only one underlying stock there naturally arise a large number of funds.

Some evidence regarding the above results can be found in Table 1. It shows that since 1996 the number of funds capable of investing in equities has exceeded the number of available stocks in the U.S., until today when there are approximately eleven funds for every eight stocks. If one adds in FMIs beyond mutual funds the ratio of FMIs to stocks becomes even more impressive. While puzzling within a standard CAPM framework, this empirical fact is fully compatible with conclusions found in Theorem 1. Simply put, investors want a wide range of funds because they want a wide range of trading strategies available for their own personal use while absent from the markets.

Another facet of the model born out in Table 1 is Theorem 1's statement that while investors may not need a wide variety of price sensitive funds, such funds are desired by most of the population. Within the traditional nomenclature, Theorem 1's price sensitive fund looks like an "asset allocation fund." After all, both fund types move an investor's portfolio into and out of stocks depending on whether equity prices are relatively low or high. In 1999 while the number of funds that invested only in equities (funds 1 through *J* in the model) equaled 1,996, there were also 1,775 asset allocation funds. That is there existed 1,775 funds with a track record of varying the fraction of their portfolio in equities by at least 20%. However, more tellingly, while investors placed 934,607 million dollars in the equity only funds they put 1,802,193 million into asset allocation funds, about twice as much. This is in line with the model's prediction. While investors like a variety of equity funds, they all want to invest within asset allocation funds too.

Another issue to note is that Theorem 1 provides the investor's viewpoint on how to optimally construct the set of financial market intermediaries in the economy. Because individual investors cannot afford to remain in the market at all times they seek agents that provide services that create a *strategy* basis for the general public. Investors then create their desired strategy vectors by buying the correct number of shares in each fund. In this way a relatively small number of funds (relative to the size of the population) serve to produce the varied investment strategies a diverse population demands.

Having solved for the demands of both the period two traders and the investment funds it is now possible to solve for the equilibrium period two price. Setting supply, X_2^0 equal to demand, $X_2 + \sum_j f_j$ yields the following equilibrium equation

$$X_2^0 - \int_{i_1} \left(X_1(i) + \sum_j \alpha_j e_j + P_2 / \gamma\sigma_D^2 \right) di_1 - P_2 / \gamma\sigma_D^2 = 0 \tag{9}$$

which produces an equilibrium price of

$$P_2 = -\gamma\sigma_D^2 \left\{ \int_{i_1} \left(X_1(i_1) + \sum_j \alpha_j e_j \right) di_1 + X_2^0 \right\}. \tag{10}$$

Notice that while the demands of the first *J*+1 funds are not price sensitive the price itself looks like the price that would arise if the period one traders had to trade on the set of signals e_j rather than their own personal endowment shocks $n_1(i)$. Another feature of the equilibrium price is that its expected value does not depend upon the number of trading firms established by the period one traders. By comparison, in a standard model with risk averse intermediaries, additional firms increase the expected price as they add to the pool of traders willing to absorb risk. This difference shows that modeling financial market intermediaries from first principles is more than just an exercise with little empirical or normative consequences. In reality, firms are not people. As such they cannot absorb risk, they can only pass it along to their shareholders. Thus, it should not come as a surprise that when one models investment firms as firms they do not impact the market risk premium.

1.4 Period Two Stock Price Volatility with Financial Market Intermediation

The fact that some investors trade via mutual funds only impacts price volatility through the $\int_{i_1} \sum_j \alpha_j(i) e_j di_1$ term in **(10)**. Conversely, the other terms in **(10)** remain unaffected by the number of mutual funds or the signals they may use. Therefore if one knows how the volatility of $\int_{i_1}\sum_{j}\alpha_j(i)e_j di_1$ changes with the total number of funds then one knows how price volatility changes with the total number of funds.

From **(10)** the part of the price variance (σ^2_{PJ}) influenced by the number of mutual funds equals

$$\sigma^2_{PJ} = E\left\{\left[\int_x\sum_j\alpha_j(x)e_j dx\right]\left[\int_y\sum_j\alpha_j(y)e_j dy\right]\right\} \tag{11}$$

where *x* and *y* are dummy variables of integration. The vector α can be found via the standard arguments and equals $\Sigma_{n(i),e}{}'\Sigma_e^{-1}$, where $\Sigma_{n(i),e}$ equals the 1×*J* covariance vector between the endowment shock to period one trader *i* and the set of *J* informative signals *e*. The Σ_e term equals the *J*×*J* variance-covariance matrix for the signals *e*. Substituting out for the α terms and using the fact that $E\left[e_j e_j'\right] = \Sigma_e$ produces

$$\sigma^2_{PJ} = \int_x\int_y \Sigma_{n(x),e}{}'\Sigma_e^{-1}\Sigma_{n(y),e}dxdy. \tag{12}$$

Based upon **(12)** one can derive the following result.

Theorem 2. *Price volatility increases with the number of mutual funds.*

Theorem 2 shows that contrary to what one might intuitively think, increasing the trading options available to people via third parties increases price volatility. In fact, the lowest price volatility level arises when the period one traders are unable to trade in period two via mutual funds that acquire costly signals. Rather than smooth out prices, the mutual funds increase volatility because they are trading on behalf of their owners. These owners do not create funds to provide services to the period two traders, but rather to themselves. This "selfishness" leads the owners to produce corporations whose trading pushes prices, on average, away from the mean. Compare this with the results from a model in which the intermediaries are modeled as entities with utility functions. In that case, price volatility declines as the pool of individuals for risk sharing increases. Once again modeling financial market intermediaries as corporations changes the conclusions drawn about how the economy functions.

1.5 The Period One Risk Discount and Financial Market Intermediation

Let $V(W_1,J)$ represent the expected utility of a period one trader given his initial wealth and given that there exist J funds that trade based upon informative signals. After plugging **(8)** into **(6)** this trader's terminal wealth equals

$$W_3(i) = X_1(i)(P_2 - P_1) + \left(\eta(i) - P_2/\gamma\sigma_D^2\right)(D - P_2) + X_1^0(i)P_1 + B_1^0(i) \tag{13}$$

where D, P_2, and $\eta(i)$ are unknown at time one. Taking the expectation of $e^{-\gamma W_3(i)}$ and denoting period one wealth as $W_1(i) = X_1(i)\,P_1 + B_1(i)$ produces

$$\begin{aligned} &V(W_1, J) = E\left[U(W_3(i))\right] = \\ &-\left[\left(1+\sigma_P^2\sigma_D^{-2}\right)\left(1-\gamma^2\sigma_D^2\sigma_\eta^2(i)\right)\right]^{-1/2} \exp\left\{-\gamma W_1 - \frac{\overline{P}_2^{\,2} + 2\gamma\overline{P}_2\sigma_D^2 X_1 - \gamma^2\sigma_D^2\sigma_P^2 X_1^{\,2}}{2(\sigma_D^2+\sigma_P^2)}\right\} \end{aligned} \tag{14}$$

where the expected value of P_2 equals $\overline{P}_2$, and its variance is given by σ_P^2. Note that in equation **(14)**, the two terms which depend on the number of funds J are $\sigma_\eta^2(i)$ and σ_P^2. In particular, the introduction of a new fund will decrease $\sigma_\eta^2(i)$ and will increase σ_P^2 (see

Theorem 2).

Instead of taking X_1 as given, consider how traders optimally choose their asset holdings in period one. The time one budget constraint is given by

$$W_1 = W_0 - X_1 P_1. \tag{15}$$

Taking the first order condition of $V(W_1; J)$ and solving for X_1 yields

$$X_1 = \frac{\overline{P}_2}{\gamma\sigma_P^2} - \frac{\sigma_D^2 + \sigma_P^2}{\gamma\sigma_D^2\sigma_P^2} P_1. \tag{16}$$

Notice that this demand does not depend on a given agent's residual endowment risk from period two (this is due to the assumption that agents pay the market price for their time 2 endowment shocks of shares). Hence the expression gives yields the aggregate demand in period one as a function of the period one price. Setting this demand equal to X_1^0 (the aggregate time one supply) the market clearing price equals

$$P_1 = \frac{\sigma_D^2}{\sigma_D^2 + \sigma_P^2}(\overline{P}_2 - \gamma\sigma_P^2 X_1^0). \tag{17}$$

The period one price contains what may be called the risk-weighted mean of the time two price, as well as a risk discount for the time one supply.

Note that the number of funds trading in period two affects the period one equilibrium price. This interaction occurs via the time two price volatility σ_P^2. When new funds are introduced the volatility of the time two price increases relative to the volatility of the time three dividend. This can either increase or decrease the period one price. The following lemma, which follows from simple differentiation, states the relevant result.

Lemma 2. *The period one price decreases (or equivalently the risk discount increases) as the number of period two funds increases if and only if*

$$\overline{P}_2 > -\gamma\sigma_D^2 X_1^0. \tag{18}$$

To understand this lemma, consider what happens in the limit as σ_P^2 goes to either zero or infinity. As $\sigma_P^2 \to 0$ then $P_1 \to \overline{P}_2$ since without any time two price volatility, buying the stock at time one is identical to buying the stock at time two. On the other hand, when $\sigma_P^2 \to \infty$ one finds that $P_1 \to -\gamma\sigma_D^2 X_1^0$, since an infinitely high period two price variance induces those holding shares today to hold them (in expectation) until time three when the dividend payout occurs. Hence the time one price is equivalent to the price in a one period economy. Moving from a low to a high price volatility simply moves the price from one extreme ($\overline{P}_2$) to the other ($-\gamma\sigma_D^2 X_1^0$).

It is interesting to note that the presence of a fund at time two may affect asset prices at time one. The introduction of a fund, by increasing period two price volatility, actually serves as a wedge between the time two and time one supply shocks. More funds at time two imply that the anticipated time two supply shock will have less of an affect on period one prices. This is somewhat counterintuitive: if we know there will be a supply shock tomorrow which will decrease tomorrow's price, then investors ought to push today's price close to tomorrow's price. However, if there is sufficient price volatility tomorrow, then even though they anticipate a lower (in expectation) price, their risk aversion will prevent them from selling short today. The interesting point, however, is that the anticipation of more funds tomorrow (which makes markets more complete, but also tends to increase price volatility) actually inhibits people from pushing today's price close to tomorrow's expected price.

1.6 Comparison to Traditional Framework

Again, compare the result in Lemma 2 with what one obtains from a framework in which the intermediaries are modeled as humans with utility functions. In such a model the addition of intermediaries increases the period one price, since there are additional people to share the market wide risk. Here, additional firms cannot add to the risk sharing capacity of the economy. As a result, the impact of additional firms has an ambiguous influence on the period one risk premium since it depends on the current risk-return tradeoffs already available.

Summary 2: A Comparison with the Equilibrium Properties In a Standard Model of Financial Market Intermediation

Result	Standard Model	First Principle's Model
Demands of the Period Two Traders	Typical mean variance demands.	Same.
Equilibrium Clearing Condition	Market supply equals the period one holdings of the FMIs and the period two trader endowment.	Market supply equals the period two trader endowment.
Demands of the FMIs in Period Two	Typical mean variance demands.	Three different types. Type 1 trades a constant amount. Type 2 trades based upon the price. Type 3 trades based upon a private signal, and ignores prices.
Equilibrium Prices	Equal to that derived from a representative agent holding the sum of the FMIs period one holdings, plus the endowment of the period two traders.	Equal to that derived from a representative agent holding the sum of the estimated endowment of the period one traders based upon the signals obtained by the FMIs, plus the endowment of the period two traders.
Price Volatility as the Number of FMIs Increase	Declines	Increases
Period One Price as the Number of FMIs Increase	Increases	Depends upon the model's parameters.

2 Social Welfare and the Entry of New Funds

Agents in the model need to give up a part of their time one wealth in order to finance a new period two fund. Thus, they will choose to do so only if the new fund makes them better off net of the fixed startup costs. On the plus side, introducing a fund increases the ability of each agent in the economy to hedge his time two endowment shock. It may therefore seem that social welfare always increases if an additional period two fund can be created at no cost. However this intuition is incorrect. While introducing a new fund expands the hedging opportunities available to the agents in the economy, it also increases the period two price volatility. The former effect is beneficial. The latter effect may not be. Which dominates determines the welfare implications of new fund entry.

2.1 Welfare Impact of a New Fund

Consider a situation where J funds with endowment signals already exist at time two. The decision problem of the J+1st fund is to choose a trading strategy which maximizes the following objective function

$$G_{J+1} = \int_i V(W_1(i); J+1)\, di - \int_i V(W_1(i); J)\, di. \tag{19}$$

This objective function has the obvious interpretation of maximizing social welfare from introducing the J+1st fund into an economy with only J funds. It is possible however that the maximum value of this function is negative. In this case, the J+1st fund will choose not to enter.

To gain some insight into how G_{J+1} behaves consider the effect of an additional fund on the time one welfare of an individual investor. Although not strictly true, assume for the moment that the number of funds J can be chosen from the positive real numbers (hence allowing for fractional funds). Differentiating $V(W_1(i);J)$ with respect to *J* yields

$$\frac{dV(W_1,J)}{dJ}=\frac{\partial V}{\partial\sigma_\eta^2}\frac{\partial\sigma_\eta^2}{\partial J}+\frac{\partial V}{\partial\sigma_P^2}\frac{\partial\sigma_P^2}{\partial J}. \qquad \textbf{(20)}$$

Clearly, $\sigma_{\eta(i)}^2$ must be decreasing in *J* since additional signals can only reduce residual endowment uncertainty. Furthermore, from Theorem 2, the period 2 price volatility is increasing in *J*. This leaves only the signs of $\partial V/\partial\sigma_{\eta(i)}^2$ and $\partial V/\partial\sigma_P^2$ unknown in equation **(20)**. Both of these are established in the following lemma.

Lemma 3. *The expected utility of a trader* $V(W_1(i);J)$ *is concave and decreasing in* σ_η^2*, i.e.* $\partial V/\partial\sigma_\eta^2<0$ *and* $\partial^2 V/\left[\partial\sigma_\eta^2\right]^2<0$. *Also a trader's expected utility increases in the price variance* $\partial V/\partial\sigma_P^2$ *if*

$$\sigma_P^2+\sigma_D^2\geq(\overline{P}_2+\gamma\,\sigma_D^2\,X_1(i))^2\,. \qquad \textbf{(21)}$$

If this condition holds for some J, then for fund I (with I > J) social welfare will be higher than it was with only J funds (i.e. $G_I > 0$ for I > J).

The last result in the lemma has a particularly nice economic interpretation. Consider the simplest case when $\overline{P}_2=0$ and $X_1(i)=0$. Here an increase in time two price volatility aids the period one investors. The reason for this is that in a full information setting price volatility (due to non-dividend information) changes the expected return of the stock. With the ability to go long or short at will, investors always profit from anticipated variability in an asset's expected return. Consider, for example, a large positive shock to the time two price. Since the price is now high relative to fundamentals, and the expected return is low, investors will choose to short shares of the stock. This selling must be beneficial or they would not do it. The same logic allows one to conclude that large negative price shocks are also beneficial. The investors essentially have a free option to long or short in period two, and the value of this option increases in the volatility of the time two price.

In the case where $\overline{P}_2 \neq 0$ or $X_1(i) \neq 0$ an increase in price volatility may hurt or help investors. It helps for the same reasons given previously. However, consider the agents' time three wealth given in equation **(13)**. If $X_1(i) \neq 0$, then time 2 price volatility adds noise to agents' time three wealth. This decreases their utility. If $\overline{P}_2 \neq 0$, then from the wealth equation we see that the uninsurable part of the time two endowment shock will be

$$\eta(i)\left(-\overline{P}_2+\varepsilon_P\right) \tag{22}$$

where ε_P is the price uncertainty. As the fixed part of the time two price increases, the negative utility impact of ε_P will become more pronounced (this is because after taking an expectation with respect to $\eta(i)$ an agent's welfare will be proportional to $e^{\sigma^2_{\eta(i)}(\overline{P}_2-\varepsilon_P)^2}$). Overall then, the effect of a new fund on agents occurs through two channels. First the improved hedging opportunities against time two endowment shocks make investors better off at time one. Second, the increased price volatility from that very use makes them worse off.

An immediate consequence of Lemma 3 is that once price variability becomes sufficiently large, the entry of any new fund has to make society strictly better off. The reason is easily seen from equation **(21)**. The right hand side does not depend on the number of time 2 funds. But the left hand side, namely the price variability, does. Hence once enough funds have entered to allow **(21)** to hold, it will continue to hold for all new funds. Note, however, that it is possible for society to be better off with no funds at all. The reason for this is that **(21)** may never hold, and the negative effects of increased price variability may dominate the positive effects of increased ability to hedge endowment shocks.

2.2 Optimal Signal Choice

Lemma 3 has two additional implications for the signal a fund *j* will acquire. First, consider a candidate equilibrium in which each fund has been assigned a particular signal. Then if any fund can change its signal in such a way that it will reduce $\sigma_\eta^2(i)$ for all *i*, its investors will vote for the change. Second, the concavity of *V* in $\sigma_\eta^2(i)$ implies that funds can profitably trade off endowment risk across investors. Consider how much money a set of potential investors can raise to start a particular fund, with a particular signal. Now imagine changing the signal that the fund receives. Suppose that change reduces $\sigma_\eta^2(i)$ for individuals with a high value of $\sigma_\eta^2(i)$ and equally increases $\sigma_\eta^2(i)$ for those with low values of $\sigma_\eta^2(i)$. Then the investors can in fact raise more money with which to begin the fund. Essentially, this means the greatest sums can be raised for new trading firms by having the firms select strategies that act to equalize the residual endowment risk faced by their owners.

For the following analysis, assume that investors are evenly spread in characteristic space on an *n*-dimensional sphere. One interpretation is that of physical location. In this case, investors in Chicago receive endowment shocks that have a higher correlation with each other than with investors in California. The primary reason for assuming that investors are on a sphere is simply to eliminate end point problems when funds are attempting to determine what signal to acquire. While the influence of endpoints on fund signal choice and entry are potentially interesting they are also sufficiently complex that a thorough analysis would take the current paper far afield.

Also assume there *I* groups of investors, indexed via 1, ..., *I*. Each group of investors *i* receives time 2 endowment shocks *n(i)* with a mean of 0 and a variance of σ_n^2 (we assume for simplicity that all endowment shocks have the same variance). The correlation between endowments *i* and *j* is given by $e^{-\delta(i,j)}$, where δ(*i*,*j*) measures the distance between the two groups, and is given by

$$\delta(i,j) = \left\| \begin{bmatrix} \cos(\frac{i\,2\pi}{I}) \\ \sin(\frac{i\,2\pi}{I}) \end{bmatrix} - \begin{bmatrix} \cos(\frac{j\,2\pi}{I}) \\ \sin(\frac{j\,2\pi}{I}) \end{bmatrix} \right\|, \tag{23}$$

where $\|\bullet\|$ is the Euclidean norm. For example the correlation between groups 1 and I is the same as the correlation between groups 2 and 3. Funds choose the signal of some investor group *i*, and will trade *n(i)* shares of the risky asset at time two on behalf of their shareholders.

Figure 1 shows that for a large number of investor classes with total share endowment at times 1 and 2 of zero, the funds which they find optimal in period two will indeed be as "spread out" as possible. The circles in the figure represent the characteristic space. Circles with larger radii correspond to economies with more time 2 funds (for example, the second innermost circle corresponds to a 2 fund economy – which are the dots on the circle). Notice that as long as the total time 2 supply is 0, in each economy (i.e. the economies have 1, 2, 3, and 4 funds respectively), the funds are optimally positioned in the most "spread out" way. That is, they choose signals in a way which minimizes the total amount of cross-signal correlation. By choosing minimally correlated signals, the funds maximize the fraction of society's time 2 endowment shocks which is ex-ante insurable.

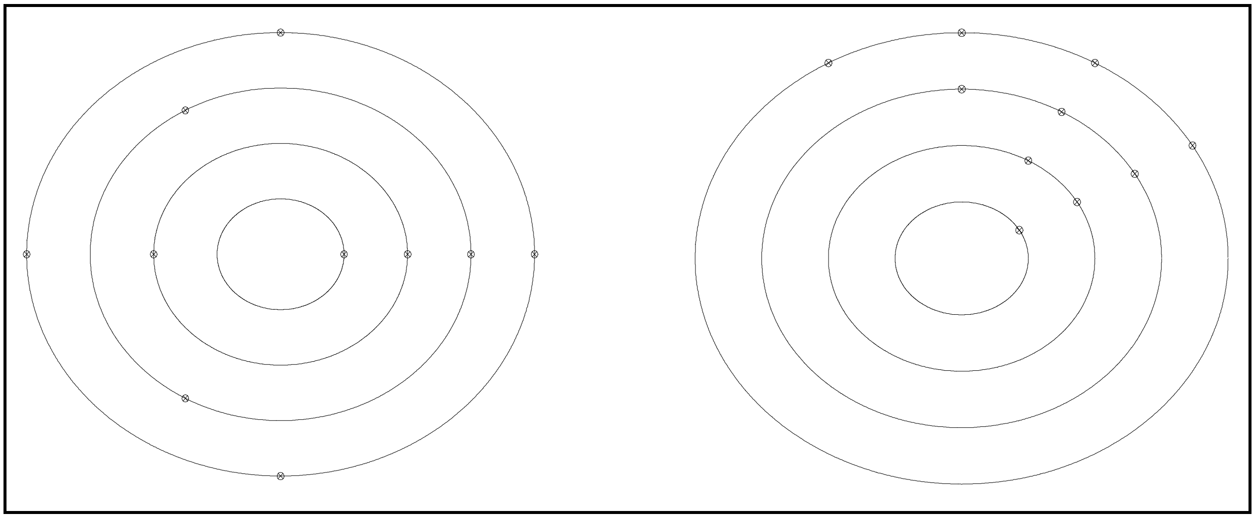

Figure 1. This figure shows the way that investors would choose to position the signals received by the optimal funds in the agents' spherical characteristics space. The innermost circle corresponds to the one fund economy, and the outermost circle corresponds to the 4 fund economy. The parameter values for the right graph are $\gamma = 1, \sigma_D^2 = 0.1, X_2^0 = 10$, and the number of investor groups is 12. The left graph has $X_2^0 = 2$.

Consider however what happens when time two supply is non-zero. As has already been pointed out, now there are two factors at play when funds decide on whether or not to enter the market. First of all funds want to provide the maximal amount of time two hedging service for the time one investors. This would lead them to spread out as much as possible on the characteristics sphere. Second, however, recall that the entry of a new fund increases the price variability and this can lead them to wish to cluster. To see when and why this will occur note that when there is a positive supply which will enter the market at time two, the time two risk discount in the price will be higher. Recall that the time two unhedged endowment shock is proportional to

$$\eta(i)\left(X_2^0 + \int_i \alpha(i)' e\, di \right). \tag{24}$$

Thus, **(24)** implies that increasing the time two supply X_2^0 decreases the willingness of agents to bear time two price risk. That is, variability coming from $\int_i \alpha(i)' e\, di$ hurts investors more. Funds therefore have an incentive to choose signals e_j which are minimally informative (and hence which minimize the magnitude of the $\alpha(i)$ vectors). This is exactly the oppositive incentive that funds had when there was no time 2 supply.

For a given agent, the introduction of a new fund, while decreasing the amount of non-hedgeable risk, can still make that agent worse off because so many of the other agents in the economy will use that fund to hedge their endowment shocks, thereby increasing price variability. Indeed, from **Figure 1**, when time two supply is non-zero, funds optimally cluster, even though investors are still evenly distributed on a spherical characteristic space.

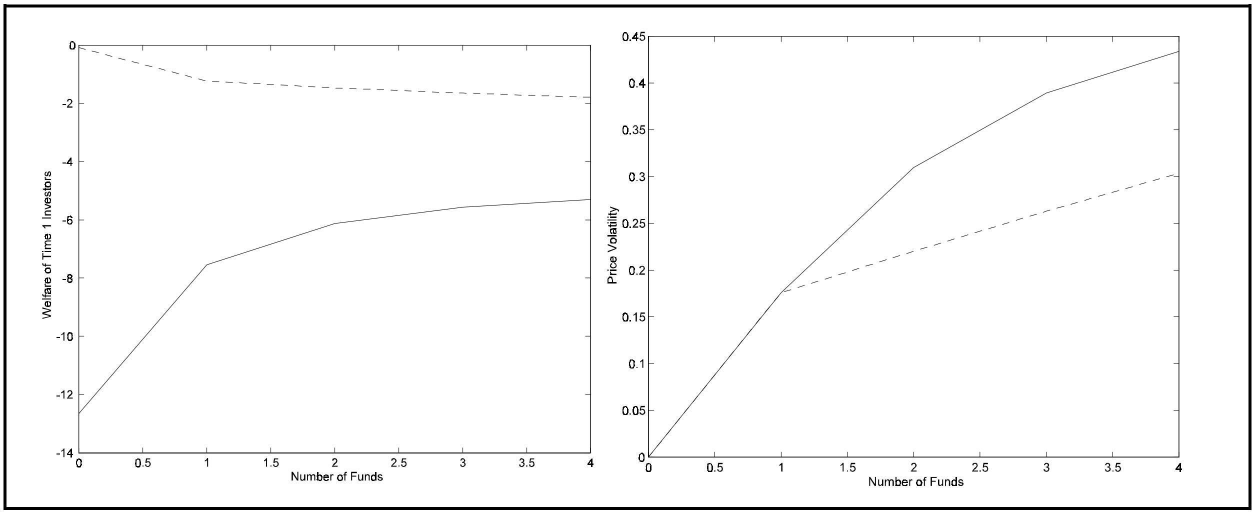

Figure 2. The left graph shows the behavior of social welfare as a function of the number of funds in the economy. The right graph shows the volatility of the period 2 price as more funds enter the economy. The dashed line refers to the economy with non-zero time 2 supply, and the solid line refers to the economy with zero time 2 supply. The parameter values for the dashed lines are $\gamma = 1,\ \sigma_D^2 = 0.1, X_2^0 = 10$, and the number of investor groups is 12. For the solid line $X_2^0 = 0$.

This clustering, by minimizing the amount of new hedging service that agents will have access to, also minimizes the amount of price variability that each new fund creates. **Figure 2** shows an example of this effect. In the economy with no time two supply, the price variability increases rapidly with new fund entry, as does social welfare. However, in the economy with a sufficiently large time two supply, new fund entry decreases social welfare. This can be seen from the downward sloping dashed line in the left graph of **Figure 2**. To minimize the welfare cost of new fund entry, additional funds choose to position themselves in a way which minimizes the price volatility (and which leads to the clustering evident in **Figure 1**). Indeed, time two price volatility (the dashed line in the right graph of **Figure 2**) in the non-zero supply case increases much more slowly than the price volatility in the zero supply economy.

2.3 Comparison to the Standard Model

The following table compares some of the results of our model to those in the standard paradigm.

Summary 3: A Comparison with the Social Welfare Properties In a Standard Model of Financial Market Intermediation

Result	Standard Model	First Principle's Model
Utility of Period One Traders with an Increase in FMIs	Increases due to improved risk sharing.	Ambiguous, depending upon whether the improved hedging opportunities outweigh the cost arising from additional price volatility.
Number of FMIs with Costless Entry	Infinite	Infinite or zero.
Signal Choice in the Absence of Period Two Supply	Select divergent signals to minimize competition.	Select divergent signals to maximize informativeness.
Signal Choice with Period Two Supply	Select divergent signals to minimize competition.	Cluster on single choice to reduce period two price volatility.

3 Equilibrium With Multiple Securities

3.1 The Agent's Problem

So far the analysis has only examined the case where a single risky security exists. While this restricted setting suffices to explore a wide range of issues, it necessarily makes it impossible to address questions regarding the relative prices of risky assets. To that end assume that there now exist K risky stocks and a single risk free bond. Dividends are now normally distributed with means of zero and variance-covariance matrix Σ_D. Similarly, the endowment shock for individual i equals the $K\times 1$ vector $N(i)$, and this vector is normally distributed.

Like the previous sections of the paper, period one investors can set up trading firms to act on their behalf in period two. As before, firms 1 through J obtain an informative normally distributed signal. The $J\times 1$ vector e of such signals includes e_j for firm j's signal. The $K\times J$ covariance matrix of the signals e and the endowment vector $N(i)$ equals $\Sigma_{eN(i)}$. From Bayes' rule, given the set of signals e, i's endowment vector can be written as

$$N(i) = A(i)\,e + \eta(i) \tag{25}$$

where $A(i)$ is a $K\times J$ matrix and $\eta(i)$ a $K\times 1$ residual vector. Thus, one can now write the trader's final wealth as

$$X_1(i)'[D - P_1] + \left[\eta(i) + A(i)'e + f\right]'[D - P_2] + X_1^0(i)'P_1^0 + B_1^0(i), \tag{26}$$

where $X_1(i)$, P_1, and f are now $K\times 1$ vectors, and Σ_η the $K\times K$ variance covariance matrix of the residuals η. Note, that f now represents the vector of aggregate holdings across all funds that the investor entrusts with his period two trades (i.e. $f = \sum_j f_j$, where f_j is a $K\times 1$ vector).

The first order conditions from this program yield equilibrium restrictions that are similar to those with one risky security and thus the analysis is not repeated here. The primary difference lies in the interpretation of the results, since demands are now vectors. Rather than demand a single constant fund each investor now wishes to use K such funds, one for each stock. Similarly, for the technical trading funds (those based only upon the price vector P_2). Rather than having one such fund, the investors now wish to invest in K such funds.

What separates out the multi security case is that the demands assigned across funds with informative signals by investor i now equals $A(i)e$. This opens up an argument for "fund families." The paper defines a fund family as a group of funds that share a common signal but have different trading strategies. This is economical. By sharing a common signal the funds save on research costs. At the same time their heterogenous trading strategies enable investors to produce a richer portfolio of customized strategies by appropriately investing across the funds in a particular family. Thus, the demand that investor i wishes to give to fund j equals the row vector $(a_{1j}, a_{2j,} \ldots, a_{Kj})' e_j$.

However, the above analysis opens up something of a dilemma. In general, one expects investors to outnumber funds in a particular fund family (or even the aggregate number of funds). But, in this case every individual investor can (in general) only satisfy his particular demands if there exist K funds within each fund family. That is separation now requires each fund family to offer one fund for each stock that exists! While fund families may endeavor to provide this level of service, it appears too costly to do so given the current technological environment. Still, the result does indicate that mutual funds have an incentive to lever off of their research department by providing more than one type of fund.[6]

The discussion, so far, raises the question as to whether or not a separation theorem of some sort exists when fund families contain fewer than K funds each. It turns out, the answer is yes. Assume that fund family j offers $K_j \leq K$ funds. Let W_j represent the $K \times K_j$ weighting matrix used by fund family j. Thus, w_{kmj} represents the relative number of shares of stock k fund m in fund family j will purchase given a signal e_j. For technical simplicity assume the rank of the weighting matrix W_j equals K_j. (If not, then simply eliminate the redundant funds and reduce K_j accordingly.) Thus, an investor in fund family j can independently set his holdings in K_j stocks given any signal e_j or any other variable, such as the price of the risky securities in the market. Having done so the remaining K-K_j stocks are then set according to the weighting matrix used by that fund family.

Let $W_{K_+,j}$ represent the $K_j \times K_j$ matrix encompassing the first K_j rows of W_j. Note, that this square matrix is of full rank, and thus invertible. Let $W_{K_-,j}$ equal the $[K\text{-}K_j] \times K_j$ matrix encompassing the last K-K_j rows of W_j. Thus, if investor i holds $\hat{a}_j(i)$ shares in the funds offered by fund family j he will then own $f_{j+} = W_{K_+,j}\hat{a}e_j$ shares of stocks one through K_j and $f_{j-} = W_{K_-,j}\hat{a}e_j$ shares in the remaining stocks via fund family j. (Implying $\left[f_{j+}{}' \;\; f_{j-}{}' \right] = f_j{}'$). Thus, one can write

$$f_{j-} = W_{K_-,j} W_{K_+,j}^{-1} f_{j+}, \tag{27}$$

---

[6] As of December 31, 1999 Morningstar lists the Merrill Lynch Group as the fund family containing the greatest number of individual funds with 286 as compared to the 8,435 stocks listed as available for trade by CRSP. Looking at the problem from the opposite end, out of the 630 fund families identified by Morningstar 462 had more than one fund.

since $\hat{a}\, e_j = W_{k+,j}^{-1}\ f_{j+}$.

In addition to the restriction given by **(27)** the analysis has to deal with the fact that the orders assigned to fund family *j* can only depend upon signal e_j and the publicly observed prices. This problem can be solved via a somewhat indirect approach. Imagine that an investor has given trading orders to every fund family other than *j* and now wishes to maximize his utility via *j*'s orders. In this case one can take as given the orders assigned to every other fund family, and treat the resulting endowment shocks as exogenous events. That is the (f_{i1}, $f_{i2,}$ . . ., f_{iK}) terms for fund family $i \neq j$ are taken as given when searching for the optimal values for (f_{j1}, $f_{j2,}$ . . ., f_{jK}). Also define $A_{\cdot j}(i)$ as the column vector used to update the investor *i*'s endowment shock given the signal e_j, and

$$\Omega_j = \begin{bmatrix} I \\ W_{K_-,j} W_{K_+,j}^{\ \ -1} \end{bmatrix} \tag{28}$$

as the matrix yielding f_j given holdings f_{j+}, that is $f_j = \Omega_j f_{j+}$. The resulting solution to the investor's optimization problem is characterized in the next lemma.

Lemma 4: *In a multiple security environment in which some or all fund families have fewer funds than stocks the solution to the investor's optimization problem for funds available in family j must satisfy*

$$f_{j+} = -\left[\Omega_j' M^{-1} \Omega_j\right]^{-1} \Omega_j' \left(M^{-1} A_{\cdot j}(i)\, e_j + \left(M^{-1} \gamma\, \Sigma_\eta + I\right) P_2\right) \tag{29}$$

where $M = \frac{1}{\gamma}\Sigma_D^{-1} - \gamma\Sigma_\eta$ *, and where we have assumed for simplicity that* $X_1(i) = 0$.

To obtain some insight regarding the solution to j_{j+} first consider the case where the fund family contains as many funds as there are stocks. In this case, Ω_j is simply a $K \times K$ identity matrix, and the solution to f reduces to $-A_j(i)e_j - (\gamma \Sigma_D)^{-1} P_2$. Thus, in the "complete markets" case the weights assigned to each stock returns to the previously discussed condition $(a_{1j}, a_{2j}, \ldots, a_{Kj})' e_j$.

Consider, however, what happens when markets are incomplete. Recall that the endowment vector of investor i may be written as $N(i) = A(i)e + \eta(i)$, where $A(i)$ is a K by J vector. Let us write $A_c(i)$ for the c^{th} column of $A(i)$. Ideally, as we have seen, the investor would like to be able to ask each fund family to trade the vector $A_c(i)e$. However, due to an incomplete set of funds, this may not be possible. The investor therefore solves the following problem

$$\min_f \text{ Distance Between } [\Omega\, f, A_c(i)] \qquad \textbf{(30)}$$

where the appropriate measure of distance has yet to be determined. One solution is to simply project, in a least squares sense, $A_c(i)$ into the space spanned by Ω to get

$$f = (\Omega'\Omega)^{-1}\Omega' A_c(i), \qquad \textbf{(31)}$$

the standard regression solution. Equation **(29)** seems similar to equation **(31)**.

However, this is not quite optimal. Since the projection cannot set the distance to zero, investors' risk preferences may cause them to deviate away from the least squares solution. In particular, consider a security which pays a very variable dividend at time three. Clearly it is more costly to leave oneself unhedged with respect to endowment shocks in this security than it is to leave oneself unhedged to endowment shocks in a security whose time three dividend is known for sure. Optimally an investor would therefore sacrifice a precise hedge of the latter security's shock, if this would allow him or her to better hedge the former security's endowment shock. A second effect is also in play. Consider two securities whose residual endowment variances (i.e. the part which is not spanned by the signals collected by the fund families) are small and large, respectively. It would be better to hedge one's endowment shock with respect to the less variable security than the shock with respect to the more variable security. In a sense hedging the more variable security is pointless because the residual variance of the endowment shock will still induce the investor to be very exposed to shocks in that security. However, by hedging out one's exposure to the security with the less variable endowment shock, the investor is able to almost fully eliminate any risk coming from this security exposure.

Hence, risk averse investors find it optimal to deviate away from the least squares solution. In fact, they optimally choose a weighting on the fund family's offered portfolios given by

$$f = (\Omega' M^{-1}\Omega)^{-1}\Omega' M^{-1} A_c(i) \qquad \textbf{(32)}$$

where the matrix $M = \left(\gamma\,\Sigma_D\right)^{-1} - \gamma\,\Sigma_\eta$ reflects investors' responses to the two issues just discussed: hedging distances from stocks with more variable dividends are penalized more, and hedging distance from stocks with more variable residual endowment shocks are penalized less. Furthermore, we see that as investors approach risk neutrality (i.e. as $\gamma \to 0$), their optimal portfolio weights reflect more the dividend hedging motive. And as investors become more risk averse (i.e. as $\gamma \to \infty$), their optimal weights reflect more their desire to only hedge those securities which have lower residual endowment variances.

Further note, that separation continues to hold. As long as there exist price funds and a constant fund, traders can satisfy their individual demands by holding a fixed number of shares in each managed mutual fund. Separation regarding the orders given by investors to the fund managers also continues to hold. Even though the price vector contains information otherwise unavailable to each fund manager the investors do not want the fund to trade on the basis of that information. Instead, investors only ask the fund manager to trade in response to that manger's signal. The investors then create the portfolio strategy they prefer by combining funds that gather information with those that do not. Instead of asking the managers gathering signals to trade on the basis of the revealed price vector, investors prefer to adjust their holding weights (the A matrix) to account for whatever information the market may reveal in period two.

Here, as in previous sections of the paper, one sees that modeling financial market intermediaries as corporations has a substantial impact on the conclusions one draws about the economy. In a model where such intermediaries maximize a risk averse utility function, each firm seeks to trade into the market portfolio. Thus, effectively all intermediaries pursue the same strategy, which is not surprising since they are assumed to act like identical human beings. By contrast, when such firms are modeled as corporations their investors demand that they pursue a wide range of strategies. That way the investing public can craft individual trading strategies via the appropriate purchase of positions in the individual investment funds.

3.2 The Equilibrium with an Incomplete Set of Funds

To gain further insight into the model's equilibrium characteristics this section adds the simplifying assumption that each fund family provides the same set of L funds (i.e. $\Omega_j = \Omega$ for all fund families). Funds thus differentiate themselves by trading on the basis of their specific signal e_j.

Given this simplification, and recalling the fact that one can write the endowment vector of each individual as $N(i) = A(i)e + \eta(i)$, the following result describes each individual's optimal trade.

Lemma 5: *In a K security environment in which all fund families offer the same set of L traded funds (with L < K), the optimal fund demand of investor i is given by*

$$f(i) = -\left[\Omega' M^{-1}\Omega\right]^{-1}\Omega'\left(M^{-1}\left(A_1(i)e_1 + \cdots + A_J(i)e_J\right) + \left(M^{-1}\gamma\Sigma_\eta + I\right)P_2\right) \qquad \textbf{(33)}$$

where $M = \left(\gamma\Sigma_D\right)^{-1} - \gamma\Sigma_\eta$.

Notice that this is an L by 1 vector which specifies which funds in each fund family investor i will choose. The trade can be implemented by giving the L by 1 vector

$$-\left[\Omega' M^{-1}\Omega\right]^{-1}\Omega' M^{-1}A_j(i) \qquad \textbf{(34)}$$

to each family j, and the L by K vector (recall there are K securities)

$$-\left[\Omega' M^{-1}\Omega\right]^{-1}\Omega'(M^{-1}\gamma\Sigma_\eta + I) \qquad \textbf{(35)}$$

to the price fund family (which allows trades in each of its L available funds to be some function of the K by 1 price vector P_2.)

The time 2 traders will optimally demand X_2 shares, where

$$X_2 = -\frac{1}{\gamma_2}\Sigma_D^{-1}P_2 \qquad \textbf{(36)}$$

Hence the market clearing condition sets

$$\int_i \Omega\, f(i)\, di + X_2 = X_2^0 \tag{37}$$

If one makes the further assumption that $\Sigma_{\eta(i)}$ is the same for all investors, then the time 2 price vector satisfies

$$\left(\Omega(\Omega' M^{-1}\Omega)^{-1}\Omega'(M^{-1}\gamma\Sigma_\eta + I) + \frac{1}{\gamma_2}\Sigma_D^{-1}\right)P_2 = -X_2^0 - \Omega(\Omega' M^{-1}\Omega)^{-1}\Omega' M^{-1}\bar{A}e \tag{38}$$

where $\bar{A} = \int_i A(i)di$ is a *J* by 1 vector.

This equation reveals the following insight: since Σ_D is of full rank, a *K* dimensional price vector exists which will support the equilibrium for any supply vector X_2^0. This would not necessarily be true if time 2 traders were not present in the market since demands of the time 1 traders only have *L* (with $L<K$) degrees of freedom. In this case an equilibrium price would exist only in the case where the time 2 supply was restricted to be a linear combination of the offered funds (i.e. one would need $X_2^0 = \Omega\,\tilde{X}_2^0$ for some *L* dimensional vector $\tilde{X}_2^0$).

Some straightforward manipulations allow us to express the time 2 price equation as follows

$$\Omega'\left(I - \gamma^2\Sigma_D\Sigma_\eta\right)^{-1}\left(P_2(1+\gamma/\gamma_2) + \gamma\Sigma_D(X_2^0 + \bar{A}e)\right) = 0. \tag{39}$$

Notice that if Ω were the *K* dimensional identity matrix, the price would be given by

$$P_2 = -\frac{\gamma\Sigma_D}{1+\gamma/\gamma_2}(X_2^0 + \bar{A}e) \tag{40}$$

which is exactly the multidimensional analog of the price in the 1 security world. However, given the fact that fund families offer only a subset of the possible funds, this standard *K* dimensional equilibrium condition does not have to hold. An *L* dimensional restriction of this condition must hold instead.

3.3 Transforming an Incomplete Economy into a Complete One

Define a *K*-*L* economy to be an economy with *K* securities and *L* (with *L*<*K*) fund vectors. Given the previous discussion, we see that this economy is incomplete in the sense that fewer funds exist than securities. While we are able to characterize the equilibrium of this economy, we have seen that the incompleteness of the fund space makes analysis somewhat difficult. In some circumstances it is possible, however, to transform an incomplete *K*-*L* economy into a complete *L*-*L* economy. Then equation **(40)** gives the equilibrium price in the *L*-*L* economy.

Let us rewrite equation **(39)** as follows

$$\widetilde{P}_2 \equiv \Omega' M^{-1}\Sigma_D^{-1}P_2 = -\frac{\gamma}{1+\gamma/\gamma_2}\Omega' M^{-1}(X_2^0 + \overline{A}e). \qquad \textbf{(41)}$$

This is an *L* dimensional restriction on the *K* dimensional price vector. Consider the *L* by *K* matrix

$$R \equiv \Omega' M^{-1}\Sigma_D^{-1}. \qquad \textbf{(42)}$$

This matrix gives us the *L* dimensional combination of the *K* securities in the incomplete economy whose price $\widetilde{P}_2$ is given by equation **(41)**. We will say that two economies are *identical* if, after trade at time one, the distribution of agents' residual endowment shocks at time two are everywhere equal (note that this is a stronger condition than saying that the endowment shocks should be equal with probability one).

Given certain conditions on the endowment shocks of the agents and on the time two supply vector, X_2^0, it is possible to create an *L*-*L* economy which is identical to the original *K*-*L* economy. We assume that for some *L*<*K* there exists an *K* by *L* dimensional funds matrix Ω such that the following conditions hold

$$\begin{aligned} X_2^0 &= \Sigma_D^{-1} M^{-1} \Omega \tilde{X}_2^0, \\ A(i) &= \Sigma_D^{-1} M^{-1} \Omega \tilde{A}(i) \qquad \forall i, \\ \eta(i) &= \Sigma_D^{-1} M^{-1} \Omega \tilde{\eta}(i) \qquad \forall i. \end{aligned} \tag{43}$$

where $\tilde{X}_2^0, \tilde{A}(i), \tilde{\eta}(i)$ have dimensions *L* by 1, *L* by *J*, and *L* by 1 respectively. These conditions say that for a sufficiently large *L* the endowment shocks which agents receive at time two, as well as the time two supply, are really *L* dimensional. The following theorem states the transformation result.

Theorem 3: *Assume that condition* ***(43)*** *holds. Then a* K-L *economy can be transformed into an equivalent* L-L *economy where the payoffs of the* L *securities is given by*

$$\tilde{D} = R\,D \tag{44}$$

(for the matrix R *from equation* ***(42)****), where agents' endowments are given by*

$$\tilde{N}(i) = \tilde{A}(i)\,e + \tilde{\eta}(i), \tag{45}$$

where the funds matrix is given by

$$\tilde{\Omega} = (\Omega' M^{-1} \Sigma_D^{-1} M^{-1} \Omega)^{-1} (\Omega' M \Omega), \tag{46}$$

and where the L *dimensional price vector is given by*

$$\tilde{P}_2 = R\,P_2 = \Omega' M^{-1} \Sigma_D^{-1} P_2 \qquad \textbf{(47)}$$

The significance of this theorem lies in how we can interpret the number of funds in an economy relative to the number of securities in an economy. First, given this theorem we can think of **(41)** not as an *L* dimensional restriction on the *K* prices in the *K*-*L* economy, but rather as the prices of the appropriately defined *L* securities in a complete *L*-*L* economy. More importantly, though, we see that the true dimensionality of an economy is not in the number of securities which exist (even if the payoffs of all of these are independent), but rather in the endowment structure of the economy. The number of fund vectors which are needed for full spanning is not equal to the number of independent security payoffs, but rather is equal to the number of factors which drive endowment shocks.

3.4 Comparison to the Standard Model

The following table contrasts the results of our intermediation model to those which apply in the traditional setting.

Summary 4: A Comparison of the Equilibrium FMI Strategies In a Standard Model of Financial Market Intermediation with Multiple Risky Securities

Result	Standard Model	First Principle's Model
Do Fund Families Exist?	No. Firms do not have an incentive to create additional competitors.	Yes. Funds families offer investors a richer strategy space for the utilization of any particular signal.
Strategies	Identical, all FMIs hold the market portfolio.	Divergent. Funds obtaining an informative signal trade only on the basis of their signal, and ignore prices. Other fund types that trade only on the basis of prices also exist, as well as funds that trade constant amounts independently of any signal.
Prices	Satisfy unrestricted mean-variance equilibrium conditions.	Prices depend both upon the dividend and supply variance via a risk weighted projection into the space of trading strategies available via existing FMIs.

4 Empirical Evidence on Fund Families

Morningstar maintains a list of 48 categories for mutual funds. The categories are an attempt to classify funds according to their investment styles and objectives. Table 2 provides evidence regarding the number of mutual fund families and funds in each Morningstar category. In general, it appears fund families try to spread their funds across a wide variety of categories. While some categories are more popular than others, there are approximately two and a half to three funds per fund family in each. Consider now the investor's optimization problem. The primary restriction comes from the Ω matrix. To the degree that a fund family only offers a few funds, or funds that form a poor strategy basis, Ω will restrict an investor's ability to use that fund family's funds to enhance his welfare. Thus, fund families have an incentive to set up a wide range of funds, all of which select investments via very different strategies. Thus, the finding in Table 2 that fund families appear evenly spread out, is equivalent to saying that they are trying to provide investors with strategies that cause Ω to impose as few restrictions as possible.

Further evidence that fund families try to provide a wide strategy basis for their investors can be found in Table 3. The first column sorts fund families by the number of funds they offer, while the second column sorts them by the number of Morningstar categories their funds fall into. The column labeled “# of Fam.” displays the number of fund families meeting the criteria given by the first two columns. Cells in which fund families have 75% or more of their funds in separate Moringstar categories have been shaded to show just how spread out funds within a family tend to be. While the very big funds have a great many funds within the same set of categories, this is largely due to the fact that there are only 48 categories available. Thus, fund families with say 100 or more funds will necessarily have on average over two funds per category. However, fund families with more modest offerings clearly attempt to spread out the strategies they offer to investors. This can also be seen in the following graph.

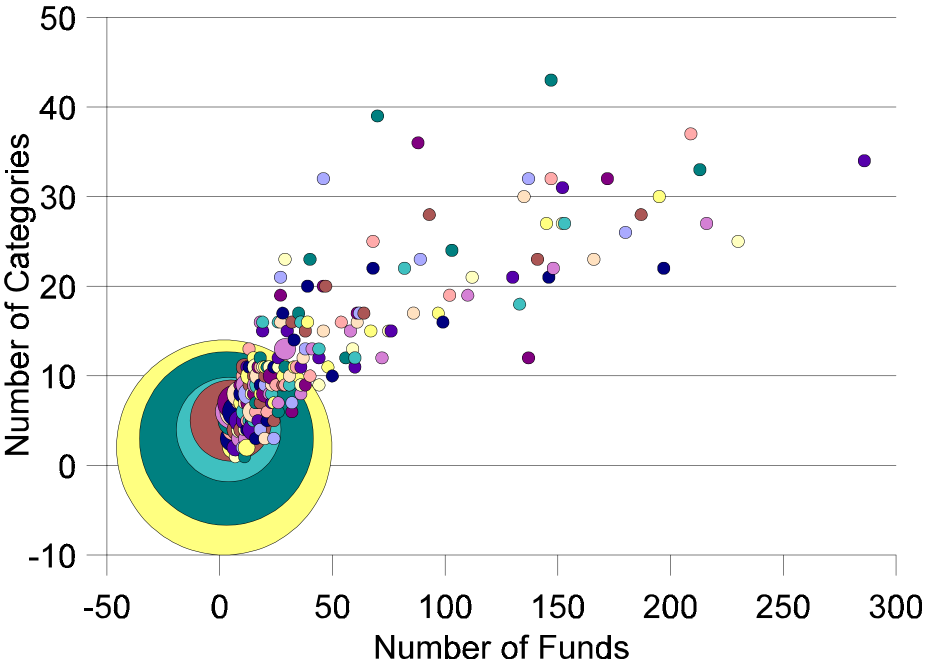

Figure 3

Figure 3 displays the data from Table 3 in graphical form.[7] Each circle represents a set of fund families with a particular number of funds and Morningstar categories. A circle's size indicates the number of funds at a particular point. As one can see fund families quickly seek to fill in about thirty or so Morningstar categories with their offerings. Translating this to the model's parameters, a fund family that offers funds within about thirty Morningstar categories provides a sufficiently rich Ω matrix that it becomes unprofitable to fill in additional parts of the strategy space. Indeed, only 13.3% of all fund families offer more than thirty funds altogether.

Within the model, one can also explain the existence of small fund families. In reality it

---

[7] The graph includes data on fund families with more than 32 funds, data which due to space constraints would not fit in Table 3.

seems unlikely that investors are arrayed uniformly within characteristic space. Thus, fund families that cater to investors in the less populated parts of the characteristic space are going to have insufficient demand for their product to profitably offer more than a few individual funds. These (numerous) fund families then show up in the graph as part of the large circles along the 45 degree line near the axis. Still, even though such fund families may not be able to offer many funds, the model indicates that those they do offer should span as much of the strategy space as possible. This corresponds well with the fact that such families tend to have most of their funds located in separate Morningstar categories.

4.1 Hypotheses and Tests Regarding Time Varying Fund Betas

4.1.1 Hypotheses

Another test of the model can be found in the time varying betas of each fund. Both Theorem 1 and Lemma 4 indicate that investors seek funds and fund families that offer to both obtain informative signals, and provide a rich set of trading strategies via their funds with which to make use of the signals. Among other places these asset allocation decisions should show up via systematic changes in fund betas. If the model is correct then managers go into and out of stocks on the basis of the information produced by their fund family’s research department and the rules governing their fund’s investment objectives. There are thus two sources of change for a fund’s beta. One is the information provided by the family’s research department, and this information naturally impacts every fund within a family. Two is a fund’s stock selection (vector in Ω), which varies across funds both within and across fund families.

Hypothesis 1: Time varying betas should show a higher level of correlation for funds within a family, than for funds across families.

Another more subtle hypothesis can also be found within the model's results. Sections 2.2 and 3.1 show that investors will value fund families that spread out in strategy space. Thus, one expects the first fund introduced by a family to be relatively unique in terms of it strategy (vector in Ω). This should in turn show up as a relatively low correlation between such a fund's time varying beta, and that of other funds. Now consider the second fund introduced by a fund family. Since the goal is to complete Ω, this fund must do something for investors that the first fund does not. In essence then since the first fund takes up the most unique spot available in the strategy space, the second fund should take up the next most unique, and the third fund the third most unique. This produces a ranking of the expected correlations for the time varying betas across funds based on the order in which the funds were introduced.

Let *O#* represent the organization number of a fund within a family. Thus, the first fund started by a family has a *O#* of 1, the second 2, etcetera. Also define $\rho_\beta(i, j)$ as the correlation between the time varying betas of two funds (in different fund families) with organization numbers *i* and *j*. Then empirically one expects to see:

Hypothesis 2: The correlation among time varying betas across fund families should increase with organization number of the funds being compared. That is $\rho_\beta(1,1) < \rho_\beta(1,2) < \rho_\beta(1,3)$, $\rho_\beta(2,2) < \rho_\beta(2,3)$, and $\rho_\beta(1,1) < \rho_\beta(2,2) < \rho_\beta(3,3)$.

4.1.2 Methodology

Data from the Morningstar January 2000 CD was combined with the CRSP mutual fund database. The Morningstar database includes active funds as of December 31, 1999. For each fund Morningstar lists its fund family, its origination date, and its ticker symbol, if any. These origination dates were then used to produce an organization number for each fund. This does mean that the organization numbers suffer from survivorship bias. However, at most this simply adds noise to one of the independent variables, which can only bias the results away from finding any particular patterns.

Using the ticker symbols the Morningstar data was then matched with the CRSP data to produce a return series for each fund. Month-by-month a beta was calculated for each fund using a 10 month weighted rolling regression. Only the five months before and after a particular month were used. No estimate includes the month in question. Thus, for example, a June beta was estimated via a weighted regression using data from January to May, and July to November. Weights were selected using a tri-cube kernel.

Once a beta series was calculated for each fund, funds were paired with each other. The database includes a fund pair only if they overlap in time for a sufficient number of years to produce at least 13 months of comparable data. Using the overlapping months during which beta estimates exist a correlation coefficient for the pair's time varying beta is calculated.

4.1.3 Results

Tables 4 and 5 contain the results. Each table reports the average time varying beta for pairs that meet the criteria for each cell. Below that is the standard deviation of the correlations and the number of observations per cell. Note that the tables support both Hypothesis 1 and Hypothesis 2. The beta correlations in Table 4 are uniformly higher than those in Table 5. In fact, the magnitude of those in Table 4 are nearly double those in Table 5. In terms of Hypothesis 2, the average correlation across funds increases as one moves to the right or down in the table. Exactly as predicted.

Panels B, and C of Table 5 break out the fund pairs that do and do not include index funds. As noted in section 1.3.2 while the model predicts such funds will exist, there is no reason to believe they provide particularly unique strategies for the investing public as they do not trade based upon informative signals. Thus, one expects their time varying beta correlations to be much higher than those of managed funds. Panel C shows that the methodology picks this up.

Clearly Table 5‘s panels do not provide firm estimates of the correlation between any fund pair’s time varying betas. A ten month weighted regression is simply uses too little data to believe that most of the time variation is anything but noise. However, Panel C does show just how high even this noisy measure can get. It thus provides a benchmark against which one can measure the results in both Table 4 and panels A and B of Table 5. The fact that the index funds have estimated correlation coefficients two to three times higher than those observed by the managed funds indicates support for the hypotheses put forward in the paper. While the estimators are no doubt noisy they do come out in the predicted order.

5 Relationship to the Existing Literature

For the most part papers that explore the topic of intertemporal trading take the existence of trading firms as given, along with some utility function for those firms. Instead research has looked at the design of the institutions through which such trade takes place. For example, Glosten (1994) asks whether trade via a limit order book dominates all other forms. A recent paper by Pirrong (1999) looks at the design of the exchange itself. In his paper individuals with risk averse utility functions form exchanges on which trade takes place.

In terms of the design of mutual funds there are two recent working papers in the area.

Ou-Yang (1997) looks at the optimal contract risk averse investors should hand to a single manager. In contrast to the current paper his model assumes that the investors only invest in the one mutual fund. That paper's major interest is the contract that emerges and the incentive scheme needed to invoke the necessary actions from the agent. The closest paper to the current article is probably that of Massa (2000). His paper assumes that investors have preferences over mutual fund types and then asks how the industry will form to satisfy those preferences. This paper differs from that one in that here preferences are homogenous (in that everybody is identical). Nevertheless, the current setting still produces a wide range of funds in equilibrium.

Previous research has also looked at the effects of the way in which investment managers are compensated. Das and Sundaram (2000) analyze the effects that different fee structures for investment advisors have on investor welfare. They find that "incentive" type fee structures are best from an investor's view point. Nanda, Narayanan, and Warther (2000) analyze the interaction of different types of fee structures and the ability of investment managers. They find that fund managers of differing abilities have an incentive to customize the fee structure that their fund charges in order to maximize the rents they are able to extract from their own skills. It is worth noting that this paper compliments the above papers. Where they look at the agency problem to be solved, this paper takes the existence of a solution as given and instead looks at the strategy the investors wish management to pursue.

The security design literature is also related to the present paper. Duffie and Jackson (1990) study the optimal design of futures contracts by exchanges which are trying to maximize their trading volume. They find that exchanges have an incentive to offer contracts which are maximally correlated with a linear combination of the unspanned portion of investors' endowment shocks. Our paper differs in that new funds choose their trading strategies so as to maximize investor welfare, rather than trading volume. Allen and Gale (1989,1991) consider the structure of financial markets when firms optimally issue costly new securities. They find that firms have an incentive to split up their income streams to allocate payoffs in a given state to those investors which most value them. Unlike in our setting where firms are assumed to be set up by individuals, they find that the private benefits of new securities to firms are different from the private benefits of new securities to society. Willen (1999) and Davis and Willen (2000a,b) analyze the social benefits of adding a security to a population of investors who have heterogeneous endowment shocks.

Many empirical papers have focused on the ability of mutual fund managers to outperform some benchmark, and on the effects that such performance has on future fund flows (see for example Carhart (1997), Chevalier and Ellison (1999), Lynch and Musto (2000), Khorana and Servaes (2000), Ackerman, McEnally, and Ravenscraft (1999), among others). The general conclusions of this literature are that very few fund managers can outperform consistently once risk loadings have been accounted for. Funds which have done well in the past tend to attract investors in the future. Also fund families which charge lower fees and offer a wide range of products tend to have a higher market share. Our papers sheds light on the proper basis for analysis of fund performance: risk adjusted alphas may not be the proper measurement of manager performance when the manager’s objective may dictate that he or she pursue a very different strategy than the one which would have maximized alpha. Furthermore, our model suggests that fund families have an incentive to differentiate themselves by their product offerings exactly so as to attract investors. Similarly researchers have documented a wide range of fund styles and many funds (see for example Gruber (1996), Brown and Goetzmann (1997), and Fung and Hsieh (1997)). These findings are roughly consistent with our model in that we predict that many funds ought to exist in order to provide investors with a maximally spanning set of trading strategies.

6 Conclusion

This paper provides the first model that examines the organization of financial market intermediaries as corporations. Unlike other articles in the area the model assumes investment banks do not come endowed with utility functions, rather they simply carry out instructions given by their shareholder. As a result the paper provides either a new explanation or the first explanation for a number of observed phenomena:

- Mutual funds outnumber traded securities.
- Mutual funds have a number of objective functions which lead to observably different trading styles. These styles include:
    - A wide variety of firms that trade on news regarding the endowment shocks of individuals. In the real world this might correspond to fundamental research about how various parts of the economy are doing.
    - A price fund that trades only on the equilibrium price. In essence a technical trading fund.
    - A fund that simply trades a fixed amount of stock. Basically an index fund.
- Adding mutual funds to the economy makes stock prices more volatile. This result contrasts sharply with models where funds act like people with utility functions. In those

papers additional funds reduce volatility.

- Adding mutual funds does not impact the risk premium. Again this contrasts with models where funds are basically risk averse traders. In those papers additional funds reduce the risk premium.
- There are incentives for funds that trade on market fundamentals to “spread out.” This produces divergent trading styles that investors which are not currently active in the market can use to better hedge out the risks they face.
- Multiple security markets encourage the formation of mutual fund “families.” Fund families allow several funds to use the same researched information, and this allows investors a more varied set of potential trading strategies based upon that information. In the multiple security case this improves investor welfare since investors will differ in how they wish to adjust their relative holdings in response to any particular signal.

References

Allen, F. and D. Gale, 1989, “Optimal Security Design,” *Review of Financial Studies,* 1 (3), 229–263.

Allen, F. and D. Gale, 1991, “Arbitrage, Short Sales, and Financial Innovation,” *Econometrica*, 59 (4), 1041–1068.

Ackerman, C., R. McEnally, and D. Ravenscraft, 1999, “The Performance of Hedge Funds: Risk, Return, and Incentives,” *Journal of Finance*, 54 (3), 833–874.

Brown, S. and W. Goetzmann, 1997, “Mutual Fund Styles,” *Journal of Financial Economics*, 43, 373-399.

Campbell, John, and Albert Kyle, 1993, “Smart Money, Noise Trading and Stock Price Behaviour,” *Review of Economic Studies*, 60, 1-34.

Carhart, M., 1997, “On Persistence in Mutual Fund Performance,” *Journal of Finance,* 52 (1), 57–82.

Chevalier, J. and G. Ellison, 1999, “Are Some Mutual Fund Managers Better Than Others? Cross-Sectional Patterns in Behavior and Performance,” *Journal of Finance*, 54 (3), 875–899.

Das, S.R. and R.K. Sundaram, 2000, “Fee Speech: Signalling, Risk-Sharing, and the Impact of Fee Structures on Investor Welfare,” working paper.

Davis, S.J. and P. Willen, 2000a, “Occupation-Level Income Shocks and Asset Returns: Their Covariance and Implications for Portfolio Choice,” *NBER working paper 7905.*

Davis, S.J. and P. Willen, 2000b, “Using Financial Assets to Hedge Labor Income Risk: Estimating the Benefits,” working paper.

Duffie, D. and M.O. Jackson, 1990, “Optimal Innovation of Futures Contracts,” *Review of Financial Studies*, 2 (3), 275–296.

Ferson, W.E. and R.W. Schadt, 1996, "Measuring Fund Strategy and Performance in Changing Economic Conditions," *Journal of Finance*, 51, 425-461.

Fung, W. and D.A. Hsieh, 1997, “Empirical Characteristics of Dynamic Trading Strategies: The Case of Hedge Funds,” *Review of Financial Studies,* 10 (2), 275–302.

Glosten, Larry, 1994, “Is the Electronic Open Limit Order Book Inevitable?,” *Journal of Finance*, 49, 1127-61.

Gruber. M., 1996, "Another Puzzle: The Growth in Actively Managed Mutual Funds," *Journal of Finance,* 51, 783–810.

Ho, Thomas and Hans Stoll, 1983, "The Dynamics of Dealer Markets Under Competition," *Journal of Finance*, 38,, 1053-1074.

Khorana, A. and H. Servaes, 2000, "What Drives Market Share in the Mutual Fund Industry," working paper.

Kyle, Albert, 1985, "Continuous Auctions and Insider Trading," *Econometrica*, 53, 1315–1335.

Lynch, A.W. and D.K. Musto, 2000, "How Investors Interpret Past Fund Returns," working paper.

Massa, Massimo, 2000, "Why So Many Mutual Funds? Mutual Fund Families, Market Segmentation and Financial Performance," INSEAD working paper.

Nanda, V., M.P. Narayanan, and V.A. Warther, 2000, "Liquidity, Investment Ability, and Mutual Fund Structure," *Journal of Financial Economics,* 57, 417–443.

Ou-Yang, Hui, 1997, "Optimal Contracts in Continuous Time Principal-Agent Problems," University of North Carolina at Chapel Hill working paper.

Pirrong, Craig, 1999, "The Organization of financial Exchange Markets: Theory and Evidence," *Journal of Financial Markets*, 2, 329-358.

Ross, S.A., 1973, "The economic theory of agency: The principal's problem," *American Economic Review,* 63 (2), 134–139.

Sharpe, W., 1966, "Mutual Fund Performance," *Journal of Business,* 39, 119–138.

Sharpe, W., 1992, "Asset Allocation: Management Style and Performance Measurement," *The Journal of Portfolio Management*, 18, 7-19.

Theil, Henri, 1971, *Principles of Econometrics,* John Wiley & Sons, New York.

Whittle, Peter, 1990, *Risk-Sensitive Optimal Control*, John Wiley & Sons, New York.

Willen, P., 1999, "Welfare, Financial Innovation and Self Insurance in Dynamic Incomplete Market Models," working paper.

Appendix

Lemma 1*: Investor i sets f_j to solve*

$$X_1(i) + \left[\sum_j \alpha_j e_j + f_j \right] + \frac{1}{\gamma} \sigma_D^{-2} P_2 = 0. \tag{8}$$

Proof: By using **(7)** in **(6)** and then employing the risk sensitive certainty equivalence principle from Whittle (1990) the solution to the investor's problem for the f_j can be found via the following program

$$\underset{f_j, D, \eta}{\text{ext}} \quad \begin{aligned} & X_1(i)[D - P_1] + \left\{ \eta(i) + \sum_j \alpha_j e_j + f_j \right\} [D - P_2] + X_1^0(i) P_1^0 + B_1^0(i) \\ & + \frac{1}{2\gamma} \left[\sigma_D^{-2} D^2 + \sigma_\eta^{-2} \eta^2 \right]. \end{aligned} \tag{48}$$

Borrowing Whittle's terminology the expression *ext* stands for "extremization" and implies that the objective is maximized with respect to the controls (f_j) and minimized with respect to the unknowns (*D*, and η).

The resulting first order conditions with respect to the f_j,

$$D - P_2 = 0, \tag{49}$$

the dividend *D*,

$$X_1(i) + \eta(i) + \sum_j \alpha_j e_j + f_j + \frac{1}{\gamma} \sigma_D^{-2} D = 0, \tag{50}$$

and finally the residual endowment shock η,

$$D - P_2 + \frac{1}{\gamma}\sigma_\eta^{-2}\eta = 0 \qquad \textbf{(51)}$$

are found via differentiation of **(48)**. Using the above three equations to eliminate *D* and η leads to the solution for f_{j} given in the lemma.

Q.E.D.

Theorem 2. *Price volatility increases with the number of mutual funds.*

Proof: Consider the change in price volatility when going from *J*-1 funds to *J* funds. Let ${}_{J-1}\Sigma_{n(i),e}$ represent the *J*-1×1 covariance vector between the endowment shock *n*(*i*) and the first *J*-1 signals. Let ${}_{J-1}\Sigma_e$ represent the *J*-1×*J*-1 variance-covariance matrix of the first *J*-1 signals, and ${}_{J-1,J}\Sigma_e$ the *J*-1×1 covariance vector between the first *J*-1 signals and the *J*th signal. Then one can write

$$\sigma_{PJ}^2 = \int_x\int_y \begin{bmatrix} {}_{J-1}\Sigma_{n(x),e} \\ \sigma_{n(x),e_J} \end{bmatrix}' \begin{bmatrix} {}_{J-1}\Sigma_e & {}_{J-1,J}\Sigma_e \\ {}_{J-1,J}\Sigma_e' & \sigma_{e_J}^2 \end{bmatrix}^{-1} \begin{bmatrix} {}_{J-1}\Sigma_{n(x),e} \\ \sigma_{n(y),e_J} \end{bmatrix} dxdy. \qquad \textbf{(52)}$$

Using the formula for the partitioned inverse this reduces to

$$\begin{aligned} \sigma_{PJ}^2 = {} & \int_x\int_y {}_{J-1}\Sigma_{n(x),e}{}'\ {}_{J-1}\Sigma_e\ {}_{J-1}\Sigma_{n(y),e} + \\ & \left({}_{J-1}\Sigma_{n(x),e}{}'\ {}_{J-1}\Sigma_e\ {}_{J-1,J}\Sigma_e - \sigma_{n(x),e_J}\right)\left(\sigma_{e_J}^2 - {}_{J-1,J}\Sigma_e{}'\ {}_{J-1}\Sigma_e\ {}_{J-1,J}\Sigma_e\right)^{-1} \times \\ & \left({}_{J-1}\Sigma_{n(y),e}{}'\ {}_{J-1}\Sigma_e\ {}_{J-1,J}\Sigma_e - \sigma_{n(y),e_J}\right) dxdy. \end{aligned} \qquad \textbf{(53)}$$

Recall that $\Sigma_{n(i),e}'\Sigma_e^{-1}$ equals α(i). Let ${}_{J\text{-}1}\alpha(i)$ represent the vector $\alpha(i)$ when only the first J-1 signals are available. Then **(53)** further simplifies to

$$\begin{aligned}\sigma_{PJ}^2 = \int_x\int_y & {}_{J-1}\Sigma_{n(x),e}' \; {}_{J-1}\Sigma_e \; {}_{J-1}\Sigma_{n(y),e} + \\ & \left({}_{J-1}\alpha(x)' \; {}_{J-1,J}\Sigma_e - \sigma_{n(x),e_J}\right)\left(\sigma_{e_J}^2 - {}_{J-1,J}\Sigma_e' \; {}_{J-1}\Sigma_e \; {}_{J-1,J}\Sigma_e\right)^{-1} \times \\ & \left({}_{J-1}\alpha(y)' \; {}_{J-1,J}\Sigma_e - \sigma_{n(y),e_J}\right) dxdy.\end{aligned} \tag{54}$$

The first term in the integral of **(54)** equals the price variance when there exist J-1 mutual funds. Thus, one only needs to sign the second term in the integral to determine if the price variance increases or decreases in the number of mutual funds.

The center term in the product of **(54)**, $\left(\sigma_{e_J}^2 - {}_{J-1,J}\Sigma_e' \; {}_{J-1}\Sigma_e \; {}_{J-1,J}\Sigma_e\right)^{-1}$, is the J^{th} diagonal element of the inverse of Σ_e. Since Σ_e is a positive definite matrix, the diagonal terms of its inverse must be positive and thus this term is positive. Also, note that $\left(\sigma_{e_J}^2 - {}_{J-1,J}\Sigma_e' \; {}_{J-1}\Sigma_e \; {}_{J-1,J}\Sigma_e\right)^{-1}$ is a positive scalar that does not vary by investor. Thus, one can change the order of integration of the product and rewrite it as

$$\begin{aligned}& \int_x\left({}_{J-1}\alpha(x)' \; {}_{J-1,J}\Sigma_e - \sigma_{n(x),e_J}\right) dx \int_y\left({}_{J-1}\alpha(y)' \; {}_{J-1,J}\Sigma_e - \sigma_{n(y),e_J}\right) dy = \\ & \left[\int_x\left({}_{J-1}\alpha(x)' \; {}_{J-1,J}\Sigma_e - \sigma_{n(x),e_J}\right) dx\right]^2 > 0.\end{aligned} \tag{55}$$

Q.E.D.

Lemma 4: *In a multiple security environment in which some or all fund families have fewer funds than stocks the solution to the investor's optimization problem must satisfy*

$$f_{j+} = -\left[\Omega'_j\left(\gamma\Sigma_\eta - \frac{1}{\gamma}\Sigma_D^{-1}\right)^{-1}\Omega_j\right]^{-1}\Omega'_j\left(\gamma\Sigma_\eta - \frac{1}{\gamma}\Sigma_D^{-1}\right)^{-1}A_{\cdot j}(i)e_j, \tag{28}$$

for the funds available in fund family *j*.

Proof: If there exist as many funds in each fund family as stocks then the objective function **(48)** can be expressed in matrix notation as

$$\underset{f,D,\eta}{ext}\quad \begin{aligned}&X_1(i)'[D-P_1]+\left[\eta(i)+A(i)e+f\right]'[D-P_2]+X_1^0(i)'P_1^0+B_1^0(i)\\&+\frac{1}{2\gamma}\left[D'\Sigma_D^{-1}D+\eta'\Sigma_\eta^{-1}\eta\right].\end{aligned} \tag{56}$$

However, when fund family *j* has fewer than *K* funds problem **(48)** further transforms to

$$\underset{f_{j+},D,\eta}{ext}\quad \begin{aligned}&X_1(i)'[D-P_1]+\left[\eta(i)+A_{\cdot j}(i)e_j+\hat{A}_{\cdot j}(i)P_2\right]'[D-P_2]+f_{j+}{}'[D_{j+}-P_{2,j+}]\\&+f_{j+}{}'W_{K_+,j}{}^{-1\prime}W_{K_-,j}{}'[D_{j-}-P_{2,j-}]+X_1^0(i)'P_1^0+B_1^0(i)+\frac{1}{2\gamma}\left[D'\Sigma_D^{-1}D+\eta'\Sigma_\eta^{-1}\eta\right]\end{aligned} \tag{57}$$

for fund *j*, where the problem now allows for the fact that only K_j stock demands can be selected as free variables. Here, D_{j+}, and D_{j-} represent the dividends paid by the first K_j stocks and the remaining K-K_j stocks respectively. The variables $P_{2,j+}$, and $P_{2,j-}$ represent the analogous period two prices for the stocks. The $A_{\cdot j}(i)$ and $\hat{A}_{\cdot j}(i)$ terms represent column vectors that convert the signals e_j and P_2 into estimates of the total shares owned by investor *i* through both his endowment shock *N*(*i*) and orders to the other investment fund families. The residual η is then the error term from these two estimates. Thus, one implicitly has

$$N(i) + \sum_{h \neq j} f_h = A_{\cdot j}(i)e + \hat{A}_{\cdot j}(i)P_2 + \eta(i). \tag{58}$$

Differentiating **(56)** with respect to f_{j+}, D, and η yields first order conditions of,

$$D_{j+} - P_{2,j+} + W_{K_+,j}^{-1\prime} W_{K_1,j}{}' (D_{j-} - P_{2,j-}) = 0, \tag{59}$$

$$X_1(i) + \eta(i) + e_j A_{\cdot j}(i)' + P_2' \hat{A}_{\cdot j}(i)' + \begin{bmatrix} f_{j+} \\ W_{K_-,j} W_{K_+,j}{}^{-1} f_{j+} \end{bmatrix} + \frac{1}{\gamma} \Sigma_D^{-1} D = 0, \tag{60}$$

and

$$D - P_2 + \frac{1}{\gamma} \Sigma_\eta^{-1} \eta = 0 \tag{61}$$

respectively.

Use equations **(59)** and **(61)** to eliminate D and η from **(60)**. Next use the definition for Ω_j given by **(28)** and write the solution to f_{j+} as **(28)**, where the assumed existence of a complete set of price funds and constant funds have been used to eliminate the X_1 and P_2 vectors. Q.E.D.

The following is the proof of Theorem 3.

Proof of Theorem 3. First we need to show that the endowment shocks in the 2 economies are indeed the same. Given an L dimensional vector x of securities in the L-L economy, their payoff will be equal to $R'x$ (a K dimensional vector) in the K-L economy. Hence we need to show that $N(i) = R'\tilde{N}(i)$, which is true by construction.

Next we need to show that the trade vector *f(i)* will be the same in the two economies. Let us define

$$\tilde{\Sigma}_D \equiv R\Sigma_D R' = \Omega' M^{-1}\Sigma_D^{-1}M^{-1}\Omega. \qquad \textbf{(62)}$$

Using **(33)** and the conditions in **(43)**, we see that the demand in the *K*-*L* economy is given by

$$f(i) = -(\Omega' M^{-1}\Omega)^{-1}\tilde{\Sigma}_D\tilde{A}(i)\,e - (\Omega' M^{-1}\Omega)^{-1}\frac{1}{\gamma}\tilde{P}_2\,. \qquad \textbf{(63)}$$

Similarly we find that since it is complete in the *L*-*L* economy an agent's demands are given by

$$f(i) = -\tilde{\Omega}^{-1}\tilde{A}(i)\,e - \tilde{\Omega}^{-1}\,\tilde{\Sigma}_D^{-1}\frac{1}{\gamma}\tilde{P}_2\,. \qquad \textbf{(64)}$$

Given the definition of $\tilde{\Omega}$ from equation **(46)**, we see that these are indeed identical.

That the price vector is given by **(47)** can be seen by observing that the complete markets price equation **(40)** in implies that

$$\tilde{P}_2 = -\frac{\gamma\tilde{\Sigma}_D}{1+\gamma/\gamma_2}(\tilde{X}_2^0 + \tilde{\bar{A}}e). \qquad \textbf{(65)}$$

The results follows from the conditions in **(43)**, as well as the definition of $\tilde{\Sigma}_D$ given above.

Q.E.D.

Table 1: Division of Equity Funds by Type.

Year	Number of Equities[1]	Equity Funds[2]		Funds Primarily Invested in Equities[3]		Equity Only Funds[4]		Equity Index Funds[5]		Asset Allocation Funds[6]	
		#	TNA[7]	#	TNA	#	TNA	#	TNA	#	TNA
1990	6,635	785	296,361	182	73,469	37	15,344	5	2,631	465	188,566
1991	6,529	738	395,528	218	118,617	39	26,232	6	5,318	430	262,492
1992	6,630	3,998	1,454,405	729	166,378	281	58,108	36	11,837	1,035	429,002
1993	6,790	5,392	1,910,069	1,049	282,405	467	95,953	52	20,347	1,297	624,560
1994	7,554	7,167	1,970,881	1,656	437,746	696	120,200	57	23,206	1,521	658,459
1995	7,982	8,262	2,562,696	2,071	560,718	991	190,866	68	40,303	1,714	871,821
1996	8,221	9,027	3,103,536	2,654	904,693	1,257	295,774	87	75,592	1,832	1,073,854
1997	8,770	10,741	3,945,730	3,604	1,444,976	2,018	476,416	134	128,992	1,907	1,339,362
1998	8,822	11,322	4,805,749	4,410	2,067,695	2,059	678,513	146	198,588	1,877	1,551,096
1999	8,435	11,882	5,809,238	4,936	2,769,340	1,996	934,607	178	296,102	1,775	1,802,193

1. Total companies with available prices on the NYSE, AMEX and NASDAQ combined as of the first trading day in January.
2. Any fund that held 5% or more of its net asset value in equities at some time during its life.
3. Any fund with 90% or more of its portfolio invested in equities in that year.
4. Funds that held 90% or more of their net asset value in equities over their entire life with a minimum of three years in operation.
5. Any fund with “Index” or “Idx” as part of its name and with 90% or more of its portfolio invested in equities in that year.
6. Any fund in which equity holdings as a fraction of all assets vary by 20 or more percentage points during the life of the fund.
7. Total net assets in millions of dollars.

Table 2: Total Number of Fund Families and Funds in Each Morningstar Category as of December 31, 1999.

Morningstar Categories	# of Fund Families[1]	# of Funds[2]	Morningstar Categories	# of Fund Families	# of Funds
Convertibles	24	57	Muni CA Long	38	111
Diversified Emerging Mkts	72	168	Muni NY Interm	18	38
Diversified Pacific/Asia Stock	21	51	Muni NY Long	39	87
Domestic Hybrid	216	747	Muni National Interm	88	181
Emerg Mkts Bond	21	46	Muni National Long	98	307
Europe Stock	55	152	Muni Short	44	103
Foreign Stock	214	668	Muni Single State Interm	58	250
High Yield Bond	105	321	Muni Single State Long	69	717
Intermediate Government	102	308	Pacific/Asia ex-Japan Stock	35	117
Intermediate-term Bond	209	554	Short Government	73	174
International Bond	63	196	Short-term Bond	94	216
International Hybrid	31	82	Small Blend	110	238
Japan Stock	24	52	Small Growth	150	375
Large Blend	242	934	Small Value	105	235
Large Growth	191	645	Specialty-Communication	12	21
Large Value	182	614	Specialty-Financial	25	65
Latin America Stock	25	53	Specialty-Health	25	59
Long Government	28	50	Specialty-Natural Res	24	65
Long-term Bond	56	120	Specialty-Precious Metals	21	42
Mid-Cap Blend	115	262	Specialty-Real Estate	53	128
Mid-Cap Growth	138	393	Specialty-Technology	48	104
Mid-Cap Value	105	275	Specialty-Utilities	29	97
Multisector Bond	43	129	Ultrashort Bond	35	54
Muni CA Interm	25	39	World Stock	83	279

Notes: Total of 48 Categories, 630 Fund Families, and 10,979 Funds.

1. Number of fund families with at least one fund in each category.
2. Number of individual funds in each category.

Table 3: Distribution of Funds within Morningstar Categories by Fund Family.

# of Funds[1]	# of Cat.[2]	# of Fam.[3]	# of Funds	# of Cat.	# of Fam.	# of Funds	# of Cat.	# of Fam.	# of Funds	# of Cat.	# of Fam.
1	1	168	10	2	1	15	4	3	22	8	1
2	1	23	10	4	2	15	5	1	22	11	1
2	2	45	10	7	2	15	6	3	23	7	1
3	1	8	10	8	2	15	8	1	23	8	1
3	2	9	10	10	1	15	9	1	23	9	1
3	3	36	11	1	1	15	10	1	23	10	2
4	1	2	11	3	1	15	12	1	24	3	1
4	2	6	11	4	1	16	3	1	24	5	1
4	3	13	11	5	2	16	6	1	24	7	1
4	4	21	11	6	2	16	7	1	24	11	1
5	2	1	11	7	1	16	9	2	25	9	1
5	3	6	11	8	1	16	11	2	26	6	1
5	4	7	11	9	2	17	5	1	26	7	1
5	5	16	11	11	2	17	9	2	26	11	1
6	2	3	12	2	2	17	10	2	26	12	1
6	3	4	12	4	1	17	11	1	26	16	1
6	4	3	12	5	1	18	4	1	27	16	1
6	5	5	12	8	1	18	7	1	27	19	1
6	6	6	12	9	3	18	8	1	27	21	1
7	1	1	12	10	1	18	9	1	28	9	1
7	2	2	12	11	1	18	11	1	28	10	1
7	4	2	13	5	1	18	12	1	28	17	1
7	6	5	13	6	1	18	16	1	29	9	1
7	7	6	13	7	2	19	11	1	29	11	1
8	3	1	13	8	3	19	15	1	29	13	3
8	4	3	13	9	1	19	16	1	29	23	1
8	5	2	13	10	2	20	3	1	30	15	1
8	6	5	13	11	1	20	8	2	31	9	1
8	7	1	13	13	1	20	9	1	31	10	1
8	8	3	14	5	1	20	10	1	32	6	1
9	3	2	14	6	1	20	11	1	32	7	1
9	4	2	14	8	1	21	5	1	32	16	1
9	5	3				21	6	1			
9	7	1				21	11	1			
9	8	4									
9	9	1									

Notes: Fund families with more than 32 funds not displayed.
1. Number of funds offered by the fund family.
2. Number of Morningstar distinct categories in which the funds appear.
3. Number of fund families within the grouping.

Light shading 100% > # of Categories/# of Funds > 75%,
Dark shading # of Categories = # of Funds.

Table 4: Beta Correlations Within Families

	Average	1 With 2	1 With 3	2 With 3
All Funds	0.2279	0.2212	0.2704	0.2116
	(0.4726)	(0.4514)	(0.4317)	(0.4541)
	239,201	1,759	340	1,423
Correlations ≠ ±1	0.2279	0.2212	0.2704	0.2116
	(0.4726)	(0.4514)	(0.4317)	(0.4541)
	239,191	1,759	340	1,423
Correlations ≠ ±1 & No Index Funds	0.2292	0.2220	0.2696	0.1995
	(0.4745)	(0.4535)	(0.4368)	(0.4309)
	231,727	1,710	318	1,366

Notes: Betas for each fund are calculated with a via a rolling tri-cubed kernel estimator. Window is set to five months before and after each date. The weighted regression excludes the month being estimated. For example, the estimated June beta for a fund uses a weighted average of the date from January to May and July to November. Funds are initially included when they have one full calendar year of data. Correlations are only estimated when there exists over one year's worth of overlapping dates with which to conduct the calculation.

Cell entries: Top – average correlation between the betas of each fund. Middle (in parentheses) – standard *deviation*. Bottom – number of observations.

Table 5: Beta Correlations Across Fund Family Funds

	Panel A: Fund Introduction Number Within a Family			
	All Pairs			
	All	1	2	3
All	0.1214 (0.6372) 34,712,391	0.0873 (0.5865) 4,470,118	0.1110 (0.6212) 6,496,828	0.1264 (0.5975) 2,122,379
1		0.0739 (0.5225) 152,157	0.0825 (0.5541) 602,592	0.0834 (0.5279) 294,149
2			0.1077 (0.6032) 331,868	0.1110 (0.5937) 542,254
3				0.1323 (0.5497) 32,959
	Panel B: All Pairs Excluding Those With an Index Fund			
All	0.1229 (0.6359) 33,450,827	0.0872 (0.5863) 4,421,938	0.1111 (0.6213) 6,390,393	0.1266 (0.5986) 2,064,742
1		0.0731 (0.5207) 149,966	0.0815 (0.5536) 590,782	0.0822 (0.5273) 287,644
2			0.1070 (0.6047) 322,569	0.1101 (0.5956) 525,117
3				0.1329 (0.5536) 31,443
	Panel C: Only Pairs in Which Both Funds are Index Funds			
All	0.3307 (0.7204) 11,313	0.3143 (0.7950) 599	0.2799 (0.6756) 1,654	0.4694 (0.6181) 893
1		-0.0102 (1.0844) 6	0.3111 (0.7959) 51	0.3358 (0.8863) 30
2			0.2591 (0.5822) 60	0.3473 (0.5849) 131
3				0.6110 (0.5423) 13

Notes: Betas for each fund are calculated via a rolling tri-cubed kernel estimator. Window is set to five months before and after each date. The weighted regression excludes the month being estimated. For example, the estimated June beta for a fund uses a weighted average of the data from January to May and July to November. Funds are initially included when they have one full calendar year of data. Correlations are only estimated when there exists over one year's worth of overlapping dates with which to conduct the calculation. Correlations are only included if both funds are in *separate* fund families. Fund introduction number refers to the order in which funds (in existence as of December 1999, and tracked by Morningstar) were introduced within a family. The oldest operating fund within a family is labeled 1, the second oldest 2, etcetera.

Cell entries: Top – average correlation between the betas of each fund. Middle (in parentheses) – standard *deviation*. Bottom – number of observations.