How do family strategies affect fund performance?
When performance-maximization is not the only game in town.

Massimo Massa*

INSEAD

Abstract

This is a first attempt to study how the structure of the industry affects mutual fund behavior. We show that industry structure matters. We show that the mutual fund families employ strategies that rely on the heterogeneity of the investors in terms of investment horizon by offering the possibility to switch across different funds belonging to the same family at no cost. We argue that this option acts as an externality for all the funds belonging to the same family, affecting the target level of performance the family wants to reach and the number of funds it wants to set-up.

By using the universe of the US mutual fund industry, we empirically confirm this intuition. We analyze the relationship between performance, fund proliferation and the degree of product differentiation in the mutual fund industry. In particular, we show that the degree of product differentiation *negatively* affects performance and *positively* affects fund proliferation.

JEL Classification: G11, G23

Keywords: Mutual funds, market structure, performance, behavioral finance.

* Finance Department, INSEAD, Boulevard de Constance 77305, Fontainebleau Cedex, France.
e-mail: massimo.massa@insead.fr, Tel: (33)(0)1 60724481, Fax : (33) (0)1 60724045.

1 Introduction

The most glaring stylized fact about the mutual fund industry is the existence of a very high number of funds, differentiated into market "categories" and belonging to relatively few families. Nowadays, the number of mutual funds in the US has already overtaken the number of stocks traded on NYSE and AMEX added together, reaching 8,171 units. In particular, in the period 1990-2000 the number of mutual funds grew from 3,081 to 8,171 while the number of families has only slightly increased, from 361 to 431. During the same time, the degree of segmentation of the industry also grew, reaching around 33 different categories.

This can hardly be explained in terms of the traditional finance literature, not only because there already exists a number of securities presumably sufficient to pursue optimal investment strategies, but also because market segmentation makes it harder to improve absolute performance. Indeed, segmentation reduces the scope and range of activity of the manager and forces him to invest only in the assets specific to the fund's category, possibly hampering his market timing skills.[i]

A second important feature is the role increasingly played by marketing and segmentation strategies in an industry traditionally based on a homogeneous product. Indeed, fund proliferation has been accompanied by an explosion in the amount of information gathered by specialized magazines and publications *(Barrons, Business Week, Consumer Reports*) in response to investors' need to evaluate funds. Between the two events there seems to be a strong two-way causality: while the rise in the number of funds has increased the public desire for more information, more information has also made it increasingly difficult for mutual fund families to differentiate themselves purely on the basis of performance. This has led them to rely increasingly on alternative non performance-related strategies, in order to make inter-fund comparison harder and to raise market dispersion. In this context, fund proliferation becomes an additional tool that can be used to limit competition and increase market coverage. Indeed, we argue that the overall strategy of the family can be seen as a typical *brand proliferation* strategy designed to "occupy" market share and deter entry by rivals. [ii]

The issues of fund proliferation and imperfect competition in the mutual fund industry have gone largely unnoticed in the financial literature. In all, few attempts to model the mutual fund industry have been made. Dermine, Neven and Thisse (1991) lay out a model where mutual funds locate themselves on the portfolio frontier along with other primitive assets. Massa (1998) argues that market segmentation and fund proliferation can be seen as marketing strategies used by the families to exploit investors' heterogeneity. Category proliferation is justified in terms of the positive "spillover" that having a "star" fund provides to all the funds belonging to the same family. Nanda, Narayanan and Warther (2000) develop a model of the mutual fund industry where management fees and load fees are determined endogenously in a competitive setting. Investors' clienteles and heterogeneity in managerial skills originate different fee structures. Mamaysky and Spiegel (2001) starting from first principles derive the first equilibrium model of the mutual fund industry where funds are shown to exist in order to overcome investors' hedging needs. Mutual funds are analysed as trading devices set-up by investors who cannot remain in the market to trade at all times.

Empirical investigation has mostly focused on the ability of mutual funds to beat the market (Brown and Goetzmann, 1995,1997, Grinblatt and Titman, 1989,1994, Elton, Gruber and Blake, 1996, Elton, Gruber, Das and Hlavka, 1993, Ferson and Schadt, 1996, Wermers, 2000), the reaction of investors to mutual fund performance (Sirri and Tufano, 1998, Zheng, 1999), mutual fund managers' compensation and risk taking (Chevalier and Ellison, 1997, Brown and Starks, 1996), the relationship between fees and performance (Ippolito, 1989) and on the effect that capital gains tax has on investor demand and fund strategies (Barclay, Pearson and Weisbach, 1998). Only recently have mutual funds been studied as an industry. Khorana and Servaes (1999, 2001) empirically analyze the determinants of mutual fund starts, identifying a series of factors that induce the family to set up new funds, such as economies of scale and scope, fund family's prior performance and the overall level of fund invested. Ivkovich (2001) and Zheng (2001) focus on the impact of a fund's performance at the family level.

However, the link between market structure and family strategies has never been directly investigated, either from a theoretical perspective or from an empirical one. Nor has anyone studied the relationship between fund performance and market structure. Indeed, the standard paradigm assumes that all funds are perfect substitutes for one another, and investors simply pick the funds that provide them with the highest level of performance. In this context, the type of market structure does not play any role. This would not be the case were funds perceived as heterogeneous products that differed on the basis of a set of family-specific as well as investor-specific characteristics.

Family-specific characteristics pertain to the way investors evaluate funds that belong to different families, but are otherwise identical in terms of performance and fees. Possibly, the most relevant of these characteristics is generated by the fee structure.

Mutual fund investors face two types of fees: load fees, which are paid only once and are assumed to be the same across all funds of the same family, and a management fee, that varies for each fund. Sometimes the fund also charges "switch or exchange fees", that is, fees the investor has to pay to move his investments over two funds belonging to the same family. To avoid arbitrage opportunities, switch fees and load fees are usually structured in a way that the investor is made to pay an "investment fee" uniform with all the funds of the same family, and then he is free to move across them. As a result, load fees can effectively be considered as a uniform price that investors pay to get into the family as a whole, and that segments the mutual fund industry in terms of family affiliation.

The possibility of moving money in and out of funds belonging to the same family at very low cost, or at no cost at all provides investors with a benefit that reduces the effective fees they pay. Indeed, this can be seen as an option that the family provides to its investors: the higher the number of funds in the family, the greater the value of the option since the effective fees decrease as a function of the number of funds.

The existence of this option affects the degree of competition among funds, as investors will evaluate them in terms of the returns provided *and* the value of the free-switching option. The greater the value of the option, the lower is the degree of competition among funds and the greater is the segmentation of the industry.

Given that investors may have different investment horizons – time in the future when they want either to monetize their investment or to switch from one type of fund to another, the value of such an option is investor-specific. If investors have short-term investment horizons and/or plan to rotate their portfolios often, they will select a fund that is part of a family that manages many other funds, as they can get the new desired portfolio allocation by reinvesting within the family without incurring additional entrance fees. In this case, the value of the option will be high. If, on the other end, the investors have a long-term horizon, or plan a very low portfolio turnover, they will care less about switching among funds of the family. In this case, the value of the option will be low.

For example, an investor who is very risk averse, but is confident about short-term market performance, may decide to invest in equity funds in the short run, thinking of switching to bond funds in few months' time. In this case, the choice of the equity funds to invest will be affected by the availability of bond funds within the family and by the possibility of switching into them at low cost.

This means that two funds that provide the same return and have the same investment profile may cater to different investors. Their appeal will depend on family specific characteristics (i.e., the number of other funds offered by the same family) as well as investor specific ones (i.e., their investment horizon).[iii]

If investors are heterogeneous and funds are perceived as products differentiated on the basis of family- as well as investor-specific characteristics, families have many alternative tools available to them, in order to attract investors.

For example, a family may try to raise performance by increasing risk taking, investing more in research or providing its managers with better incentives. Alternatively, if the family realizes that the level of performance it offers is much lower than that of all its competitors and that it would be very costly to compete on the performance dimension alone, it will focus on other ways of attracting investors, such us reducing fees or increasing the number of funds within the family. That is, each fund will position itself on the basis of the fund-specific performance/fee/family combination it offers, and will target the investors whose preferences are closer in the product space to the offered combination.

This has important implications. First, performance-maximization is not necessarily the optimal strategy, and the success of alternative strategies will depend on both the structure of the market and its degree of competitiveness. In fact, the profit maximizing mix of fees, performance and number of funds might even involve a level of performance that would otherwise be defined as "inferior" in a standard performance evaluation analysis. Moreover, the level of performance of a fund has to be directly related to the degree of product differentiation in the category the fund is in – i.e., the degree of competitiveness of the industry. Finally, if fund proliferation is one of the possible strategies for the family, the number of funds and the number of categories where the family has funds should also be directly related to the degree of differentiation.

We will empirically analyze these issues by showing the existence of a strong and statistically significant relationship between performance and the degree of product differentiation and between the latter and the degree of fund proliferation. This is a first attempt to study how mutual funds behave in a framework where the interaction between a fund family and industry structure matters. We will show that the degree of product differentiation *negatively* affects performance and *positively* affects fund proliferation.

The paper is structured as follows. In Section 2, we lay out the framework and derive the main testable restrictions between performance, fund proliferation and industry characteristics. We describe the data in Section 3, and in Section 4 we outline the econometric approach we use. In Sections 5 we provide evidence of degree of competition of the mutual fund industry, by studying whether the investment horizon affects investor choice and whether there is evidence of family-driven segmentation. In Sections 6 and 7 we test the main working hypotheses. In particular, we show that the degree of product differentiation *negatively* affects fund performance and *positively* affects the degree of fund proliferation. We also show how these effects aggregate at the market level. A brief conclusion follows.

2 A general framework

We use a standard framework of product differentiation as a justification for the subsequent empirical tests. We wish to address the following questions: if investors have heterogeneous tastes, what is the optimal strategy we expect families to follow? How do they choose between fund proliferation and performance enhancing strategies? We therefore do not start from first principles, but we just assume investors' heterogeneity and try to see how families exploit it.

We assume that there are $\{1,...,i,...,I\}$ competing families of funds in the market. Each family chooses the number of funds to offer $\{1,...,m,...,M\}$. Funds are distinguishable according to the primary type of securities in which they invest (i.e., equity funds, bond funds, balanced funds, money market funds,) and/or the investment styles (i.e., growth funds, cash funds, income funds).[iv]

Investors

Each investor selects a fund on the basis of its characteristics and his "appreciation" of them. In particular, the net utility the *ath* investor derives from the *mth* fund belonging to the *ith* family, can be described as:

$$1)\ U_{im}^{a} = fund\ characteristics + consumer\ tastes = r_{im} - p_i(1 - B_i M_i) + \xi_{im}^{a},$$

where r_{im} is the performance of the *mth* fund, belonging to the *ith* family, and p_i are the load fees charged by the *ith* family. The term ξ_{im}^{a} captures this heterogeneity in preferences.[v] $B_i M_i$ represents the reduction in the fees thanks to the free-switching option, that is, the fact that the investor can freely switch from one fund to another within the same category at low cost, or at no cost at all. The value of this benefit is proportional to the number of funds belonging to the same family (M_i) through the coefficient B_i that defines the unitary (per fund) value of the option. The more funds the family has, the greater is the possibility of switching at no cost. If the investor plans to rotate his portfolio regularly, by shifting money from one fund to another, the value of B_i will be high. On the other hand, if he has a long-term horizon, with very low portfolio turnover, B_i will be low.

We assume that investors first select a particular family and then pick the fund which best suits their investment needs. This approach is supported by the empirical observation that a huge volume of assets under managements is in retirement accounts where the plan sponsor severely limits the family choices. Furthermore, this is consistent with the assumption that investors evaluate the funds belonging to a family on the basis of the free-switching option. If we assume that ξ_{im}^{a} is i.i.d. distributed as a double exponential, investor demand can be conveniently described in close form in terms of the probability of investment in the specific fund, that is: $P_{im}=P_iP_{m|i}$, where

$$2)\ P_{m|i}=\frac{\exp\phi_{im}}{\sum_{m=1}^{M_i}\exp\phi_{im}},\quad P_i=\frac{\exp(\frac{S_i}{\mu_1})}{\sum_{i=1}^{I}\exp(\frac{S_i}{\mu_1})},\quad S_i=\mu_2\ln\left\{\sum_{m=1}^{M_i}\exp\frac{\phi_{im}}{\mu_2}\right\}.$$

where $\phi_{im}=r_{im}-p_i(1-B_iM_i)$, P_{im} is the probability of investing into the *mth* fund of the *ith* family, P_i is the probability of investing into the *ith* family and $P_{m|i}$ is the probability of investing into the *mth* fund, *conditional* upon the choice of the *ith* family.

This provides a directly testable specification of investors' demand. By aggregating the probability of investing in a specific fund across all the investors, we have the market share of the fund. In particular, the probability of choosing a specific fund can be defined in terms of two parameters μ_1 and μ_2. μ_1 describes how different the investors perceive the fund families to be one from another, while μ_2 describes how different the investors perceive the funds belonging to the same family to be one from another.[vi] We will therefore define μ_1 as the degree of heterogeneity among different families, and μ_2 as the degree of heterogeneity among funds belonging to the same family.[vii]

Families of funds

In order to satisfy investors' desires, families behave as multi-product firms who decide on the number of funds to offer (M_i), the investment to make in order to generate better performance (c_{im}) and the fees to charge (p_{im}) on the basis of their ability to provide investors with an adequate bundle of utility (e.g., performance, fees) and the costs incurred (e.g., start-up costs). Their aim is to maximize profits:[viii]

$$3)\ Max_{p_{im},c_{im},M_i}\left\{\sum_{m_i=1}^{M_i}[p_{im}(1-C_iM_i)-c_{im}]P_{im}-k_i\right\},$$

where c_{im} is the cost incurred by the family to provide a certain level of performance. From the previous definition we also know that $P_{im}=f(p_{im,}c_{im},M_i)$. We may think of the "performance-enhancing" activity as research, that is the money spent to acquire information to “beat the market”[ix]. We define $A_{im} = \partial r_{im} / \partial c_{im}$ the increment of performance per unitary cost. It measures how efficient the family is in generating performance. C_i is the unitary (per fund) cost of providing the “free-switching option” to investors. This cost covers the forfeited load fees that the family does not earn when investors switch across its funds at low or no cost, as well as the loss in economies of scale that are incurred, since the family is simultaneously operating in many categories. The variable k_i represents the fixed costs incurred to enter the mutual fund industry.[x]

It can be shown that the level of performance for the funds that are part of the *ith* family, consistent with the aforesaid assumptions on investors’ behavior and families’strategies is:

$$4)\quad r_i = -\mu_1 + A_i\left[-k_i + \frac{\mu_2}{(A_i - 1)} + \frac{\mu_2 C_i}{(B_i - A_i C_i)} \right]$$

(see Appendix). Performance is a function of the supply side (i.e., the ability to generate performance, A_i, the cost of the free-switching option for the family, C_i, the value of the free-switching option for the investors, B_i, the cost of entering the industry, k_i and the level of fees charged, p_i) as well of the demand side (i.e., heterogeneity, μ_1 and μ_2). For example, the more investors value the free-switching option (i.e., greater B_i), the less the family will invest in performance. Conversely, the higher the cost of the option (i.e., greater C_i), the more the family will concentrate on performance. The same holds for the entrance costs (k_i): the higher they are, the fewer the competing families are, and the less the existing families have to worry in order to fend off potential entrants with higher performance.

Also, performance is *positively* related to the degree of heterogeneity across funds within the same family (μ_2) and *negatively* related to the degree of heterogeneity among families (μ_1). Indeed, the more a family is able to make investors perceive the funds it is managing as differentiated products (i.e., μ_2 is high), the less it will fear that higher performance of one of its funds may cannibalize the market share of the other funds of the same family. On the other hand, if family affiliation makes a big difference (i.e., μ_1 is high), the incentive to generate performance decreases. Indeed, if families are perceived to be different one from another, market segmentation is already making demand steep enough, investors can be attracted without investing in performance-enhancing activities, and the need to generate performance is lower. Equation 4) can also be rewritten as:

$$5)\ r_i = -\mu_1 + \mu_2 A_i \left[\frac{1}{(A_i - 1)} + \frac{C_i}{(B_i - A_i C_i)} \right] - A_i \Gamma,$$

where Γ represents an index of product heterogeneity of the industry.[xi] We will define Γ as “Index of product differentiation”. Γ is a function of the absolute value of the differences between the characteristics (fees and performance) of the funds operating in a category (see Appendix). The relationship between performance and product differentiation is *negative*. This applies to the case where differentiation is defined in terms of non performance-related characteristics, as well as to the case where it is defined in terms of performance.

Let’s assume that that differentiation is defined in terms of non performance-related characteristics The intuition is the same as the one that applies to markets with differentiated products: the more the firms are able to differentiate themselves, the less they directly compete on prices. That is, the more the families are able to differentiate themselves in terms of non-performance related characteristics, (e.g., higher degree of fee differentiation), the less they need to compete in terms of performance.

Analogous rationale applies where differentiation is defined in terms of performance. Each family decides the investment in performance enhancement, conditioning on the performance of the other families. If the competitors display a higher performance, the family has a lower incentive to invest in performance, and relies on a fund proliferation policy. At the other extreme, if a family’s performance is higher than that of its competitors, the incentive to further increase performance will drop. In the intermediate cases, when the family perceives that it can perform better that some competitors and worse than others, it will spend more on performance. Therefore, it does not make sense to invest in performance if the family is already well ahead of its competitors or if it is far behind them. In the former case, the cost would be prohibitive, in the latter, each additional investment in performance would have a negative return. It is worth investing in performance only where the investment provides the family with a cost-effective way of "beating" the competitors, i.e., where dispersion is low.

A second, direct implication is the relationship between fund proliferation (i.e., the number of funds being offered and the number of categories in which the family is simultaneously operating) and the degree of product differentiation. For exactly the opposite reasons of those provided before, we expect fund proliferation to be *positively* related to the degree of product differentiation in the category.

Therefore, our testable restrictions are: $\frac{\partial r_i}{\partial \Gamma} < 0$ and $\frac{\partial M_i}{\partial \Gamma} > 0$. We will now move on to test them.

3 The data

3.1 Category selection and variables used

The data come from the CRSP Mutual Fund files. CRSP is a survivorship-bias free dataset that contains data for 18,616 funds. CRSP contains three potential sets of categories. The first one is based on investment objectives ("OBJ") and relies on the Weisenberger categorization. The second is based on the ICDI fund objective codes (*"ICDI_OBJ")*. The third set of categories is based on the strategic insight of the (ICDI) fund's objective code ("*SI_OBJ")*.

The three types are available for different periods and with different coverage. However, not all the funds have a clearly defined category. In particular, *OBJ* was used during 1962-1990 and the CRSP dataset does not yet list all the *OBJ* codes even for this limited period. *OBJ* and *SI_OBJ* start later on, in the '90s. 1991-1992 is the period of data intersection of the three types of categories. If year 1992 is taken as the first year of introduction of *ICDI_OBJ*, the number of missing *ICDI_OBJ* is insignificant.[xii] Hence, we use this criterion to define the different categories, and we limit our analysis to the period between 1992-2000. Table 1 shows some descriptive statistics for the data. Of the 18,616 funds, 928 existed between 1961-1990 and 17,688 existed from 1993 onwards. The remaining 17,688 unique funds are categorised by *ICDI_OBJ*. Of these funds, 760 do not have a clear code in CRSP. We therefore use the Morningstar files to find the missing categories, by matching the funds in CRSP with those in Morningstar. There are 24 unique categories for these 17,688 funds. Out of these sets of funds there are 388 Index Funds. Given their peculiar characteristic, they are not considered.

The 24 categories are: Aggressive Growth (AG), Balanced (BL), High Quality Bonds (BQ), High Yield Bonds (BY), Global Bonds (GB), Global Equity (GE), Growth and Income (GI), Ginnie Mae (GM), Government Securities (GS), International Equities (IE), Income (IN), Long-Term Growth (LG), Tax-Free Money Market (MF), Government Securities Money Market (MG), High Quality Municipal Bond (MQ), Single-State Municipal Bond (MS), Taxable Money Market (MT), High-Yield Money Market (MY), Option Income (OI), Precious Metals (PM), Sector Funds (SF), Special Funds (SP), Total Return Fund (TR) and Utility Funds (UT). We drop the Option Income category as it contained only 1 fund during the period.

Performance, net asset values and flows come at a monthly frequency, while fees, total income, capital gains distribution and yields come yearly. The fees include: maximum load fees (i.e., the maximum front-end load fees), maximum contingent deferred sales, rear load fees (i.e., the redemption fees), 12b-1 fees, other load fees (i.e., maximum deferred sales charge and redemption fees).

We consider data at monthly and yearly frequencies, disaggregated at both the fund level and the category level. For all the variables (returns, risk, different types of fees) we construct averages across funds belonging to the same family, as well as averages across all funds within a specific category, regardless of family affiliation. We aggregate across funds by using simple and weighted averages. Average yearly return mean, standard deviation and skewness are, respectively, the mean, standard deviation and skewness of the returns over the previous twelve months with rolling windows. Monthly flows are constructed as *Flow = TNA – TNAPrior – TNAPrior * Ret,* where *TnaPrior* is the total net asset value of the previous month, *TNA* is the net asset value at the end of the month and performance is the return on the managed assets in the specific month. Descriptive statistics of the categories in terms of performance, expense ratio and number of funds and families active, are provided in Table 1, Panels A, B and C.

3.2 Construction of the index of product differentiation

We exploit the cross-sectional dimension of the mutual fund industry by considering each category as a different market. That is, we construct an index of product dispersion for each category, that proxies for the dispersion of the characteristics of the funds belonging to the *cth* category at time *t*. The index has to capture the main features of Γ, at the category level. That is, it has to be low when dispersion is low, and increase as dispersion increases.

In order to construct the index, we use the observable characteristics that the funds use to differentiate themselves: non performance-related characteristics (fees), and performance-related ones (performance). In particular, we construct three different proxies for Γ. The first proxy uses only non performance-related characteristics, that is the dispersion of fees across all the funds in the category. The second proxy uses the performance-related characteristics, that is the dispersion of the returns and standard deviations of the returns of the funds in the category. The third proxy uses both the performance-related characteristics and the non performance-related ones, as defined before, along with the dispersion of distributed income, the dispersion of the capital gain distributions and the dispersion of the average yields.

The proxies are constructed as follows. First we calculate the standard deviation of the characteristics (e.g., fees charged) across all the funds active in the category. This is done separately for each type of characteristics every year. Then, the resulting components of the index are standardized so as to make them homogeneous in terms of range of variation. In particular, we consider two possible standardizations. According to the first one, we divide the single components by their maximum value over the whole period. Alternatively, we divide them by the average value of the series constructed as the mean of the characteristics across all the funds active in the category. The first standardization provides a new series that represents a measure of deviation over time from the long-run average, while the second one simply rescales on the basis of the mean value. We will report the results for the two measures in Table 1, Panels A and B respectively ("market differentiation I and II").

We also consider different ways of aggregating the components of the index, using the simple average of its components as well as the weighted average, with the total net assets of the funds as weights. We will refer to the first type as “weighted index” and to the second as “non-weighted index”.

The fees we consider are: maximum front-end load fees, maximum contingent deferred sales charges, rear load or redemption fees, 12b-1 fee, maximum deferred sales charges and redemption fees. We also include the expense ratio. The indexes are constructed at monthly and yearly frequencies.

One possible objection is that our measure of differentiation is actually proxying for the *ex post* dispersion of returns across funds due, for instance, to the degree of riskiness of the particular category instead of the degree of product dispersion. Indeed, the more uncertain the returns in a particular category are, the more likely it is that different funds will display different level of performance *ex post*, just by fluke. This would generate spurious correlation between performance and our measure of dispersion. To control for this, we consider specifications where we directly include among the control variables the standard deviation of returns in the previous twelve months of both the specific fund and all the funds active in the same category. This should account for the uncertainty of the category. Furthermore, the specification with non performance-related characteristics is unlikely to be a direct proxy for riskiness. Descriptive statistics of the indexes are reported in Table 1, Panel D.

4 Econometric issues

Before turning to the empirical section, let us briefly consider the econometric complexity of the problem we face, its potential implications and the methodology we use to tackle it. We will test our hypotheses by using a panel data. However, in our case the econometric estimation is particularly complex, as we face four different problems: possible endogeneity of the explanatory variables, the existence of a lagged dependent variable and company specific effects (fixed effects), serial correlation and cross-sectional heteroskedasticity. Let us see how to deal with them. The first issue is that of endogeneity. Let us consider the equation:

$$6)\ y_{i,t} = \beta x_{i,t} + \eta_i + \varepsilon_{i,t}\,,$$

where η_i represents a company specific effect (fixed effect). Exogeneity of x_{it} would require that $E\left(\varepsilon_{it} \mid x_i^T\right) = 0$ for (t = 1,…,T), that is, x_{it} to be uncorrelated with past, present and future values of the disturbance $\varepsilon_{i,t}$. In other words, for x_{it} to be exogenous we need it to cause y_{it}, but not be affected by y_{it}.

However, the economic problem we are facing does not allow us to determine a priori which variables are exogenous. In our case y_{it} will alternatively be different measures of fund performance (at fund level, company level and category level) as well as measures of fund proliferation (number of funds, number of categories the family is active in, fund turnover in the category), while x_{it} will represent our measures of product differentiation. Indeed, it is possible that higher performance induces funds to segment and differentiate themselves. But it is also possible that differentiation leads performance.

If investor funds flow into securities with higher performance, we would also expect fund families to respond by creating funds in those areas. Therefore, if fund dispersion is measured in a dataset that includes stock performance from either the same or earlier periods, it is possible that stock's return is positively correlated to the diversity of the funds trading it.[xiii]

In the last section we will perform Granger causality tests and show that dispersion Granger leads performance and not viceversa. However, in order not to make the results of the paper dependent on these Granger tests and to increase the robustness of our estimations, we will not make any assumption about endogeneity now. We will only make the milder assumption of predeterminedness. x_{it} is defined as being predetermined if $E\left(\varepsilon_{it} \mid x_i^t, y_i^{t-1}\right) = 0$ for (t = 2,…T), that is "current shocks are uncorrelated with past values of *y* and with current and past values of *x*, but feedback effects from the lagged dependent variables (or lagged errors) to current and future values of the explanatory variable, are not ruled out" (Arellano, 2001). That is, we assume the explanatory variables not to be correlated with future realizations of the error term. This implies that current explanatory variables can be affected by past and current values of performance, but not by future ones.

This allows for the possibility of reverse causality and simultaneity. Therefore, in order to account for possible feedback effects from past values of the dependent variable, we will estimate the following dynamic panel specification:

$$7)\ y_{i,t} = \alpha y_{i,t-1} + \beta x_{i,t} + \eta_i + \varepsilon_{i,t}\ .$$

While this specification properly controls for endogeneity and simultaneity issues, it raises an additional concern. Indeed, the existence of the lagged dependent variable, coupled with company specific effects (i.e., η_i), makes the standard panel estimation techniques - i.e., OLS, Within Estimators ("fixed effect estimation") and Between Estimators ("random effect") - *biased and inconsistent*. This is due to the fact that the lagged dependent structure induces a correlation between the $x_{i,t}$ and the η_i. Therefore, a different approach is required.

One way of dealing with company specific effects in the presence of lagged dependent variables is to take first differences and then apply the instrumental Generalized Method of Moments (GMM) estimator (Anderson and Hsiao 1981). By taking first differences, we eliminate the company specific effect. That is:

$$8)\ y_{i,t} - y_{i,t-1} = \alpha(y_{i,t-1} - y_{i,t-2}) + \beta(x_{i,t} - x_{i,t-1}) + (\varepsilon_{i,t} - \varepsilon_{i,t-1})\ .$$

However, in the differenced form, the new errors $(\varepsilon_{i,t} - \varepsilon_{i,t-1})$ are now correlated to the differenced lagged dependent variable $(y_{i,t-1} - y_{i,t-2})$ by construction. Therefore, instrumental variable estimation is required.

The endogeneity issues we mentioned before further complicate the task of finding proper instrumental variables. While current values of the explanatory variables can be used as instruments only in the case where such variables are exogenous, only lagged values can be used in the case where the variables are predetermined. In particular, in the case where $x_{i,t}$ are predetermined, but not strictly exogenous, only lagged values of $x_{i,t}$ are good instruments. If, on the other hand, $x_{i,t}$ are strictly exogenous, we can use both current and lagged values of $x_{i,t}$ as instruments.

We will therefore use only lagged values of $x_{i,t}$ as instruments. We rely on the findings of Arellano (1989), Arellano and Bond (1991) and Kiviet (1995) who show that the best instruments are provided by the lagged explanatory variables, *provided that no autocorrelation exists*. In this case, with no serial correlation, we can define the following set of moment conditions:

$$9) \quad \begin{cases} E\left[y_{i,t-s}(\varepsilon_{i,t}-\varepsilon_{i,t-1})\right]=0 \\ E\left[x_{i,t-s}(\varepsilon_{i,t}-\varepsilon_{i,t-1})\right]=0 \end{cases} \quad \text{for } s\geq 2,\ t=3,\ldots T$$

The GMM estimate of this set of moment conditions delivers unbiased and consistent estimates, even in the presence of endogeneity of the explanatory variables and in the case of simultaneity and reverse causality.

However, these conditions require the absence of serial correlation. This brings us to the next additional problem: *serial correlation*. Given that the very process of first differencing may induce serial correlation, in order to address this issue, we use an estimator based on Orthogonal Deviations Estimator. This methodology, first suggested by Arellano (1988) and Arellano and Bover (1995), consists of constructing for each observation the deviation from the average of future observations in the sample for the same individual and weighting each deviation in order to standardize the variance. The new “ortogonalized” variable is:

$$10) \quad x^{*}_{i,t}=\left(x_{i,t}-\frac{x_{i,t+1}+\ldots+x_{i,T}}{T-t}\right)\left(\frac{T-t}{T-t+1}\right)^{1/2} \quad \text{for} \quad t=1,\ldots,T\text{-}1.$$

This provides additional advantages, as, unlike First Differences, Orthogonal Deviations preserve the lack of correlation if the original variables are not autocorrelated. In particular, the transformed errors are serially uncorrelated and homoskedastic if the original errors are serially uncorrelated and homoskedastic. That is, constructing Orthogonal Deviations does not induce the serial correlation that is induced by the process of differencing. “Orthogonal Deviations can be regarded as the result of doing first differences to eliminate fixed effects, plus a GLS transformation to remove the serial correlation induced by differencing” (Arellano and Honore, 2001).

Furthermore, in order to check for the serial correlation, we estimate and report the test statistics designed by Arellano (1990, 1993) and Arellano and Bond (1988). These tests are based on the standardized average residual covariances which are asymptotically distributed N(0,1) under the null of no autocorrelation. We report the tests of first-order and second-order serial correlations which are constructed on the differenced variables. The absence of serial correlation would require the differenced residuals display existence of first order negative serial correlation and absence of second order correlation.

In order to test the robustness of the estimation, we also report the Sargan test of over-identifying restrictions. The Sargan test is asymptotically distributed as a chi-square with as many degrees of freedom as the over-identifying restrictions under the null hypothesis of the validity of the instruments.

We will see that in all the specifications, there is evidence of first order negative serial correlation and absence of second order correlation, as required, and Sargan fails to reject the null of over-identification. This supports the robustness of our instrumental estimation and the validity of the chosen methodology.

On the basis of the aforesaid reasons, in order to perform our estimation, we use a Generalized Method of Moments Dynamic Panel estimator that controls for unobserved company and fund specific characteristics, the endogeneity of the explanatory variables and the use of lagged dependent variables. Finally, in order to control for heteroskedasticity, we use a consistent variance-covariance matrix, by applying White's correction.

5 Investors' heterogeneity and degree of competition in the mutual fund industry

Our approach to understanding the behavior of a family of funds relies on a critical assumption: investors differ in terms of their investment horizons. This leads them to evaluate load fees differently, depending on the frequency with which they plan to rotate their portfolios (e.g., from bond funds to equity funds). The shorter their investment horizon, the more they try to avoid funds with high load fees, and the more they prefer funds belonging to families that are part of a bigger family, so that they can change their investment profile at low or zero cost. This implies family-driven heterogeneity. That is, otherwise identical funds operating in the same category and with the same investment policy are perceived to be different, depending on the family to which they belong.

We will now focus on these two hypotheses. We will assess whether the reaction of investors is indeed related to their investment horizon and whether there is evidence of family-driven heterogeneity among mutual funds.

5.1 Investment horizon and investment decision

If an investor evaluates the attractiveness of a fund on the basis of the "effective fees", i.e., the fees he pays during his investment, his choice of a fund should depend on the number of times he thinks he will rebalance his portfolio and the number of funds among which he can choose at no cost (i.e. funds offered by the same family). That is, we expect that investors who are planning to reallocate their assets more frequently will tend to invest in funds with lower load fees and in funds that belong to bigger families.

Given that we cannot directly observe investors' investment horizon, we will use as proxy for it the standard deviation of the flows in the funds. The hypothesis to test is whether the funds that either charge lower load fees or are members of big families attract less stable investors. If this were the case, we would expect the standard deviation of the flows in and out of the funds to be *negatively* related to the level of the load fees and *positively* related to the number of funds belonging to the same family. We therefore estimate:

$$11)\ VF_{m,i,t} = \alpha + \beta F_{i,t} + \gamma N_{i,t} + \boldsymbol{\delta}\mathbf{Controls}_{\mathbf{m,i,t}} + e_{m,i,t}\,,$$

where the subscripts, "*m*", "*i*" and "*t*" refer, respectively, to the *mth* fund, belonging to the *ith* family at time *t*. $VF_{i,t}$ is the standard deviation of the flows of the fund over the twelve months in the respective year, while $N_{i,t}$ is the number of funds of the family across all the categories. $F_{i,t}$ represents the total load fees charged by the *ith* family at time *t*. We consider alternatively, the total load fees charged by the fund (specifications I and IV), the average total load fees charged by the family for all the funds belonging to the same category (specifications II and V) and the average total load fees charged by the family for all the funds it has across all categories (specifications III and VI). $\mathbf{Controls}_{\mathbf{m,i,t}}$ is a vector of control variables. They include: the total net assets of the fund as well as the overall total net asset of the family the fund belongs to, the average flows in and out of the fund, the performance and risk of the fund as well as the average overall performance of the family and the performance and risk of the competing families. Risk is constructed for each fund as the standard deviation of the returns in the previous 12 months.

Including total net assets allows us to properly control for size. Indeed, larger funds may have fund flows with a larger standard deviation and may be more likely to belong to larger fund families. It is also worth noting that by including both total net assets managed by the family in the category and total net assets across categories, together with the level of flows in and out of the fund, we are making the test very demanding.

Risk is also particularly relevant as the variation of the flows may be partly due to the volatility of the returns of the fund and of the family. In order to account for market-wide fluctuations, we also include the average performance and risk of the funds belonging to the competing families, both within the same category and overall across all categories.

We test the joint hypothesis that the standard deviation of the flows is *negatively* related to the level of load fees and *positively* related to number of funds belonging to the same family, i.e.:

$$H_o : \beta = 0 \text{ and } \gamma = 0 \text{ and } H_A : \beta < 0 \text{ and } \gamma > 0.$$

The results are reported in Table 2, Panel A. All the results, for all specifications, support our hypothesis. Funds belonging to families with high load fees tend to have more stable investors ($\beta < 0$). The relationship is strongly significant and holds in the case where the fees are constructed as averages across funds belonging to the same family and active in the same category (Specifications I and III). It also holds in the case where they are constructed as averages across funds belonging to the same family and active in the different categories (Specifications II and IV). This also holds at the overall level (Specification V). This seems to suggest that load fees affect the stability of the flows in and out of a particular fund.

Also, there is evidence that funds belonging to families that offer more funds tend to have less stable investors ($\gamma > 0$). Indeed, in all the specifications, there is a strong *positive* correlation between the number of funds belonging to the family and the standard deviation of the flows. These results support our hypothesis at the fund level.[xiv]

Let us now try to see whether these results aggregate at the family level. In other words, we want to find out whether a family that charges higher load fees or has the bigger market coverage experiences higher stability of flows. We therefore re-estimate equation 11) considering as a dependent variable either the standard deviation of the flows of the funds managed by the *ith* family in the *cth* category ($VF_{i,t|c}$) or the standard deviation of the flows of all the funds of the family across all categories (VF_{it}). $F_{i,t}$ represents the Total Load Fees charged by the family on the funds in the category or on all the funds of the family overall. $N_{i,t}$ is the degree of market coverage. That is, it represents the number of categories in which the family is operating. This variable proxies for the ability of the family to span many different investment dimensions.

We use a set of control variables analogous to those used before.[xv] The results are reported in Table 2, Panel B, when we focus on the family at the category level, and Panel C, when we focus on the family overall. The results are consistent with those at fund level. Indeed, there is evidence of a strong *negative* correlation between load fees and the standard deviation of the flows and of a *positive* correlation between the market coverage of the family and the standard deviation of the flows. It is worth noting that these results hold even after we control for fund and family size.

These results support the first of our hypotheses. That is, investors with a shorter and/or more volatile investment horizon tend to go for the funds with lower load fees and which are part of big families. Let us now move on to see whether there is also evidence of family-driven heterogeneity in the mutual fund industry.

5.2 Family-driven heterogeneity in the mutual fund industry

Testing for the existence of family-driven heterogeneity implies testing whether family affiliation affects investors' demand. Therefore, unlike previous studies (Sirri and Tufano, 1998, Warther, 1996, Zheng, 1999), we have to estimate the demand *by explicitly accounting for the possibility of heterogeneity.*

We start from the reduced form equation that represents the probability of the purchase of a fund. It can be shown (see Appendix) that:

$$12)\ \ln(P_{im}) - \ln(P_0) = \frac{1}{\mu_1}\delta_{im} + \frac{\mu_1 - \mu_2}{\mu_1}\ln(P_{m|i}),$$

where P_{im} is the probability that investors choose the *mth* fund of the *ith* family, $P_{m|i}$ is the probability that investors choose the *mth* fund *conditional* upon the choice of the *ith* family, and P_0 is the probability that investors choose other financial investment opportunities alternative to investing in mutual funds. Let us for instance, assume that investors are willing to invest in money market funds and are attracted by Fidelity. In this case, P_{im} is the *unconditional* probability of choosing, let's say, Fidelity Money Market Fund, while $P_{m|i}$ is the probability that the investor chooses Fidelity Money Market Fund, *conditional* on having decided to invest in Fidelity. P_0 represents the probability that the investor chooses another form of investment (e.g., direct investment in equity).

The vector δ_{im} contains all the characteristics of the funds on which investors condition their decisions, while the parameters μ_1 and μ_2 represent, respectively, the degree of heterogeneity across families and the degree of heterogeneity across funds run by the same family. If $\mu_1 = \mu_2$, the degree of heterogeneity among funds of the same family is the same as the degree of heterogeneity among families. If $\mu_1 > \mu_2$, the degree of heterogeneity among funds run by the same company is greater than the degree of heterogeneity among families. This implies that investors perceive two funds operating in the same category but run by different companies as more different than two funds run by the same company, but operating in different categories. In other words, family affiliation matters more than investment policy. If, on the other hand, $\mu_1 < \mu_2$ the investment policy of the fund is perceived as the discriminating factor, as opposed to family affiliation.

Most of the studies of mutual fund demand have implicitly assumed that family affiliation does not matter (i.e., μ_1=0) and only investment policy and performance matters (i.e., μ_2>0) (Sirri and Tufano, 1998, Zheng, 1999). Only recently has family performance been considered as one of the factors affecting investors demand (Ivkovitz, 2001, Nanda, Wang and Zheng, 2001). However, in these studies, family affiliation is considered as just one of the many characteristics that affect investor demand, but fund specific characteristics and performance still play the major role (i.e., $\mu_2 > \mu_1 > 0$). We claim that it is possible that investors are paying more attention to the family the funds belong to than to the fund specific characteristics (i.e., $\mu_1 > \mu_2$).

The empirical analogue of equation 12) is:

$$13)\quad \ln(s_{im}) - \ln(s_0) = \mathbf{a\delta}_{im} + b\ln(s_{m|i}),$$

where s_{mi} is the market share of the *mth* fund, belonging to the *ith* family at time *t*, $s_{m|i}$ is the market share of the *ith* family at time *t* and s_0 is the market share of the "outside alternative", that is the market share of all the other possibilities of investment the investors have, alternative to investing in mutual funds. It controls for outside market variations (i.e., exogenous shocks). Equation 13) lends itself nicely to the testing of the three aforesaid hypotheses. Indeed, the alternative hypotheses are captured by the coefficient *b*. That is:

$$\mu_1 > \mu_2 \quad \text{if} \quad b > 0, \quad \mu_1 = \mu_2 \quad \text{if} \quad b = 0, \; \mu_1 < \mu_2 \quad \text{if} \quad b < 0.$$

In order to estimate equation 13) we use a panel approach. In particular, we estimate:

$$14)\quad \ln(s_{mi,t}) - \ln(s_{0,t}) = \mathbf{a\delta_{mi,t}} + b\ln(s_{m|i,t}) + c[\ln(s_{mi,t-1}) - \ln(s_{0,t-1})] + \varepsilon_{mi,t}.$$

This specification is based on the approach of Berry (1994) and Berry, Levinson and Pakes (1995) and allows us to exploit the information contained in the market share. The vector $\boldsymbol{\delta}_{\mathbf{mi,t}}$ is a matrix that contains the vectors of the characteristics of the funds that are observable to the econometrician, that is, $\delta_{mi,t} = r_{mi,t} - p_{i,t}(1 - B_i M_{i,t})$. The characteristics that we use are: the performance of the fund, its risk, the performance of the family it belongs to, the overall risk of such a family, the total load fees charged by the family. The unobservable (to the econometrician) characteristics ($\xi^a_{im,t}$) contained in equation 1) are part of the error term ($\varepsilon_{mi,t}$).

The definition of the market share of a mutual fund is quite tricky. Indeed, the market share corresponds to the probability that an investor would purchase the *mth* mutual fund belonging to the *ith* family. In a standard Industrial Organization model, this would be proxied by the actual sales of the product under consideration. However, in the case of mutual funds, investors have the option to divest from the fund at any time. Therefore, flows capture only part of the picture and a better proxy would be the total net assets of the fund. These represent not only the actual purchases, but also the potential demand of investors who, having the option, do not withdraw money from the fund.[xvi]

We work at the category level and aggregate all the funds that the family has in a particular category. We construct $s_{mi,t}$ as the ratio between the assets of all the funds the *ith* company has in the *mth* category and the overall assets (i.e., mutual funds and other investment opportunities available in the market) existing in the market at time *t* and $s_{m|i,t}$ as the ratio between the assets of all the funds the *ith* company has in the *mth* category and the overall assets of the *ith* family at time *t*. s_0 is the market share of the other investment opportunities available in the market defined as a fraction of the overall assets. We consider as a proxy for the other investment opportunities available in the market the part of the total US financial wealth not invested in mutual funds.[xvii]

Equation 14) is a Dynamic Panel Data specification. We estimate it by using GMM with instrumental variables. In order to choose the instruments, we exploiting the restrictions imposed by the Markov Nash equilibria (Appendix) and using as instruments the combinations of the characteristics of the competing fund families (Berry, 1994 and Berry, Levinson and Pakes, 1995). We consider two alternative specifications: one is based on first differences and the other on orthogonal deviations. White's corrected robust variance-covariance matrix is used to control for heteroskedasticity.

The results are reported in Table 3, Panels A and B. They strikingly agree. The value of *b* is less than 1, and in general around 0.3. This confirms our intuition that investors perceive funds as differentiated products and that family affiliation plays a major role in segmenting the market. It is also worth noting that investors' demand is strongly *negatively* related to the level of the load fees and *positively* related to the number of funds offered by the family. This provides a further confirmation of our working hypothesis and agrees with results on the demand of mutual funds (Sirri and Tufano, 1998, Zheng, 1999).

Also, in all the specifications, Sargan fails to reject the null of over-identification, supporting the robustness of our instrumental estimation. We now move on to the relationship between the degree of product differentiation, performance and family proliferation.

6 Performance and product differentiation.

We focus on equation 5), which directly links performance to the degree of product dispersion ($\Gamma_{c,t}$) in each category. If a relationship between the two variables exists and is the result of a deliberate strategy of the family, we would expect this relationship to hold both at a fund level and at the family level. That is, we expect a relationship between the average performance of the funds of a family and the degree of differentiation of the category they are in. Furthermore, we expect this relationship to aggregate at the family level, so that the performance of the family should be affected by the degree of differentiation of the categories where its funds are. Therefore, we carry out the analysis at four levels.

First, we test whether the performance of the funds is related to the degree of differentiation of the category they are in. Then, we see whether this relationship holds at the family level. That is, we see whether the performance of a family in a category is related to the degree of differentiation of the category in which it is operating. Third, we test whether the overall performance of a family is related to the average degree of differentiation of the categories in which it is operating. Finally, we assess whether the impact of product differentiation aggregates at the category level. That is, whether categories characterized by higher degree of product differentiation systematically provide lower performance. Let us start by considering the individual funds.

6.1 An analysis at the fund level.

We are interested in assessing whether there is a relationship between the performance of a fund and the degree of product differentiation of the category it operates in. The empirical analogue of equation 5) is,

$$15)\quad r_{mic,t} = \alpha + \beta\Gamma_{c,t} + \boldsymbol{\gamma}\mathbf{Controls}_{\mathbf{mic,t}} + \delta r_{mic,t-1} + e_{mic,t},$$

where the subscripts “*m*”, “*i*”, “*c*” and “*t*” refer, respectively, to the *mth* fund of the *ith* family in the *cth* category at time *t*. In particular, $r_{mic,t}$ is the performance of the *mth* fund of the *ith* family, belonging to the *cth* category at time *t,* and $\mathbf{Controls}_{\mathbf{mic,t}}$ is a vector of control variables. $\Gamma_{c,t}$ is the index of product differentiation for the *cth* category at time *t*. We consider three different ways of constructing the index of differentiation for each category ("non-performance based", "performance-based", "mixed"), two different ways of aggregating its components: ("weighted index" and "non-weighted index"), as well of two alternative standardizations ("market differentiation I and II").

Control variables are used to assess the robustness of the results. In particular, we want to see whether our measure of differentiation survives in a horse race against all the other factors that the traditional literature has related to fund performance. Therefore, the control variables include the total load fees and risk of the fund as well as the overall average performance and risk of the family the fund belongs to. Also, in order to control for market conditions, we include not only the overall average total net assets of the family the fund belongs to and those of the competing families, but also the average performance and risk of the competing families, both overall and in the category.[xviii] The average risk is constructed as the average of the standard deviation of the twelve monthly returns. Averages are taken either across the funds operating in the same category ("in the category") or across all the funds of the same family over all categories ("overall"). The observations are yearly.

The estimations are carried out by using a Dynamic Panel Data estimation, based on GMM, with instrumental variables and White's adjusted heteroskedastic consistent least-squares variance-covariance matrix. The model is estimated in Orthogonal Deviations. Lagged levels and differences of the explanatory variables are used as instruments. The hypothesis we test is:

$$H_0 : \beta = 0 \quad \text{and} \quad H_A : \beta < 0 .$$

That is, we expect a *negative* correlation between performance and the degree of product differentiation.

The results are reported in Table 4, Panels A, B and C. They show a strong and statistically significant *negative* relationship between fund performance and the degree of product differentiation in the category. Furthermore, the degree of product differentiation is also strongly significant, even after all the possible control variables are added. This is all the more striking if we consider that we are controlling for all the other variables that the literature has assumed to be related to performance. Also, this relationship holds across all the specifications, regardless of how we constructed the index of market differentiation.

It is worth noting that, while we do not have a direct measure of “extra-performance”, the inclusion of the performance of the other competing funds, as well as a measure of risk of both the family and the competing families, allows us to properly control for it. Furthermore, the results are robust to the different sets of control variables, including performances of the other funds in the category. This provides a particularly demanding test. Also, the Sargan tests fail to reject the over-identifying restrictions and the tests of serial correlation reject serial correlation. This supports the validity of the choice of the instruments.

These results provide a first evidence of the impact of the market structure on mutual fund performance. We can now move on to see whether the relationship between performance and market differentiation aggregates at the family level.

6.2 An analysis at the family level.

We now consider the family and look at whether there is a relationship between the performances of the funds it manages and the categories they operate in. We therefore estimate:

$$16)\quad r_{i,c,t} = \alpha + \beta\Gamma_{c,t} + \boldsymbol{\gamma}\mathbf{Controls}_{\mathbf{ic,t}} + \delta r_{i,c,t-1} + e_{i,c,t},$$

where $r_{i,c,t}$ is the average performance of all the funds the *ith* family has in the *cth* category at time *t*, and $\Gamma_{c,t}$ is the index of product differentiation for the *cth* category at time *t*.

As before, we use different specifications based on various ways of constructing the index of differentiation. The control variables ($\mathbf{Controls}_{\mathbf{ic,t}}$) include the risk and expense ratio of the family in the category, the overall family performance, the risk and expense ratio across all the categories, the performance, the risk and expense ratio of the competing families in the category as well as overall across categories. We also consider the total load fees charged by the family as well as those charged by the competing families. The estimations are carried out by using a Dynamic Panel Data estimation, based on GMM, with instrumental variables.[xix] The results are reported in Table 5, Panels A and B.

They display a strong and statistically significant *negative* correlation between performance and degree of product differentiation, for all the specifications and for the different ways of constructing the index of differentiation. Also in this case, the validity of the instruments is supported by the Sargan tests and the tests of serial correlation.

This provides further evidence of families' strategic behavior. We now consider the overall family performance in order to see whether this strategic behavior directly impacts family-wide performance. We therefore estimate:

$$17)\quad r_{i,t} = \alpha + \beta\Gamma_{i,t} + \boldsymbol{\gamma}\mathbf{Controls_{i,t}} + \delta r_{i,t-1} + e_{i,t},$$

where $r_{i,t}$ is the average performance of the all the funds that the *ith* family manages across all the categories at time *t*, and $\Gamma_{i,t}$ is the index that proxies for the average degree of product differentiation of all the categories the families has funds in at time *t*.

The family-specific index of product differentiation is constructed by weighting the indexes of all the categories the family is operating by the total net assets of the funds the family is managing in such categories. As before, we use different specifications based on various ways of constructing the index of differentiation and a set of control variables. [xx] The estimations are carried out by using a Dynamic Panel Data methodology based on GMM with instrumental variables.[xxi]

As in the previous tests, we expect $\beta_i < 0$. The results are reported in Table 6, Panels A and B. All the specifications agree, and support the working hypothesis of a *negative* correlation between performance and degree of product differentiation. The Sargan tests and the tests of autocorrelations support the choice of the instruments and confirm the validity of the estimation.

Furthermore, the fact that the results are the same across different specifications, regardless of the way the indexes of product differentiation have been built and after controlling for several measures of risk, suggests that our results are due to product differentiation as opposed to risk.

If product differentiation affects the strategies of the family and the optimal level of performance, the crucial question is whether this impact aggregates at the industry level. In other words, is it the case that categories characterized by lower product differentiation – i.e. higher degree of competition – generate in equilibrium higher performance? If this were indeed the case, this would have potential important implications in terms of asset pricing. We therefore move on to address this issue.

6.3 Category performance and product differentiation

We want to see *how product differentiation affects the performance of the overall category*. That is, we want to determine whether the impact of product differentiation is noticeable at the aggregate level. In particular, we test whether its impact aggregates at the industry level. This would imply that categories characterized by higher degree of product differentiation should systematically provide lower performance.

We analyze these two issues separately. In particular, first, we test whether there is a *negative* correlation between the degree of product differentiation and performance at the category level. Then, we investigate whether there is evidence of causality, i.e. if product differentiation drives performance or viceversa.

We start by focusing on the average performance across the funds within a particular category and test the relationship between the degree of product differentiation in a category and the average performance displayed by all the funds belonging to such a category. In particular, we estimate:

$$18)\ \ r_{c,t} = \alpha + \beta\Gamma_{c,t} + \boldsymbol{\gamma}\mathbf{Controls}_{\mathbf{c,t}} + \delta r_{c,t-1} + e_{c,t},$$

where $r_{c,t}$ is the average performance of the all the funds in the *cth* category at time *t* and $\Gamma_{c,t}$ is the index that proxies for the degree of product differentiation of the *cth* category at time *t*. Different ways of constructing the index of differentiation are considered.

The control variables include the three Fama and French factors (i.e., Market, SMB and HML) and the riskless rate (i.e., return on the 3-month T-Bill rate). These variables are meant to replicate market conditions, and are used as benchmarks in the estimation of mutual fund performance. We also include the average risk of the category. This captures the specific sources of idiosyncratic characteristics of the specific category. Indeed, as

Grinblatt and Moskowitz (2000) and Campbell (2000) have shown, a vast part of market volatility can be explained in terms of industry factors. In our case, the industry factors pertain to the mutual fund category. The average risk is constructed as the average of the standard deviation of the twelve monthly returns of all the funds in the category. Averages are taken across all the funds in the category.

We perform a Dynamic Panel Data estimation, based on GMM, with instrumental variables using the same methodology outlined before. However, unlike before, we now use monthly data. This makes the test particularly demanding. Indeed, while performance and risk can be estimated with monthly frequency, the index of product differentiation based on non performance-related characteristics relies on data available only with yearly frequency.[xxii] This implies that most of the explanatory power of the index is due to the cross-sectional dimension.[xxiii] The hypothesis we test is analogous to the previous one:

$$H_0 : \beta = 0 \quad \text{and} \quad H_A : \beta < 0 .$$

This provides the alternative to the standard asset pricing framework, where no relationship is postulated between performance and the degree of product differentiation of the category. It also implicitly quantifies how distant the particular category is from pure competition.

The results, reported in Table 7, Panels A and B, strongly support the model. It seems that differentiation plays a role in explaining category returns at both the yearly and the monthly frequency. In particular, even after we use all the control variables, product differentiation remains highly significant and *positive*.

One of the main implications of these results is the fact that, if performance is related to “industry” variables, and if there is some arbitrage relationship holding in the markets, industry considerations should also affect the prices of the underlying asset in which the funds invest. Recently, Barberis and Shleifer (2001), studying asset prices in an economy where some investors classify risky assets into different styles, show that news about one style can affect the prices of other apparently unrelated styles. Assets belonging in the same style will comove more than assets in different styles. Their interpretation is “demand driven”, that is the comovement is induced by the fact that investors evaluate the assets on the basis of relative performance.

In our case, performance is “supply driven”. In other words, it is the industry itself that first groups the assets into different categories and then optimally chooses the average rate of return in each category on the basis of the type of competition existing in such a category. Categories where the degree of differentiation among funds is higher are also those that generate in equilibrium a lower performance. This would require lower returns for the assets (stocks, bonds) the funds in the particular category invest. To provide additional evidence of the differentiation performance relationship, we will now test whether the degree of product differentiation causes performance in the category.

6.4 Evidence of causality between performance and product differentiation

The tests of causality are performed by estimating the following system of equations:

$$\text{19)}\quad \begin{cases} r_{c,t} = \sum_{s=1}^{K} \alpha_{1,s} r_{c,t-s} + \sum_{s=1}^{K} \beta_{1,s} \Gamma_{c,t-s} + \eta_{c,s} + \varepsilon_{1,c,t} \\ \Gamma_{c,t} = \sum_{s=1}^{K} \alpha_{2,s} r_{c,t-s} + \sum_{s=1}^{K} \beta_{1,s} \Gamma_{c,t-s} + \eta_{c,s} + \varepsilon_{2,c,t} \end{cases}$$

where $r_{i,t}$ is the average performance of the funds belonging to the *cth* category at time *t*, $\Gamma_{c,t}$ is the degree of product differentiation of the *cth* category at time *t*, and η_c represents a category specific effect. *K* lags are used and different specifications based on alternative ways of constructing the index of differentiation are employed.

The panel structure and the existence of category specific characteristics (fixed effect), makes the estimation different from a standard VAR one. In order to address the potential endogeneity of the explanatory variables and the correlation between explanatory variables and residuals induced by the category specific effects (η_c), we use a dynamic panel GMM estimation with instrumental variables. In particular, the estimations are carried out by using Orthogonal Deviations and lags of the explanatory variables, as well as those of the dependent variable are employed as instruments (Arellano 1990, 1993 and Arellano and Bond 1988). The frequency of the observations is yearly. White's adjusted heteroskedastic consistent least-squares variance-covariance matrix is used to control for heteroskedasticity.

We allow for three lags. The choice is determined by the trade-off between the power of the estimation and the need to have enough free lags available to construct the set of instrumental variables. Also, we expect a period of three years to be a long enough in the mutual fund industry.

The results are reported in Table 8, Panels A and B. These show that, in the case where the index of dispersion is not performance based, there is strong evidence of causality from the degree of differentiation to performance, and no evidence of causality from performance to differentiation. This seems to suggest that market structure Granger causes fund performance. Evidence on the validity of the instruments is provided by the Sargan tests that always and consistently fail to reject the null of over-identifying restrictions. Further evidence on the validity of the instruments is provided by the tests for autocorrelation that reject the null of autocorrelation of the residuals.

However, bi-directional causality exists between dispersion and performance in the case where the index is related to performance. Also, these results are merely indicative, as the size of the sample is limited by the need to use yearly data. We also performed Granger causality tests by using monthly observations of 12 months' previous returns. While the size of the sample increases twelve times, the added persistence generated by the rolling twelve month observation windows is not enough to make up for the fact that fees are observed at a lower frequency. Also in this case, the results (not reported) agree with those reported, displaying Granger causality from market structure to returns.

Let us now consider the second testable restriction, the relationship between product differentiation and fund proliferation.

7 Fund proliferation and product differentiation

We consider the issue of fund proliferation in terms of market structure and degree of product differentiation. We will first consider test whether product differentiation affects the decision of the family to set up a new fund and/or to enter a new category. Then, we will see whether this aggregates at the category level, so that it affects the overall fund turnover of the category.

7.1 Is there a fund proliferation strategy at the family level?

If fund proliferation were an alternative way of attracting investors, we would expect fund proliferation to be higher in the very cases where performance is lower. That is, we would expect performance to be *negatively* related to the degree of product differentiation and fund proliferation to be *positively* related to it.

We consider two alternative measures of fund proliferation: the number of funds offered by a family within the category and the number of categories in which the family is active. Indeed, while product differentiation of a category may increase the number of funds the family offers within that category, the overall level of product differentiation of all the categories in which the family is active may induce the family to enter new categories. This would increase its overall level of market coverage.

Let us start by considering the relationship between fund proliferation within a category and product differentiation within it. We estimate:

$$20)\ N_{i,c,t} = \alpha + \beta\Gamma_{i,c,t} + \boldsymbol{\gamma}\mathbf{Controls_{i,t}} + \delta N_{i,c,t-1} + e_{i,t},$$

where $N_{i,c,t}$ is the number of funds of the *ith* family in the *cth* category at time *t*, and $\Gamma_{i,c,t}$ is the index that proxies for the degree of product differentiation of the *cth* category at time *t*. As before, we use various specifications based on different ways of constructing the index of differentiation. $\mathbf{Controls_{i,t}}$ is a vector of control variables that include the performance, expense ratio and total load fees of the family, as well as the performance, expense ratio and total load fees of the competing families. We perform a Dynamic Panel Data estimation, based on GMM with instrumental variables.[xxiv] We test the hypothesis:

$$H_0 : \beta = 0 \quad \text{and} \quad H_A : \beta > 0 .$$

A positive value of β would suggest that families tend to offer more funds in categories characterized by a higher degree of product differentiation.

The results, reported in Table 9, Panels A and B, show a statistically strong *positive* relationship between the degree of product differentiation and the number of funds. These results hold across all the different specifications and are very robust to the way the index of product differentiation has been constructed. Also, evidence on the validity of the instruments is provided by the absence of serial correlation and by the fact that the Sargan tests consistently fail to reject the null of over-identifying restrictions.

Let us now consider whether this relationship holds also for the overall number of categories in which the family is offering funds, i.e., the degree of market coverage. We therefore estimate

$$21)\ \ C_{i,t} = \alpha + \beta\Gamma_{i,t} + \boldsymbol{\gamma}\mathbf{Controls}_{\mathbf{i,t}} + \delta C_{i,t-1} + e_{i,t},$$

where $C_{i,t}$ is the degree of market coverage of the *ith* family at time *t*. It represents the number of categories in which the family is operating at time *t*. $\Gamma_{i,t}$ is the index that proxies for the average degree of product differentiation of all the categories the families has funds in at time *t*. This family-specific index of product differentiation is constructed by weighting the indexes of differentiation of all the categories the family has funds in by the total net assets of the funds its is running in such categories. The control variables are defined as before. We use various specifications based on different ways of constructing the index of differentiation for each.

The hypothesis to test is the same as before. However, now a positive value of β would suggest that families that pursue a strategy of category proliferation tend to operate in categories characterized by a higher degree of product differentiation. The results, reported in Table 10, Panels A and B, show a statistically strong *positive* relationship between the number of categories the family is operating in and their average degree of differentiation. This holds across all the different specifications, and regardless of the way we constructed the index of product differentiation. Also, the Sargan tests and the tests of autocorrelation provide evidence in favor of the specification.

7.2 Fund turnover and product differentiation

We now focus on the relationship between degree of differentiation and fund turnover. If families resort to fund proliferation in order to attract investors, we would expect the number of new and dead funds to be higher in the very categories where the degree of differentiation is stronger. That is, we expect that the strategies of fund proliferation would, in aggregate, induce more fund turnover. We therefore estimate

$$22)\ \ T_{c,t} = \alpha + \beta\Gamma_{c,t} + \boldsymbol{\gamma}\mathbf{Controls}_{\mathbf{c,t}} + \delta T_{c,t-1} + e_{c,t},$$

where $T_{c,t}$ is a measure of fund turnover, contructed as the number of new and dead funds in the *cth* category at time *t,* and where $\Gamma_{c,t}$ is the index that proxies for the degree of product differentiation in such a category. As before, different ways of constructing the index of differentiation are considered. The control variables are the average perfomance and risk of the category, the average overall performance of the overall mutual fund industry and the market risk premium (i.e., excess rate of return of the market on the riskless rate). Lagged number of funds and lagged turnover are also employed. We perform a Dynamic Panel Data estimation, based on GMM, with instrumental variables using the same methodology outlined before. The test is the same as before, i.e., we expect: $\beta > 0$.

The results, reported in Table 11, Panels A and B, show a strong *positive* correlation between fund turnover and the degree of product differentiation of the category. They suggest that families exploit the possibility offered by product differentiation by resorting to fund proliferation and the scaling down of the investment in performance enhancement. This generates a higher number of new funds, as well as a higher number of fund failures. Indeed, the mere fact that funds are perceived as differentiated products reduces the possibility of negative spillover that the disappearance of a fund may have on the reputation of the family.

The fact that more funds would attract investors and negative performance or failure does not necessarily feedback and negatively affects the family concur in increasing the incentive to resort to a fund proliferation strategy at the expense of performance. This is in line with earlier findings of Sirri and Tufano (1998) and our previous results (Table) that suggest that the demand for a fund is scarcely affected by the performance of the other funds of the family.

It is important to note that the econometric methodology eliminates the concerns due to simultaneity or spurious correlation. Indeed, the Sargan tests and the tests of autocorrelations support the choice of the instruments.

Conclusion

We showed that market structure affects mutual fund performance. We argued that performance is only one of the dimensions along which mutual fund families compete. In particular, we showed that performance enhancement and category proliferation are alternative ways of attracting investors. The choice between them is directly related to the degree of product dispersion in the market, measured as the dispersion in the “services” (fees, performance) that the competing funds offer. By testing our theory on the US mutual fund industry, we showed that the performance of the mutual funds is *negatively* related to the degree of product differentiation of the category in which they are active, while the degree of category proliferation is *positively* related to it.

Appendix

Performance and degree of product differentiation

We follow the approach pioneered by Anderson, De Palma and Thisse (1994). We solve the decision problem of the family as a two-stage game. At the first stage, we solve the sub-game in terms of investment costs (c_{im}). The managing company maximises:

$$23)\ Max_{c_{im}} \left\{ \sum_{m_i=1}^{M_i} \left[p_{im}(1 - C_i m_i) - c_{im} \right] P_{im} - k_i \right\} = Max_{c_{im}} \left\{ \sum_{m_i=1}^{M_i} mu_{im} P_{im} - k_i \right\}$$

with respect to c_{im}. mu_{im} is the mark-up that the *ith* family charges on the *mth* fund. This yields:

$$24) \frac{\partial \pi_i}{\partial c_{im}} = \left\{ \begin{array}{l} -P_{im} + mu_{im} P_{im} A_{im} [\frac{1}{\mu_1}(P_{m|i} - P_{im}) + \frac{1}{\mu_2}(1 - P_{m|i})] + \\ P_{im} \sum_{\substack{n=1 \\ n \neq m}}^{M} mu_{in} P_{in} A_{in} [\frac{1}{\mu_1}(P_{n|i} - P_{in}) - \frac{1}{\mu_2} P_{n|i}] \end{array} \right\} = 0,$$

where $A_i = \frac{\partial r_i}{\partial c_i}$. Equation 24) can also be rewritten as:

$$25)\ mu_{im} = \frac{\mu_2}{A_i} \left\{ P_{im} - A_i \sum_{n=1}^{M} mu_{in} P_{in} [\frac{1}{\mu_1}(P_{n|i} - P_{in}) - \frac{1}{\mu_2} P_{n|i}] \right\}.$$

This implies that the mark-up is the same across all the funds, that is, $mu_{im} = mu_{in} = mu_i$, for all *m = 1,…m,…, n,…M* . Also, $mu_{im} = \frac{\mu_1}{(1 - P_i)} A_i^{-1}$. The investment in performance-enhancement is: $c_i = p_i(1 - C_i M_i) + \frac{\mu_1}{(P_i - 1)} A_i^{-1}$.

To solve the next stage of the game, we substitute the optimal value of c_i into equation 23) and optimize in terms of the number of funds and of the fees charged. This yields: $p_i = \frac{\mu_2}{(A_i - 1)}$ and $M_i = \frac{A_i - 1}{(A_i C_i - B_i)}$. We concentrate on the case where $B_i > C_i$, that is the investors value the switching option more than it costs the family to generate it. If we impose the zero profit condition, we find that:

$$26)\ P_i = \frac{A_i k_i}{(A_i k_i + \mu_1)} \quad \text{and} \quad r_i = -\mu_1 + A_i \left[-k_i + \mu_2 \left(\frac{1}{(A_i - 1)} + \frac{C_i}{(B_i - A_i C_i)} \right) \right].$$

From the zero profit condition it also follows that:

$$27)\ \mu_1 A_i^{-1} \left[M_i e^{\frac{\phi_i}{\mu_2}} \right]^{\frac{\mu_2}{\mu_1}} \left[\sum_{\substack{j=1 \\ j \neq i}}^{I-1} [M_j e^{\frac{\phi_j}{\mu_2}}]^{\frac{\mu_2}{\mu_1}} \right]^{-1} = k_i,$$

where $\phi_{im} = r_{im} - p_i(1 - B_i M_i)$.

Let's assume that the entrance costs are identical for all the families, i.e. $k_i = k_j = k$, for all *i = 1…j…I.* Rearranging terms, solving for M_i and averaging over all the funds held by the different families, we have:

$$28)\ \Gamma = \mu_1^{-\frac{\mu_1}{\mu_2}} \frac{1}{\sum_{i=1}^{I} M_i} \left\{ \sum_{i=1}^{I} \left(A_i \sum_{\substack{j=1 \\ j\neq i}}^{I-1} [M_j \frac{e^{\frac{\phi_j}{\mu_2}}}{e^{\frac{\phi_i}{\mu_2}}}]^{\frac{\mu_2}{\mu_1}} \right)^{\frac{\mu_1}{\mu_2}} \right\}^{-\frac{\mu_2}{\mu_1}}.$$

By using the fact that M and ϕ are negatively related, it can be shown that, Γ is increasing in the degree of product differentiation (i.e., difference between ϕs, or $\phi_i - \phi_j$, for each $i \neq j$) if μ_2 is big enough, while, it decrease if μ_2 is small. After substituting for Γ, we have:

$$29)\ r_i = -\mu_1 + A_i\mu_2 \left[\frac{1}{(A_i - 1)} + \frac{C_i}{(B_i - A_i C_i)} \right] - A_i\Gamma.$$

Finally, if we take the derivative of M with respect to Γ, we find that:

$$30)\ \frac{\partial M}{\partial \Gamma} = \frac{\mu_1\mu_2(C-B)P(1-P)}{BC[I\mu_2(P-1)+\mu_1 P\Gamma]^2} > 0.$$

Testable restrictions of the family-driven heterogeneity

We follow the approach of Berry (1994) and Berry, Levinson and Pakes (1995) who derive a testable specification of the demand function of a product in terms of the market share of the product. From the specification of the demand function expressed in terms of the the probability of investment into a fund, we have:

$$31)\ P_{im} = P_{m|i}P_i \text{ where } P_{m|i} = \frac{e^{\frac{\delta_m}{\mu_2}}}{D_i},\ D_i = \sum_{m=1}^{M} e^{\frac{\delta_m}{\mu_2}},\ \text{and } P_i = \frac{D_i^{\frac{\mu_2}{\mu_1}}}{\sum_{i=1}^{I} D_i^{\frac{\mu_2}{\mu_1}}}.$$

That is, the probability of choosing the *mth* fund of the *ith* family can be decomposed into the *unconditional* probability that the investor selects the *ith* family (P_i) and the probability of choosing the specific *mth* fund *conditional* upon the choice of the *ith* family ($P_{m|i}$). The parameter δ_j is a vector containing all the characteristics of the funds that are observable to the econometrician. Taking logs, we have:

$$32)\ \ln(P_{im}) = \frac{1}{\mu_2}\delta_j - \frac{\mu_2 - \mu_1}{\mu_1}\ln(D_i) - \ln\sum_{i=1}^{I} D_i^{\frac{\mu_2}{\mu_1}} \text{ and } \ln(P_{m|i}) = \frac{1}{\mu_2}\delta_m - \ln(D_i).$$

We can also define the "outside alternative" as the only component of group zero, with $\delta_0 = 0$ and $D_0 = 1$. Investors can decide to invest into this alternative investment opportunity. It differs from the other "inside" products (i.e., the funds) as the price (i.e., financial characteristics) of this assets are not set in response to the ones of the inside goods. This implies that: $\ln(P_0) = -\ln\left(\sum_{i=1}^{I} D_i^{\frac{\mu_2}{\mu_1}}\right)$. Using equations 31) and 32) we have:

$$33)\ \ln(P_{im}) - \ln(P_0) = \frac{1}{\mu_1}\delta_m + \frac{\mu_1 - \mu_2}{\mu_1}\ln(P_{m|i})\,.$$

If we equate the observed market share with the probability of purchase of the specific fund, we have:

$$34)\ \ln(s_{im}) - \ln(s_0) = \frac{1}{\mu_1}\delta_m + \frac{\mu_1 - \mu_2}{\mu_1}\ln(s_{m|i}).$$

References

Admati R. A., Pfleiderer P. and J. Zechner (1994) “Large shareholder activism, risk sharing and financial market equilibrium” *Journal of Political Economy, 102, 1097-1130.*

Anderson S. P., De Palma A., Thisse J. (1994) “Discrete choice theory of product differentiation” *The MIT Press.*

Anderson, T.W. and C. Hsiao (1981), ”Formulation and estimation of dynamic models using panel data”, *Journal of Econometrics*, 18, 47-82.

Arellano, M. (1990), “Testing for autocorrelation in dynamic random effects models”, *Review of Economic Studies,* 57, 127-134.

Arellano, M. and S.R. Bond (1991), “Some tests of specification for panel data” Monte Carlo evidence and an application to employment equations”, *Review of Economic Studies*, 58, 277-297.

Arellano, M. and O. Bover (1995), “Another look at instrumental variable estimation of error-components models”, *Journal of Econometrics,* 68, 29-51.

Barberis, N. and Shleifer, A., (2001) “Style investing”, Working Paper.

Barclay, M. J., Pearson, N.D. and Weisbach M.S., (1998), “Open-end mutual funds and capital gains taxes”, *Journal of Financial Economics*, 49, 3-43.

Baumol, W. J., Goldfeld S. M., Gordon L. A., Koehn M. F. (1990), “The economics of mutual fund markets: competition versus regulation”, *Kluwer Academic Publishers.*

Berry, S. (1994) “Estimating discrete choice models of production differentiation”, *Rand Journal of Economics,* 25, 242-261.

Berry. S, Levinsohn, J. and Pakes, A, (1996), “Automobile prices in market equilibrium”, Econometrica.

Bresnahan T. F., “Empirical studies of industries with market power”, *Handbook of Industrial Organization*, Vol. 2, Chapter 17.

Brown, S.J. and Goetzmann, W.N. (1995), “Performance persistence”, *Journal of Finance*, 50, 853-873.

Brown, S.J. and Goetzmann, W.N. (1997), “Mutual fund styles”, *Journal of Financial Economics,* 43, 373-399.

Carhart, M., (1997), “On persistence in mutual fund performance”, *Journal of Finance*, 52, 57-82.

Chevalier, J. and Ellison G. (1997), “Risk taking by mutual funds as a response to incentives”, *Journal of Political Economy*, 105, 1167-1200.

Dermine J., Neven D. and Thisse J. F. (1991), “Towards an equilibrium model of the mutual funds industry”, *Journal of Banking and Finance,* 15, 485-499.

Elton E. J., Gruber M. J., Blake, C. R. (1996), “The persistence of risk-adjusted mutual fund performance”, *Journal of Business,* 69, 133-157.

Elton E. J., Gruber M. J., Das S. and Hlavka, M. (1993), “Efficiency with costly Information: a reinterpretation of evidence from managed portfolios”, *The Review of Financial Studies*, 6, 1-22.

Ferson, W. E., Schadt, R.W. (1996), “Measuring fund strategy and performance in changing economic conditions”, *Journal of Finance*, 51, 425-461.

Goetzmann, W.N. and Ibbotson, R.G., (1994), “Do winners repeat? Patters in mutual fund performance”, *Journal of Portfolio Management*, 20, 9-17.

Grinblatt, M. and Moskowitz T., (1999), “Do industries explain momentum?”, *Journal of Finance*, 54, 1249-1290.

Grinblatt, M. and Titman S., (1994), “The persistence of mutual fund performance”, *Journal of Finance*, 47, 1877-1984.

Gruber. M, (1996), “Another puzzle: the growth in actively managed mutual funds”, *Journal of Finance*, 51, 1127-1161.

Ippolito R. A. (1989), “Efficiency with costly information: a study of mutual fund performance, 1965-1984”, *The Quarterly Journal of Economics*, 104, 1-23.

Ivkovich Z. (2001), “Is blood thicker than water: spillovers in mutual fund families”, Working Paper.

Khorana A., Servaes H. (1999), “The determinants of mutual fund starts”, *The Review of Financial Studies*, 12 (5).

Khorana A., Servaes H. (2000), “What drives market share in the mutual fund industry”, Working Paper.

Kiviet, J.F. (1995), “On bias, inconsistency and efficiency of various estimators in dynamic panel data models”, *Journal of Econometrics*, 68, 53-78.

Mamaysky, H. and Spiegel M., (2001), “ A theory of mutual funds: optimal fund objectives and industry organization”, Working Paper.

Massa, M., (1998) “ Are there too many mutual funds? Mutual fund families, market segmentation and financial performance”, Working Paper.

Nanda, V., Narayanan, M.P. and Warther, V.A. (2000), “Liquidity, investment ability and mutual fund structure”, *Journal of Financial Economics*, 57, 417-443.

Nanda, V., Wang, Z. and Zheng L. (2001), “Family values and the star phenomenon”, Working Paper.

Roll R., Ross S. (1980), “An empirical investigation of the Arbitrage Pricing Theory”, *The Journal of Finance*.

Sirri R. E., Tufano P. (1998), “Costly search and mutual fund flows”, *Journal of Finance*, 53(5), 1589-1622.

Starks, L. and Brown.K., (1996), “Of tournaments and temptations: an analysis of managerial incentives in the mutual fund industry”, *Journal of Finance*, 51, 85-110.

Wermers, R., (2000), “Mutual fund performance: an empirical decomposition into stock-picking talent, style, transaction costs and expenses”, *Journal of Finance*, forthcoming.

Zheng, L., (1999), “Is money smart? A study of mutual fund investors’ selection ability”, *Journal of Finance*, 54, 901-933.

---

[i] The existence of a mutual fund industry can be justified if it helps to reduce transaction costs, thanks to the possibility of conducting large-size transactions at low costs. Even so, a limited number of mutual funds would be more than enough to complete the task of mimicking the risky portfolios at low costs for most investors.

[ii] The existence of strategies such as the creation of many funds in order to hide the poor performers merging them into the best ones, or on the use of "incubator fund strategies" support this view.

[iii] We may also think of other types of heterogeneities: investors may have different expectations on the returns of the different categories – i.e., tips or inside information about which category will generate higher returns. Alternatively, investors may have different hedging needs that lead them to evaluate the performance/fee combination in terms of their overall portfolio and wealth/income profile. However, these are not strictly necessary in order to generate segmentation.

[iv] These characteristics define the *category* (i.e., market category) the fund belongs to. For simplicity, we assume for simplicity that each family may enter each category with only one fund. That is, the choice of each family consists of the decision of setting up a fund with specific characteristics. In the empirical section, we will relax this assumption by working with the average characteristics of all the funds belonging to a particular family and active in a specific category.

[v] See Anderson, De Palma and Thisse (1994).

[vi] "μ_1 can be viewed as a measure of inter-firm heterogeneity, whereas μ_2 represents intra-firm heterogeneity". (Anderson, De Palma and Thisse, 1994, p. 251)

[vii] Both measures are bounded between zero (when all the funds are perceived to be same) and one (when all the funds are perceived as completely different products).

[viii] We adopt a profit-maximization approach that, in line with recent literature, Chevalier and Ellison (1997), defines the main goal of a family in terms of the maximization of the managed assets rather than the maximization of performance. In fact, given that most of the fees paid by the unit holder are proportional to funds' managed assets, profits are maximised when net fund-raising is maximised.

[ix] For example, Chacko and Das (1999) model the decision of the mutual fund manager to purchase information usable in order to enhance performance. In their model, funds optimally choose to buy information in order to improve their market timing or selectivity skills. Alternatively, c_{im} may be interpreted as the fees and the performance bonuses paid to the fund managers. These are variable costs linked to the desired target of performance (under the assumption that higher remuneration can "buy" better fund managers). Alternatively, given that investors only care about performance *net* of fees and a way of boosting net performance is simply reducing the management fees, we can think of c_{im} as the cost of such a fee-waiver policy.

[x] We are implicitly assuming that the fixed costs incurred to set up an additional fund are zero. Indeed, in general the start-up costs for each new fund are very small and the mutual fund industry is a typical example of an industry characterized by very low barriers to entry.

[xi] In particular,

$$\Gamma = \mu_1^{-\frac{\mu_1}{\mu_2}} \frac{1}{\sum_{i=1}^{I} M_i} \left\{ \sum_{i=1}^{I} \left(A_i \sum_{\substack{j=1 \\ j\neq i}}^{I-1} [M_j \frac{e^{\frac{\phi_j}{\mu_2}}}{e^{\frac{\phi_i}{\mu_2}}}]^{\frac{\mu_2}{\mu_1}} \right)^{\frac{\mu_1}{\mu_2}} \right\}^{-\frac{\mu_2}{\mu_1}}$$

[xii] Indeed, during the 1961-1982 period 156 funds did not have an OBJ code, that is 16.81% of the total number of funds. On the contrary, for the period 1992-2000 760 funds did not have a ICDI_OBJ code assigned, that is 4.296% of the total number of funds.

[xiii] We thank an anonymous referee for pointing this out.

[xiv] We also estimated a specification where the deviation of the flows at the fund level was related to the degree of family coverage (i.e., the number of categories in which the family is active). The results agree with those reported and are available upon request.

[xv] They include: the total net assets managed by the family as well as the flows in and out of the funds of the family, both in the category and overall across categories. We also include the average performance, risk and expense ratio of the family and of the competing families, both in the category and overall across categories and the average total load fees of the competing families. Averages are taken across all the funds in the category ("in the category") and across all the funds the family has overall ("overall").

[xvi] This choice is not above criticism as total net assets are a function of both investors' demand and the management policy of the fund. Higher performance increases the assets under management. However, the increase of the assets under management due to improvement of the performance is of second order if compared to the one due to the flows. Furthermore, we can control for it by directly netting the assets of the accumulated performance. This would imply

---

expanding equation by adding additional terms that account for the increment in performance of the fund and of the family. The omission of such a variable would only change the value of the coefficients of investors' reaction to past performance as this coefficient would proxy for both factors (i.e., increment in the value of the assets due to higher performance as well as higher flows). However, it would not affect the estimate of *b*, which is the one we care about.

[xvii] This is derived from the "Flow of Funds Accounts of the United States, Annual Flows and Outstandings", issued by the Board of Governors of the Fed. This implicitly assumes that the investors are mostly US investors. Alternatively, we use as proxy the overall market capitalization. Given that the results do not differ, we only report the former.

[xviii] Alternative specifications inclusive or not inclusive of the total net assets are considered as a robustness check.

[xix] As before, the model is estimated in Orthogonal Deviations. Lagged levels and differences of the explanatory variables are used as instruments. The estimation is done using White's adjusted heteroskedastic consistent least-squares.

[xx] The control variables include the past performance of the family, its average risk, the average performance and of the competing families, the average total load fees and expense ratios of both the family and the competing families. Averages are taken across all the funds of the same family over all categories ("overall").

[xxi] As before, the model is estimated in Orthogonal Deviations. Lagged levels and differences of the explanatory variables are used as instruments. The estimation is done using White's adjusted heteroskedastic consistent least-squares.

[xxii] For homogeneity and to make the results across specification comparable, we will construct also the index based on performance-related characteristics by using yearly performance and risk.

[xxiii] At monthly frequency, to compensate for the lower explanatory power of the control variables we include a more detailed set of control variables. In particular, in Panel A, the control variables include the average returns of all the funds in the market (across all categories) in the previous twelve months. The average standard deviation of all the funds in the same category as well as in the market in the previous twelve months ("Risk"), the average fees charged by all the funds in the same category as well as in the market in the previous twelve months. In Panel B, the controls variables include the average returns of all the funds in the market in the previous twelve months. The average standard deviation of the returns on all the funds in the same category ("Risk"), the average skewness of the returns of all the funds in the same category, the average fees charged by all the funds in the same category. We consider various types of fees: the total expense ratio, the mb-12 fees, the total load fees, the rear-end load fees and other load fees.

[xxiv] As before, the model is estimated in Orthogonal Deviations. Lagged levels and differences of the explanatory variables are used as instruments. The estimation is done using White's adjusted heteroskedastic consistent least-squares.

Table 1 Descriptive statistics

We report the main descriptive statistics of the 24 categories. In Panel A, we report the number of funds and the number of families active within each category (average value in and standard deviation in the period). In Panel B we report the average performance of the funds belonging to the category and the standard deviation of their performances over time ("Risk"). We report the average value and the standard deviation in the period. In Panel C we report the value of the net total assets ("Size") of the funds belonging to the category and the average value of their expense ratios. We report the average value and the standard deviation in the period. In these three Panels, we aggregate across funds by using weighted averages calculated by using as weights the total net asset values for each fund. In Panel D, we report the average value of the three types of indexes of product heterogeneity. We consider three ways of constructing the index. The first way uses only non performance-related characteristics, that is: the dispersion of fees across all the funds in the category. The second way uses the performance-related characteristics, that is: the dispersion of fund performances and standard deviation of the performances in the previous 12 months across all the funds in the category. The third way uses both performance-related characteristics and non performance-related ones, as defined before, along with the dispersion of the distributed incomes, capital gain distributions and average yields. The indexes are constructed as follows. First we calculate the standard deviation of the characteristics (e.g., fees charged) across all the funds active in the category. This is done separately for each type of characteristics every year. Then, the resulting components of the index are standardized so as to make them homogeneous in terms of range of variation. The standardization is based on dividing the single components by their maximum value over the whole period. The fees we consider are: maximum front-end load fees, maximum contingent deferred sales charges, rear load or redemption fees, 12b-1 fee, maximum deferred sales charges and redemption fees. We also include the expense ratio.

Panel A

Categories	Fund Number (Average)	Fund Number (Variation)	Family Number (Average)	Family Number (Variation)
Aggressive Growth	754	379	215	43
Balanced	343	138	122	18
High Quality Bonds	664	205	181	22
High Yield Bonds	196	98	70	16
Global Bonds	292	95	77	10
Global Equity	305	136	79	14
Growth and Income	642	253	188	18
Ginnie Mae	192	30	61	6
Government Securities	447	88	149	10
International Equities	772	424	160	41
Income	161	57	63	9
Long-Term Growth	1066	488	275	37
Tax-Free Money Market	435	46	109	3
Govt Securities Money Market	412	58	115	4
High Quality Municipal Bond	449	111	127	7
Single-State Municipal Bond	1149	332	115	6
Taxable Money Market	494	71	171	4
High-Yield Money Market	45	13	22	1
Precious Metals	40	145	25	2
Sector	248	145	63	25
Special	14	5	3	2
Total Performance	243	84	84	9
Utility	79	23	30	3

Panel B

Categories	Performance (Average)	Performance (Variation)	Risk (Average)	Risk (Variation)
Aggressive Growth	0.015	0.052	0.052	0.018
Balanced	0.009	0.022	0.024	0.007
High Quality Bonds	0.005	0.009	0.009	0.002
High Yield Bonds	0.007	0.016	0.014	0.006
Global Bonds	0.004	0.012	0.015	0.004
Global Equity	0.012	0.036	0.036	0.014
Growth and Income	0.012	0.033	0.033	0.012
Ginnie Mae	0.004	0.008	0.008	0.002
Government Securities	0.004	0.010	0.010	0.001
International Equities	0.009	0.041	0.043	0.013
Income	0.011	0.028	0.028	0.010
Long-Term Growth	0.014	0.039	0.041	0.014
Tax-Free Money Market	0.002	0.0004	0.0003	0.0001
Govt Securities Money Market	0.003	0.0008	0.0003	0.0002
High Quality Municipal Bond	0.004	0.012	0.011	0.003
Single-State Municipal Bond	0.004	0.012	0.012	0.004
Taxable Money Market	0.003	0.0008	0.0003	0.0001
High-Yield Money Market	0.004	0.011	0.011	0.003
Precious Metals	0.0002	0.091	0.088	0.031
Sector	0.018	0.050	0.057	0.022
Special	-0.001	0.023	0.022	0.012
Total Performance	0.009	0.021	0.023	0.007
Utility	0.010	0.027	0.028	0.008

Panel C

Categories	***Total Net Assets (Average)***	***Total Net Assets (Variation)***	***Expense Ratio (Average)***	***Expense Ratio (Variation)***
Aggressive Growth	225.78	81.43	0.012	0.0003
Balanced	286.34	41.77	0.008	0.0004
High Quality Bonds	195.21	33.30	0.007	0.0003
High Yield Bonds	341.22	43.22	0.011	0.0002
Global Bonds	156.70	51.82	0.013	0.0009
Global Equity	311.10	95.34	0.013	0.0007
Growth and Income	594.48	183.91	0.007	0.0004
Ginnie Mae	437.62	138.69	0.008	0.0008
Government Securities	136.31	31.18	0.009	0.0006
International Equities	172.75	38.57	0.013	0.0007
Income	505.07	105.01	0.008	0.0002
Long-Term Growth	399.20	130.10	0.010	0.0003
Tax-Free Money Market	313.79	58.33	0.005	0.0001
Govt Securities Money Market	514.15	98.04	0.005	0.0003
High Quality Municipal Bond	251.68	48.29	0.007	0.0002
Single-State Municipal Bond	102.68	23.36	0.007	0.0004
Taxable Money Market	1052.11	332.35	0.005	0.0001
High-Yield Money Market	514.42	74.63	0.008	0.0002
Precious Metals	89.75	35.87	0.014	0.0012
Sector	180.62	79.15	0.013	0.0005
Special	42.85	46.13	0.012	0.0019
Total Performance	209.80	38.90	0.009	0.0003
Utility	271.05	68.49	0.0119	0.0008

Panel D

	Non performance-related Index		**Performance-related Index**		**Mixed Index**	
	(Average)	**(Variation)**	**(Average)**	**(Variation)**	**(Average)**	**(Variation)**
Categories						
Aggressive Growth	0.7857	0.0362	0.3875	0.0890	0.4158	0.0328
Balanced	0.7940	0.0531	0.2205	0.0321	0.2547	0.2427
High Quality Bonds	0.6354	0.0535	0.2786	0.2510	0.3026	0.0535
High Yield Bonds	0.8732	0.0057	0.2239	0.0553	0.2622	0.1281
Global Bonds	0.8125	0.0200	0.3578	0.1321	0.3884	0.0720
Global Equity	0.8828	0.0108	0.3711	0.0741	0.4050	0.0414
Growth and Income	0.7862	0.0643	0.2318	0.0404	0.2653	0.0438
Ginnie Mae	0.7507	0.0188	0.2198	0.0456	0.2518	0.0277
Government Securities	0.7128	0.0343	0.2071	0.0282	0.2375	0.0797
International Equities	0.8041	0.0537	0.3994	0.0810	0.4282	0.0776
Income	0.8036	0.0487	0.2745	0.0799	0.3075	0.0649
Long-Term Growth	0.7882	0.0598	0.2970	0.0643	0.3284	0.0359
Tax-Free Money Market	0.1557	0.0457	0.1996	0.0365	0.2012	0.0537
Govt Securities Money Market	0.2136	0.0853	0.1912	0.0564	0.1960	0.0374
High Quality Municipal Bond	0.7628	0.0323	0.1494	0.0393	0.1844	0.0367
Single-State Municipal Bond	0.7848	0.0513	0.1318	0.0386	0.1685	0.0501
Taxable Money Market	0.4428	0.0843	0.2198	0.0516	0.2357	0.0304
High-Yield Money Market	0.8918	0.0434	0.1467	0.0318	0.1885	0.0956
Precious Metals	0.8106	0.0403	0.3210	0.0987	0.3527	0.1460
Sector	0.8034	0.0138	0.5169	0.1512	0.5419	0.1402
Special	0.1282	0.0161	0.4361	0.1453	0.4284	0.0603
Total Performance	0.8252	0.0299	0.2819	0.0614	0.3157	0.0439
Utility	0.8724	0.0143	0.2180	0.0456	0.2564	0.0328

Table 2, Panel A: Stability of investors' flows: fund-level analysis

We estimate: $VF_{m,i,t} = \alpha + \beta \mathbf{F_{i,t}} + \gamma N_{i,t} + \boldsymbol{\delta}\mathbf{Controls_{m,i,t}} + e_{m,i,t}$, where the subscripts, "*m*", "*i*", and "*t*" refer, respectively, to the *mth* fund, belonging to the *ith* family at time *t*. The frequency is yearly. $VF_{m,i,t}$ is the standard deviation of the flows in the fund over the twelve months in the respective year. $\mathrm{F_{i,t}}$ represents the Total Load Fees charged by the *ith* family at time *t*. $N_{\mathbf{i,t}}$ is the overall number of funds of the family. We consider alternatively, the Total Load Fees charged by the fund (specification I and IV), the average Total Load Fees charged by the family for all the funds belonging to the same category (specifications II and V) and the average Total Load Fees charged by the family for all the funds it has across all categories (specifications III and VI). $\mathbf{Controls_{m,i,t}}$ is a vector of control variables. They include: the Total Net Assets managed by the fund, the Total Net Assets managed by the family the fund belongs to, the Flows in and out of the fund, the Performance of the fund, the Risk of the fund, the Performance of the family the fund belongs to. Risk for each fund is constructed for each fund as the standard deviation of the performances in the previous 12 months. We also include the average Performance and Risk of the funds belonging to competing families. We use White's adjusted heteroscedastic consistent least-squares regression.

	I		*II*		*III*		*IV*		*V*		*VI*	
Explanatory Variables	*Value*	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*
Constant	24.66	13.50	24.46	13.79	25.19	11.64	20.63	16.77	20.76	17.21	20.74	12.87
Fund Total Load Fees	-3.20	-15.52	-	-	-	-	3.26	-15.27	-	-	-	-
Total Load Fees of the family in the categ.	-	-	-4.37	-18.77	-	-	-	-	-4.51	-18.45	-	-
Total Load Fees of the family overall	-	-	-	-	-2.83	-7.18	-	-	-	-	-2.75	-7.06
Number of Funds	0.02	4.63	0.03	6.23	0.02	4.83	0.02	4.57	0.03	6.24	0.02	4.74
Control Variables												
Fund Total Net Assets	0.01	6.06	0.01	6.03	0.01	6.06	0.01	6.04	0.01	6.01	0.01	6.04
Family Total Net Assets	-0.003	-1.38	-0.003	-1.37	-0.003	-1.31	-0.004	-1.45	-0.004	-1.43	-0.004	-1.52
Fund Average Flows	0.27	0.77	0.27	0.78	0.27	0.78	0.27	0.78	0.28	0.79	0.28	0.79
Fund Performance	-122.58	-1.06	-104.16	-0.90	-120.08	-1.04	-55.26	-2.13	-142.90	-1.96	-167.09	-2.31
Fund Risk	111.22	3.38	102.15	3.16	95.37	2.94	-205.38	-8.83	-199.82	-8.49	-252.27	-10.53
Family Performance	0.01	0.04	-0.009	0.02	0.01	0.04	0.02	0.06	0.01	0.04	0.02	0.07
Average Performance of comp.families in the category	55.07	0.57	33.19	0.34	28.27	0.29	-	-	-	-	-	-
Average Ovearall Performance of comp. Families	-348.20	-2.87	-315.21	-2.61	-338.75	-2.79	-	-	-	-	-	-
Average Risk of comp.families in the categ.	-404.42	-8.68	-385.17	-8.30	-439.52	-9.01	-	-	-	-	-	-
Adjusted R Square	0.14		0.14		0.13		0.13		0.14		0.13	

Table 2, Panel B: Stability of investors' flows: analysis at the family level within the category

We estimate: $VF_{i,t|c} = \alpha + \beta F_{\mathbf{i,t}} + \gamma C_{i,t} + \boldsymbol{\delta}\mathbf{Controls}_{\mathbf{i,t}} + e_{i,t|c}$, where the subscripts, "*c*", "*i*", and "c" refer, respectively, to the funds of the *ith* family in the *cth* category at time *t*. The frequency is yearly. $VF_{i,t|c}$ is the standard deviation of the flows of the funds managed by the *ith* family in the *cth* category over the twelve months in the respective year. $\mathrm{F}_{\mathrm{i,t}}$ represents is the average value of the Total Load Fees of the family in the category . $C_{\mathbf{i,t}}$ is the degree of market coverage, that is the number of categories the family is operating in. $\mathbf{Controls}_{\mathbf{i,t}}$ is a vector of control variables. They include: the Total Net Assets and the Flows in and out of the funds managed by the family within the *cth* category, the average Performance and Expense Ratio of the family and of the competing families, both in the category and overall across categories and the average Total Load Fees of the competing families. We also include the Risk of the family and of the competing families in the category. Risk for each fund is constructed for each fund as the standard deviation of the performances in the previous 12 months. We use White's adjusted heteroscedastic consistent least-squares regression.

	I		*II*		*III*		*IV*		*V*		*VI*	
Explanatory Variables	*Value*	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*
Constant	-69.19	-2.57	21.59	5.94	19.49	5.88	18.05	1.12	22.79	8.91	21.77	8.74
Total Load Fees of the family in category	-2.03	-9.18	-2.54	-10.49	-2.54	-10.47	-2.57	-11.23	-2.56	-11.64	-2.60	-11.59
Family Coverage	0.29	2.05	0.31	2.24	0.32	2.45	0.29	2.21	0.29	2.25	0.31	2.42
Control Variables												
Family Total Net Assets in the category	0.001	3.13	0.001	3.20	0.001	3.20	0.001	3.26	0.001	3.26	0.001	3.27
Family Total Net Assets overall	-0.82	-0.50	-0.55	-0.36	-0.73	-0.47	-1.09	-0.69	-1.09	-0.69	-1.28	-0.81
Family Average Flows in the category	0.87	1.60	0.86	1.59	0.86	1.59	0.83	1.57	0.83	1.58	0.83	1.58
Family Average Flows overall	-0.11	-0.57	-0.12	-0.62	-0.12	-0.59	-	-	-	-	-	-
Family Performance in the category.	-299.84	-1.56	-322.09	-1.67	-183.10	-1.89	-352.38	-2.14	-352.57	-2.15	-220.35	2.75
Family Performance overall	33.22	0.25	-24.11	-0.18	-58.57	-0.61	-	-	-	-	-	-
Family Risk in the category	52.50	2.11	77.10	3.01	-118.91	-3.79	87.64	3.39	87.80	3.39	-140.02	-5.69
Family Expense Ratio in the category	-163.52	-0.84	-781.63	-3.90	-876.98	-4.58	-383.03	-3.85	-384.59	-3.83	-373.59	-3.51
Family Overall Expense Ratio	-58.71	-0.24	513.60	2.35	650.03	3.21	-	-	-	-	-	-
Average Performance of comp. families in categ.	318.08	1.75	227.09	1.26	-	-	199.43	1.23	199.39	1.23	-	-
Average Overall Performance of comp. families	-451.96	-2.12	-135.30	-0.74	-	-	-	-	-	-	-	-
Average Risk of comp. families in the category	-95.78	-1.80	-248.49	5.70	-	-	-275.16	-8.09	-276.12	-8.13	-	-
Average Expense Ratio of comp.families in categ.	-1,861.56	-10.16	-	-	-	-	-	-	-	-	-	-
Average Overall Expense Ratio of comp.families	4,335.36	3.80	-	-	-	-	-	-	-	-	-	-
Average Overall Total Load Fees of comp families	28.16	3.11	-	-	-	-	2.26	0.29	-	-	-	-
Adjusted R Square	0.12		0.11		0.11		0.11		0.11		0.11	

Table 2, Panel C: Stability of investors' flows: family level analysis

We estimate: $VF_{i,t} = \alpha + \beta F_{\mathbf{i,t}} + \gamma C_{i,t} + \boldsymbol{\delta}\mathbf{Controls}_{\mathbf{i,t}} + e_{i,t}$, where the subscripts "*i*", and "*t*" refer, respectively, to the *ith* family at time *t*. The frequency is yearly. $\mathrm{VF}_{\mathrm{it}}$ is the standard deviation of the flows in the fund over the twelve months in the respective year. $F_{i,t}$ represents the average of the Total Load Fees the family charges across all categories. $C_{\mathbf{i,t}}$ is the degree of market coverage, that is the number of categories the family is operating in. $\mathbf{Controls}_{\mathbf{m,i,t}}$ is a vector of control variables. They include: the Total Net Assets managed by the family, the average Flows of the funds run by the same family, the average Performance, Risk and Expense Ratio of all the funds run by the same family, as well as average Performance, Risk and Expense Ratio of the competing families. Risk for each fund is constructed for each fund as the standard deviation of the performances in the previous 12 months. We use White's adjusted heteroscedastic consistent least-squares regression.

	I		*II*		*III*		*IV*		*V*		*VI*	
Explanatory Variables	*Value*	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*
Constant	-53.08	-4.70	9.35	10.18	15.32	12.28	-50.54	-4.50	10.58	14.06	17.49	18.53
Family Overall Total Load Fees	-1.47	-12.62	-1.79	-14.98	-2.03	-15.22	-1.32	-9.33	-1.65	-11.29	-1.83	-11.63
Family Coverage	0.15	3.13	0.16	3.42	0.23	5.56	-	-	-	-	-	-
Control Variables												
Family Total Net Assets	5.76	8.04	5.98	8.16	6.44	8.46	6.32	9.74	6.59	9.91	7.35	10.65
Family Flows	0.86	4.42	0.88	4.55	0.88	4.44	0.86	4.42	0.88	4.56	0.86	4.45
Family Performance	-132.36	-2.45	-94.76	-1.70	-216.18	-6.64	-136.44	-2.52	-96.73	-1.74	-219.66	-6.74
Family Risk	-172.62	-4.52	-243.83	-6.35	-84.98	-4.27	-179.26	-4.83	-254.05	-6.95	-92.78	-4.85
Family Expense Ratio	-297.97	-3.80	-	-	-	-	-310.96	-3.93	-	-	-	-
Average Performance of comp. families	-54.33	-0.69	-71.35	-0.97	-	-	-46.87	-0.60	-65.06	-0.89	-	-
Average Risk of comp. families	334.92	5.31	401.11	8.24	-	-	343.21	5.54	414.32	8.98	-	-
Average Expense Ratio of competing families	1056.69	2.04	-	-	-	-	1063.92	2.05	-	-	-	-
Average Total Load Fees of competing families.	25.68	7.12	-	-	-	-	25.06	6.95	-	-	-	-
Adjusted R Square	0.17		0.16		0.15		0.17		0.16		0.15	

Table 3: Family-driven heterogeneity in the mutual fund industry

We estimate $\ln(s_{mi,t})-\ln(s_{0,t})=\mathbf{a}\boldsymbol{\delta}_{\mathbf{mi,t}}+b\ln(s_{m|i,t})+c[\ln(s_{mi,t-1})-\ln(s_{0,t-1})]+\varepsilon_{mi,t}$, where $s_{ic,t}$ is the ratio between the assets of all the funds the *ith* company has in the *cth* category and the overall assets in the market at time *t,* $s_{c|i,t}$ is the ratio between the assets of the *ith* family in the *cth* category and the overall assets of the *ith* family at time t, while s_0 is the market share of the other alternative (to investing into mutual funds) possibilities of investment the investors have and the overall assets in the market at time *t*. We use as alternative investment opportunities of the investors the part of US financial wealth not invested in mutual funds. The vector $\boldsymbol{\delta}_{\mathbf{ic,t}}$ contains all the characteristics of the funds. They are the performance of the fund, its risk, the performance of the family it belongs to, the overall risk of such a family, the total load fees charged by the family and the total net assets of all the funds belonging to the family in the different categories. We estimate the model by using GMM with instrumental variables using two alternative specifications: one is based on First Differences (Panel A) and one is based on Orthogonal Deviations (Panel B). White's robust variance covariance matrix is employed. The Sargan test of over-identifying is reported. * means that its *p-value* > 0.1

Panel A: First Differences

	I		*II*		*III*		*IV*		*V*		*VI*	
	Value	***t-stat***	*Value*	***t-stat***	*Value*	***t-stat***	*Value*	***t-stat***	*Value*	***t-stat***	*Value*	***t-stat***
$\ln(s_{m\|i,t})$	0.27	8.61	0.29	7.10	0.18	1.58	0.42	2.68	0.29	3.08	0.17	2.31
Family Performance in categ	1.49	3.51	1.52	3.61	2.05	2.10	0.41	0.32	1.43	4.24	1.73	6.66
Overall Family Performance	-0.82	-0.92	0.57	1.01	-0.32	-0.43	1.32	1.29	-	-	-	-
Family Total Load Fees in the category	-19.45	-1.99	-	-	-26.31	-4.00	-	-	-27.31	-3.41	-	-
Overall Family Total LoadFees	-	-	-38.76	-3.48	-	-	-44.22	-3.76	-	-	-33.20	-4.12
Overall Family Risk	-0.07	-0.23	0.00	0.02	-0.10	-0.37	0.16	0.46	-0.12	-0.32	0.23	0.72
Number of funds of the family in other categories	0.00	10.96	0.00	11.09	0.00	4.70	0.00	3.04	0.00	2.29	0.00	1.39
Number of funds of the family in the same category	-	-	-	-	-	-	-	-	-0.00	-0.28	0.02	1.44
$\ln(s_{mi,t-1})-\ln(s_{0,t-1})$	0.43	5.37	0.48	5.85	0.57	3.83	0.32	1.57	0.52	2.89	0.31	1.99
Family Total Net Assets	-	-	-	-	-0.00	-0.78	0.00	0.91	-	-	-	-
Sargan test	3.88*		1.03*		0.42*		0.001*		0.82*		0.22*	

Panel B: Orthogonal Deviations

	I		*II*		*III*		*IV*		*V*		*VI*	
	Value	***t-stat***	*Value*	***t-stat***	*Value*	***t-stat***	*Value*	***t-stat***	*Value*	***t-stat***	*Value*	***t-stat***
$\ln(s_{m\|i,t})$	0.27	8.65	0.29	7.13	0.18	1.60	0.42	2.69	0.29	3.30	0.17	2.35
Family Performance in categ	1.50	3.51	1.52	3.63	2.04	2.12	0.41	0.32	1.43	4.54	1.73	6.76
Overall Family Performance	-0.82	-0.93	0.57	1.02	-0.32	-0.43	1.32	1.29	-	-	-	-
Family Total Load Fees in the category	-19.38	-1.98	-	-	-26.32	-4.06	-	-	-27.32	-3.67	-	-
Overall Family Total LoadFees	-	-	-39.77	-3.50	-	-	-44.22	-3.76	-	-	-33.17	-4.19
Overall Family Risk	-0.07	-0.24	0.00	0.02	-0.10	-0.37	0.16	0.46	-0.13	-0.36	0.23	0.74
Number of funds of the family in other categories	0.00	10.93	0.00	11.13	0.00	4.74	0.00	3.04	-0.01	-0.30	0.02	1.46
Number of funds of the family in the same category	-	-	-	-	-	-	-	-	0.00	2.46	0.00	1.41
$\ln(s_{mi,t-1})-\ln(s_{0,t-1})$	0.43	5.37	0.48	5.88	0.57	3.88	0.32	1.57	0.52	3.12	0.31	2.03
Family Total Net Assets	-	-	-	-	-0.00	-0.78	0.00	0.91	-	-	-	-
Sargan test	3.87*		1.04*		0.43*		0.001*		0.94*		0.22*	

Table 4: Fund performance and market dispersion

We estimate: $r_{mic,t} = \alpha + \beta\Gamma_{c,t} + \boldsymbol{\gamma}\mathbf{Controls_{mic,t}} + \delta r_{mic,t-1} + e_{mic,t}$, where the subscripts "*m*", "*i*", "*c*" and "*t*" refer, respectively, to the *mth* fund, of the *ith* family, in the *cth* category at time *t*. $pe_{mic,t}$ is the performance of the *mth* fund of the *ith* family belonging to the *cth* category at time *t*. $\mathbf{Controls_{mic,t}}$ is a vector of control variables we define below. $\Gamma_{c,t}$ is the index of product dispersion for the *cth* category at time *t*. We consider three ways of constructing the index ("non-performance based", "performance-based", "mixed"). We also consider two different ways of aggregating its components: weighted average ("weighted index") and simple average ("non-weighted index"). We consider two alternative standardizations of the index of dispersion: the single components of the index are standardized either using their long-term value (Panel C, "Performance and changes of market dispersion") or using the average value of the components (Panels A and B, "Performance and market dispersion"). The control variables include the risk of the fund, the overall average performance of the family the fund belongs to and of the competing families, the average risk of the family and of the competing families both overall and in the category and the average total net assets of the family the fund belongs to and of the competing families. The average Risk is constructed as the average of the standard deviations of the twelve monthly performances. Average are taken either across the funds operating in the same category ("in the category") or across all the funds of the same family over all categories ("overall"). The observations are yearly. The estimations are done using a Dynamic Panel Data estimation, based on GMM with instrumental variables and White's adjusted heteroscedastic consistent least-squares variance-covariance matrix. The model is estimated in Orthogonal Deviations. Lagged levels and differences of the explanatory variables are used as instruments. The Sargan test of over-identifying restrictions as well as tests of First Order Autocorrelation and Second Order Autocorrelation are reported. These tests are based on the standardized average residual covariances which are asymptotically distributed N(0,1) under the null of no autocorrelation. The Sargan test is asymptotically distributed as a chi-square with as many degrees of freedom as the over-identifying restrictions under the null hypothesis of the validity of the instruments. Total net assets are standardized by dividing them by, 1,000,000. * means *p-value* > 0.1, ** means *p-value* < 0.1

Panel A: Fund performance and market dispersion I

	Specifications											
	Weighted Index						*Non-Weighted Index*					
Explanatory Variables	*Non-performance related Index*		*Performance related Index*		*Mixed Index*		*Non-performance related Index*		*Performance related Index*		*Mixed Index*	
	Value	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*
Γ_{ct}	-0.08	-5.38	-0.02	-2.26	-0.02	-2.28	-0.20	-4.64	-0.01	-2.31	-0.02	-2.31
Control Variables												
Lagged Fund Performance	-0.03	-0.46	-0.31	-5.71	-0.31	-5.65	-0.32	-6.54	-0.37	-7.57	-0.38	-7.57
Family Performance	0.002	4.05	0.001	4.64	0.001	4.63	0.002	5.36	0.001	6.52	0.001	6.52
Fund Risk	-2.34	-8.83	-3.26	-6.92	-3.26	-6.94	-2.52	-11.09	-3.34	-6.91	-3.34	-6.91
Family Risk	1.17	2.17	0.09	0.18	0.11	0.21	1.61	3.77	0.31	0.81	0.31	0.81
Fund Total Load Fees	4.28	1.49	-2.59	-4.88	-2.59	-4.88	5.72	2.07	-2.90	-5.64	-2.90	-5.65
Average Overall Performance of competing families	0.67	5.13	0.46	4.32	0.47	4.35	0.33	3.65	0.38	4.58	0.38	4.58
Average Risk of competing families in the category	0.93	3.25	1.97	3.40	1.96	3.37	1.84	6.74	1.94	4.53	1.94	4.53
Average Overall Risk of competing families	-0.24	-0.55	0.80	3.09	0.80	3.07	-1.35	-3.10	0.65	2.49	0.65	2.49
Overall Total Net Assets for the competing families	-	-	-	-	-	-	-	-	-	-	-	-
Overall Total Net Assets for the family	-	-	-	-	-	-	-	-	-	-	-	-
Sargan test	0.29*		11.06*		11.01*		3.09*		12.30*		12.30*	
First Order Autocorrelation	-6.57**		-6.54**		-6.56**		-7.02**		-6.92**		-6.92**	
Second Order Autocorrelation	-1.37*		0.59*		0.61*		-0.59*		0.63*		0.63*	

Panel B: Fund performance and market dispersion I

	Specifications											
	Weighted Index						*Non-Weighted Index*					
Explanatory Variables	*Non-performance related Index*		*Performance related Index*		*Mixed Index*		*Non-performance related Index*		*Performance related Index*		*Mixed Index*	
	Value	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*
Γ_{ct}	-0.74	-3.27	-0.01	-2.02	-0.02	-2.04	-0.55	-5.50	-0.02	-2.35	-0.02	-2.27
Control Variables												
Lagged Fund Performance	0.86	3.00	-0.31	-4.15	-0.31	-4.15	-0.63	-4.06	-0.30	-4.07	-0.30	-4.02
Family Performance	0.001	2.77	0.001	4.29	0.001	4.30	0.004	4.82	0.001	4.47	0.001	4.45
Fund Risk	-2.14	-1.71	-3.29	-6.39	-3.29	-6.40	-1.97	-3.24	-3.22	-6.62	-3.24	-6.66
Family Risk	8.28	1.07	-0.09	-0.30	-0.08	-0.26	6.19	2.36	0.23	0.64	0.16	0.31
Fund Total Load Fees	2.32	0.80	-2.44	-4.88	-2.44	-4.90	5.65	1.44	-2.38	-4.68	-2.39	-4.64
Average Overall Performance of competing families	3.99	5.92	0.48	3.76	0.48	3.77	-0.18	-0.71	0.51	4.05	0.51	4.03
Average Risk of competing families in the category	-0.77	-0.84	2.25	5.05	2.24	5.03	-0.59	-1.81	1.76	4.08	1.84	2.75
Average Overall Risk of competing families	-3.74	-0.60	0.79	3.04	0.79	3.03	-4.62	-2.13	0.84	3.13	0.85	3.12
Overall Total Net Assets for the competing families	-0.04	-0.93	0.22	0.40	0.11	0.40	-0.01	-1.21	-0.11	-0.62	-0.12	-0.42
Overall Total Net Assets for the family	3.24	0.11	-12.23	-1.41	-12.34	-1.41	105	3.62	-14.13	-1.57	-15.10	-1.60
Sargan test	0.27*		10.81*		10.75*		0.17*		8.67*		8.32*	
First Order Autocorrelation	-1.94**		-6.51**		-6.53**		-4.47**		-6.96**		-6.97**	
Second Order Autocorrelation	0.02*		0.30*		0.32*		-0.15*		0.89*		-0.56*	

Panel C: Fund performance and market dispersion II

	Specifications											
	Weighted Index						*Non-Weighted Index*					
Explanatory Variables	*Non-performance related Index*		*Performance related Index*		*Mixed Index*		*Non-performance related Index*		*Performance related Index*		*Mixed Index*	
	Value	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*
Γ_{ct}	-0.88	-2.09	-0.01	-2.07	-0.02	-2.08	-0.48	-4.79	-0.02	-2.36	-0.02	-2.38
Control Variables												
Lagged Fund Performance	1.13	1.99	-0.30	-4.12	-0.30	-4.06	-0.78	-3.95	-0.30	-4.07	-0.30	-4.11
Family Performance	0.001	1.42	0.001	4.32	0.001	4.30	0.005	4.84	0.001	4.47	0.001	4.49
Fund Risk	-3.45	-1.48	-3.27	-6.44	-3.26	-6.37	-2.21	-3.21	-3.22	-6.63	-3.24	-6.62
Family Risk	14.38	0.99	-0.07	-0.23	-0.09	-0.28	7.77	2.30	0.24	0.65	0.22	0.62
Fund Total Load Fees	2.83	0.67	-2.43	-4.89	-2.40	-4.63	6.19	1.41	-2.38	-4.68	-2.38	-4.69
Average Overall Performance of competing families	3.95	3.70	0.49	3.86	0.49	3.88	-0.55	-1.70	0.51	4.05	0.50	4.02
Average Risk of competing families in the category	-0.68	-0.51	2.22	5.04	2.21	4.99	-0.69	-2.57	1.77	4.08	1.80	4.17
Average Overall Risk of competing families overall	-8.66	-0.75	0.79	3.06	0.81	2.97	-6.10	-2.16	0.83	3.12	0.82	3.05
Overall Total Net Assets for the competing families	0.2	0.17	0.1	0.33	0.3	0.32	-0.1	-0.18	-0.1	-0.60	-0.1	-0.49
Overall Total Net Assets for the family	-14.10	-0.36	-12.45	-1.42	-12.10	-1.28	127.12	3.79	-14.78	-1.57	-15.34	-1.59
Sargan test	0.06*		10.85*		10.69*		0.02*		8.64*		8.45*	
First Order Autocorrelation	-1.31**		-6.59**		-6.55**		-4.06**		-6.97**		-6.94**	
Second Order Autocorrelation	0.79*		0.29*		0.27*		-0.06*		0.90*		0.83*	

Table 5: Family performance in the category and market dispersion

---

We estimate: $r_{ic,t} = \alpha + \beta\Gamma_{c,t} + \boldsymbol{\gamma}\mathbf{Controls}_{\mathbf{ic,t}} + \delta r_{ic,t-1} + e_{ic,t}$, where the subscripts "*i*", "*c*" and "*t*" refer, respectively, to the *ith* family, in the *cth* category at time *t*. $r_{ic,t}$ is the average performance of all the funds the *ith* family has in the *cth* category at time *t*. $\mathbf{Controls}_{\mathbf{ic,t}}$ is a vector of control variables we define below. $\Gamma_{c,t}$ is the index of product dispersion for the *cth* category at time *t*. We consider three ways of constructing the index ("non-performance based", "performance-based", "mixed"). We also consider two different ways of aggregating its components: weighted average ("weighted index") and simple average ("non-weighted index") and two alternative standardizations of the index of dispersion: the single components of the index are standardized either using their long-term value ("Performance and changes of market dispersion") or using the average value of the components ("Performance and market dispersion"). The control variables include the risk and expense ratio of the family in the category, the overall family performance, risk and expense ratio, the performance, risk and expense ratio of the competing families in the category as well as their overall performances, risk and expense ratios across categories. We also consider the total load fees charged by the family as well as those charged by the competing families. The average Risk is constructed as the average of the standard deviations of the twelve monthly performances. Averages are taken either across the funds operating in the same category ("in the category") or across all the funds of the same family over all categories ("overall"). The observations are yearly. The estimations are done using a Dynamic Panel Data estimation, based on GMM with instrumental variables and White's adjusted heteroscedastic consistent least-squares variance-covariance matrix. The model is estimated in Orthogonal Deviations. Lagged levels and differences of the explanatory variables are used as instruments. The Sargan test of over-identifying restrictions as well as tests of First Order Autocorrelation and Second Order Autocorrelation are reported. These tests are based on the standardized average residual covariances which are asymptotically distributed N(0,1) under the null of no autocorrelation. The Sargan test is asymptotically distributed as a chi-square with as many degrees of freedom as the over-identifying restrictions under the null hypothesis of the validity of the instruments. * means *p-value* > 0.1, ** means *p-value* < 0.1

Panel A: Family performance in the category and market dispersion I

	Specifications											
	Weighted Index						*Non-Weighted Index*					
Explanatory Variables	*Non-performance related Index*		*Performance related Index*		*Mixed Index*		*Non-performance related Index*		*Performance related Index*		*Mixed Index*	
	Value	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*
Γ_{ct}	-0.04	-2.38	-0.005	-2.13	-0.006	-2.61	-0.14	-2.47	-0.02	-2.32	-0.02	-2.32
Control Variables												
Lagged Family Performance in categ.	0.13	1.98	-0.01	-0.53	-0.02	-1.06	0.03	0.62	0.08	1.12	0.08	1.11
Overall Family Performance	0.81	6.07	0.27	1.43	0.31	1.79	0.80	6.17	0.79	4.55	0.79	4.56
Average Performance of competing families in categ.	0.17	0.38	0.93	24.62	0.91	23.26	0.37	0.83	0.23	0.39	0.23	0.40
Overall Average Performance of competing families	0.18	0.50	-0.21	-1.29	-0.24	-1.55	-0.12	-0.39	0.06	0.14	0.06	0.14
Family Risk in categ.	-0.76	-0.97	-0.62	-2.51	-0.67	-2.83	-1.05	-2.02	-1.19	-1.65	-1.19	-1.65
Overall Family Risk	0.69	1.06	-0.23	-1.21	-0.26	-1.72	1.05	2.63	1.01	1.94	1.01	1.95
Average Risk of competing families in categ.	0.30	1.54	0.65	2.54	0.69	2.94	0.24	1.36	0.31	1.33	0.31	1.33
Overall Average Risk of competing families	-0.24	-1.47	0.21	1.28	0.26	1.81	-0.43	-1.73	-0.12	-0.39	-0.13	-0.40
Total Load Fees of the family in categ.	0.58	0.87	0.05	0.54	0.08	0.85	1.08	1.80	1.04	1.41	1.05	1.43
Total Load Fees of the competing families in categ.	-1.81	-0.84	-0.10	-0.29	-0.23	-0.57	-5.94	-2.70	-1.99	-1.06	-2.02	-1.08
Expense Ratio of the family in categ.	0.03	0.02	-0.73	-0.70	-0.63	-0.77	0.60	0.49	0.76	0.52	0.77	0.53
Overall Expense Ratio of the family	-0.31	-0.46	0.72	0.57	0.51	0.46	-0.57	-0.72	-0.65	-0.64	-0.66	-0.87
Expense Ratio of competing families in categ.	0.53	0.06	1.44	1.47	0.97	1.14	-1.68	-0.29	6.59	0.73	6.56	0.73
Overall Expense Ratio of competing families	-0.47	-0.09	-0.92	-0.89	-0.49	-0.51	-2.13	-0.48	-4.84	-0.83	-4.82	-0.83
Sargan test	2.08*		5.48*		4.08*		4.18*		2.04*		2.06*	
First Order Autocorrelation	-2.05**		-4.54**		-5.02**		-3.82**		-1.91**		-1.92**	
Second Order Autocorrelation	0.39*		0.71*		0.81*		0.81*		0.87*		0.88*	

Panel B: Family performance in the category and market dispersion II

	Specifications											
	Weighted Index						*Non-Weighted Index*					
Explanatory Variables	*Non-performance related Index*		*Performance related Index*		*Mixed Index*		*Non-performance related Index*		*Performance related Index*		*Mixed Index*	
	Value	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*
Γ_{ct}	-0.04	-2.29	-0.005	-2.28	-0.006	-2.64	-0.10	-2.92	-0.01	-2.33	-0.01	-2.35
Control Variables												
Lagged Family Performance in categ.	0.12	1.97	-0.01	-0.59	-0.01	-1.02	0.05	0.98	0.08	1.10	0.08	1.10
Overall Family Performance	0.81	6.21	0.24	1.38	0.31	1.82	0.72	6.69	0.78	4.55	0.78	4.60
Average Performance of competing families in categ.	0.20	0.48	0.94	24.92	0.91	21.40	0.75	2.46	0.22	0.39	0.23	0.40
Average Performance of competing families overall	0.12	0.35	-0.21	-1.36	-0.24	-1.61	-0.35	-1.75	0.06	0.13	0.05	0.12
Family Risk in categ.	-0.79	-1.02	-0.50	-2.39	-0.67	-2.84	-0.66	-1.95	-1.19	-1.65	-1.19	-1.66
Overall Family Risk	0.72	1.12	-0.09	-0.58	0.26	-1.71	0.81	2.67	1.01	1.94	1.01	1.96
Average Risk of competing families in categ.	0.29	1.52	0.51	2.38	0.69	2.95	0.12	0.88	0.31	1.33	0.31	1.34
Average Risk of competing families overall	-0.24	-1.51	0.08	0.64	0.25	1.79	-0.46	-2.25	-0.12	-0.39	-0.13	-0.42
Total Load Fees of the family in categ.	0.65	0.99	0.01	0.12	0.08	0.82	1.20	2.24	1.04	1.41	1.05	1.44
Total Load Fees of the competing families in categ.	-1.58	-0.74	0.13	0.42	-0.22	-0.56	-4.78	-2.78	-1.98	-1.06	-2.01	-1.08
Expense Ratio of the family in categ.	0.08	0.06	-0.19	-0.19	-0.56	-0.68	0.44	0.39	0.76	0.53	0.80	0.56
Expense Ratio of competing families in categ.	-0.34	-0.50	0.16	0.13	0.37	0.34	-0.70	-0.88	-0.65	-0.87	-0.67	-0.90
Expense Ratio of competing families overall	0.11	0.01	0.86	0.96	0.83	0.97	-8.33	-1.46	6.57	0.73	6.45	0.72
Overall Expense Ratio of competing families	-0.04	-0.01	-0.30	-0.32	-0.28	-0.30	2.09	0.68	-4.79	-0.82	-4.71	-0.82
Sargan test	2.55*		8.28*		4.45*		4.69*		2.07*		2.15*	
First Order Autocorrelation	-2.07**		-4.95**		-5.03**		-4.48**		-1.91**		-1.94**	
Second Order Autocorrelation	0.18*		0.40*		0.76*		0.22*		0.87*		0.88*	

Table 6: Family overall performance and market dispersion

We estimate: $r_{i,t} = \alpha + \beta\Gamma_{i,t} + \boldsymbol{\gamma}\mathbf{Controls_{i,t}} + \delta r_{i,t-1} + e_{i,t}$, where the subscripts "*i*" and "*t*" refer, respectively, to the *ith* family at time *t*. $r_{i,t}$ is the average performance of the all the funds that *ith* family manages across all the categories at time *t*. $\mathbf{Controls_{i,t}}$ is a vector of control variables we define below. $\Gamma_{i,t}$ is the index that proxies for the average degree of product dispersion of all the categories the families has funds in at time *t*. We consider the three ways of constructing the index ("non-performance based", "performance-based", "mixed"), two different ways of aggregating its components: weighted average ("weighted index") and simple average ("non-weighted index"). We also consider two alternative standardizations of the index of dispersion: the single components of the index are standardized either using their long-term value ("Performance and changes of market dispersion") or using the average value of the components ("Performance and market dispersion"). For each family the family specific index of product dispersion is constructed by weighting the indexes of all the categories the family is operating in, by the total net assets of the funds its is running in such categories. The control variables include the average performance and of the competing families, the average total load fees and expense ratios of both the family and the competing families. The observations are yearly. The estimations are done using a Dynamic Panel Data estimation, based on GMM with instrumental variables and White's adjusted heteroscedastic consistent least-squares variance-covariance matrix. The model is estimated in Orthogonal Deviations. Lagged levels and differences of the explanatory variables are used as instruments. The Sargan test of over-identifying restrictions as well as tests of First Order Autocorrelation and Second Order Autocorrelation are reported. These tests are based on the standardized average residual covariances which are asymptotically distributed N(0,1) under the null of no autocorrelation. The Sargan test is asymptotically distributed as a chi-square with as many degrees of freedom as the over-identifying restrictions under the null hypothesis of the validity of the instruments. * means *p-value* > 0.1, ** means *p-value* < 0.1

Panel A: Family performance and market dispersion I

Explanatory Variables	*Specifications*											
	Weighted Index						*Non-Weighted Index*					
	Non-performance related Index		*Performance related Index*		*Mixed Index*		*Non-performance related Index*		*Performance related Index*		*Mixed Index*	
	Value	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*
$\Gamma_{i,t}$	-0.31	-2.73	-0.02	-2.21	-0.19	-2.70	-0.66	-2.56	-0.03	-4.62	-0.02	2.99
Control Variables												
Lagged Family Performance	-0.35	-2.33	-0.05	-0.97	-0.73	-2.94	-1.28	-3.17	-0.31	-4.59	0.38	1.46
Average Performance of competing families	1.66	3.54	1.21	15.67	1.13	9.99	0.85	4.49	0.97	22.74	1.69	6.21
Family Total Load Fees	-4.27	-2.88	0.04	0.30	0.03	0.10	-3.14	-1.53	-0.27	-0.95	-0.17	-0.48
Total Load Fees of competing families	-5.94	-1.58	4.43	5.38	16.14	2.51	-3.72	-2.71	-4.87	-4.39	-8.63	-4.29
Family Expense Ratio	-1.02	-0.74	0.02	0.01	2.02	1.40	2.58	0.60	0.25	0.32	-0.76	-1.77
Expense Ratio of competing families	34.77	2.72	3.49	1.56	29.24	2.79	12.54	2.49	5.41	4.14	5.86	4.36
Sargan test	0.06*		0.78*		0.13*		0.33*		0.09*		1.56*	
First Order Autocorrelation	-4.35**		-3.09**		-2.49**		-1.79**		-7.14**		-4.33**	
Second Order Autocorrelation	-0.05*		-0.42*		-0.33*		1.07*		-0.93*		0.17*	

Panel B: Family performance and market dispersion II

	Specifications											
	Weighted Index						*Non-Weighted Index*					
Explanatory Variables	*Non-performance related Index*		*Performance related Index*		*Mixed Index*		*Non-performance related Index*		*Performance related Index*		*Mixed Index*	
	Value	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*
$\Gamma_{i,t}$	-0.47	-2.67	-0.03	-3.23	-0.35	-2.18	-0.46	-2.78	-0.03	-4.61	-0.02	-3.01
Control Variables												
Lagged Family Performance	-0.22	-1.21	-0.08	-1.42	-1.03	-2.21	-1.20	-3.53	-0.31	-4.59	0.37	1.44
Average Performance of competing families	2.34	3.11	1.15	12.67	1.78	4.66	0.91	5.03	0.97	22.67	1.69	6.21
Family Total Load Fees	-3.23	-2.17	-0.01	-0.10	0.06	0.13	-2.40	-1.41	-0.27	-0.96	-0.17	-0.49
Total Load Fees of competing families	-10.86	-1.90	3.94	4.02	27.94	2.13	-30.36	-2.98	-4.90	-4.39	-8.77	-4.31
Family Expense Ratio	0.18	0.10	-0.20	-0.13	3.06	1.35	2.20	0.55	0.26	0.33	-0.76	-1.77
Expense Ratio of competing families	57.54	2.65	5.56	2.28	55.21	2.19	12.83	2.86	5.51	4.15	5.96	4.36
Sargan test	0.02*		0.50*		0.22*		0.55*		0.09*		1.55*	
First Order Autocorrelation	-3.92**		-2.67**		-2.38**		-1.85**		-7.03**		-4.32**	
Second Order Autocorrelation	-2.32*		-036*		-0.90*		-1.02*		-0.94*		0.17*	

Table 7: Category performance and market dispersion

We estimate: $r_{c,t} = \alpha + \beta\Gamma_{c,t} + \boldsymbol{\gamma}\mathbf{Controls}_{\mathbf{c,t}} + \delta r_{c,t-1} + e_{c,t}$, where the subscripts "c" and "*t*" refer, respectively, to the *cth* category at time *t*. $r_{c,t}$ is the average performance of the all the funds in the *cth* category at time *t*. $\mathbf{Controls}_{\mathbf{c,t}}$ is a vector of control variables we define below. $\Gamma_{c,t}$ is the index that proxies for the average degree of product dispersion of the *cth* category at time *t*. We consider the three ways of constructing the index ("non-performance based", "performance-based", "mixed"), and two different ways of aggregating its components: weighted average ("weighted index") and simple average ("non-weighted index"). We also consider two alternative standardizations of the index of dispersion: the single components of the index are standardized either using their long-term value ("Performance and changes of market dispersion") or using the average value of the components ("Performance and market dispersion"). The control variables include the three Fama and French factors (i.e, Market, SMB and HML) and the riskless rate (i.e., performance on the 3 month T-Bill rate). We also include the average risk of the category. The average Risk is constructed as the average of the standard deviations of the twelve monthly performances of all the funds in the category. Averages are taken across all the funds in the category. The observations are *monthly*. The estimations are done using a Dynamic Panel Data estimation, based on GMM with instrumental variables and White's adjusted heteroscedastic consistent least-squares variance-covariance matrix. The model is estimated in Orthogonal Deviations. Lagged levels and differences of the explanatory variables are used as instruments. The Sargan test of over-identifying restrictions as well as tests of First Order Autocorrelation and Second Order Autocorrelation are reported. These tests are based on the standardized average residual covariances which are asymptotically distributed N(0,1) under the null of no autocorrelation. The Sargan test is asymptotically distributed as a chi-square with as many degrees of freedom as the over-identifying restrictions under the null hypothesis of the validity of the instruments. * means *p-value* > 0.1, ** means *p-value* < 0.1

Panel A: Category performance and market dispersion I

	Specifications											
	Weighted Index						*Non-Weighted Index*					
Explanatory Variables	*Non-performance related Index*		*Performance related Index*		*Mixed Index*		*Non-performance related Index*		*Performance related Index*		*Mixed Index*	
	Value	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*
Γ_{ct}	-0.10	-3.13	-3.75	-7.54	-3.47	-8.25	-0.10	-8.39	-6.77	-4.54	-7.14	-6.41
Control Variables												
Lagged Performance	-0.26	-6.92	-0.40	-7.74	-0.43	-7.32	-0.28	-14.22	-0.39	-8.10	-0.43	-9.79
Risk	3.92	4.39	4.20	5.40	4.64	5.68	4.09	8.27	5.21	8.81	6.15	8.01
Market	0.002	2.36	0.004	10.66	0.004	10.08	0.003	6.75	0.006	9.70	0.006	8.70
SMB	0.0004	1.14	0.001	4.74	0.002	6.00	0.001	6.78	0.002	7.13	0.003	8.19
HML	-0.0007	-1.07	-0.0006	-1.45	-0.0002	-0.55	0.001	3.04	0.0006	0.93	0.0007	1.21
Riskless Rate	0.10	6.38	0.14	3.37	0.17	4.65	0.10	7.18	0.27	8.25	0.28	7.90
Sargan test	10.98*		8.65*		10.32*		17.44*		12.29*		12.32*	
First Order Autocorrelation	-2.51**		-2.99**		-3.02**		-2.50**		-3.34**		-3.28**	
Second Order Autocorrelation	0.84*		-0.32*		-0.60*		0.70*		0.92*		0.30*	

Panel B: Category performance and market dispersion II

	Specifications											
	Value	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*
Explanatory Variables												
Γ_{ct}	-0.10	-3.28	-0.58	-2.86	-0.32	-2.56	-0.10	-8.22	-1.25	-3.86	-1.20	-5.29
Control Variables												
Lagged Performance	-0.26	-6.95	-0.29	-6.94	-0.27	-8.66	-0.28	-14.06	-0.34	-3.21	-0.34	-3.38
Risk	3.85	4.19	2.73	3.14	2.31	2.98	4.06	8.19	3.67	4.48	3.71	4.75
Market	0.002	2.39	0.003	4.58	0.002	3.05	0.003	6.91	0.007	6.29	0.007	6.85
SMB	0.0005	1.39	-0.00003	-0.74	0.0008	2.71	0.001	6.51	0.0007	1.13	0.0008	1.26
HML	-0.0007	-1.19	-0.0006	-0.89	-0.001	-2.04	0.001	2.80	0.002	3.52	0.003	3.74
Riskless Rate	0.10	6.83	0.14	6.49	0.008	5.11	0.10	6.93	0.18	7.57	0.20	8.04
Sargan test	10.94*		11.29*		10.17*		17.53*		14.52*		15.12*	
First Order Autocorrelation	-2.51**		-2.41**		-2.45**		-2.48**		-3.22**		-3.33**	
Second Order Autocorrelation	0.84*		0.609*		0.67*		0.73*		1.15*		1.30*	

Table 8: Causality between performance and market dispersion

We estimate the following system of equations:

$$\begin{cases} r_{c,t} = \sum_{s=1}^{K} \alpha_{1,s} r_{c,t-s} + \sum_{s=1}^{K} \beta_{1,s} \Gamma_{c,t-s} + \eta_{c,s} + \varepsilon_{1,c,t} \\ \Gamma_{c,t} = \sum_{s=1}^{K} \alpha_{2,s} r_{c,t-s} + \sum_{s=1}^{K} \beta_{1,s} \Gamma_{c,t-s} + \eta_{c,s} + \varepsilon_{2,c,t} \end{cases}$$

where $r_{i,t}$ is the average performance of the funds belonging to the *cth* category at time *t*, $\Gamma_{c,t}$ is the degree of product dispersion of *cth* category at time *t*, η_c represents a category specific effect and *k* is the number of lags. We consider three ways of constructing the index (“non-performance based”, “performance-based”, “mixed”) and two different ways of aggregating its components: weighted average (“weighted index”) and simple average (“non-weighted index”). We also consider two alternative standardizations of the index of dispersion: the single components of the index are standardized either using their long-term value (“Performance and changes of market dispersion”) or using the average value of the components (“Performance and market dispersion”). The control variables include the past performance of the category, the three Fama and French factors (i.e, Market, SMB and HML) and the riskless rate (i.e., return on the 3 month T-Bill rate). We also include the average risk of the category. The average Risk is constructed as the average of the standard deviations of the twelve monthly performances of all the funds in the category. Averages are taken across all the funds in the category. The observations are yearly. The estimations are done using a Dynamic Panel Data estimation, based on GMM with instrumental variables and White's adjusted heteroscedastic consistent least-squares variance-covariance matrix. The model is estimated in Orthogonal Deviations. Lagged levels and differences of the explanatory variables are used as instruments. The Sargan test of over-identifying restrictions as well as tests of First Order Autocorrelation and Second Order Autocorrelation are reported. These tests are based on the standardized average residual covariances which are asymptotically distributed N(0,1) under the null of no autocorrelation. The Sargan test is asymptotically distributed as a chi-square with as many degrees of freedom as the over-identifying restrictions under the null hypothesis of the validity of the instruments. We report the χ^2 and the *p-values* of the test of causality. * means *p-value* > 0.1, ** means *p-value* < 0.1

Panel A: Causality between performance and market dispersion I

	Weighted Index						*Non-Weighted Index*					
	Non-performance related Index		*Performance related Index*		*Mixed Index*		*Non-performance related Index*		*Performance related Index*		*Mixed Index*	
	χ^2	p-values	χ^2	p-values	χ^2	p-values	χ^2	p-values	χ^2	p-values	χ^2	p-values
Dispersion => Performance	36.09	0.00	19.44	0.00	20.32	0.00	12.38	0.00	4.31	0.22	9.69	0.02
Sargan	3.80		2.22		1.80		6.76		6.09		4.44	
Test of first order autocorrelation	-2.84		-1.95		-2.15		-1.82		-5.45		-5.98	
Test of second order autocorrelation	-0.75		-0.81		-0.71		0.47		-0.82		0.38	
Performance => Dispersion	5.50	0.13	40.39	0.00	40.15	0.00	5.71	0.12	100.29	0.00	85.21	0.00
Sargan	4.45*		1.74*		1.93*		4.15*		3.52*		3.08*	
Test of first order autocorrelation	-1.75**		-4.23**		-4.28**		-3.49**		-2.23**		-2.63**	
Test of second order autocorrelation	0.19*		1.25*		1.18*		0.31*		0.56*		0.62*	

Panel B: Causality between performance and market dispersion II

	Weighted Index						*Non-Weighted Index*					
	Non-performance related Index		*Performance related Index*		*Mixed Index*		*Non-performance related Index*		*Performance related Index*		*Mixed Index*	
	χ^2	p-values	χ^2	p-values	χ^2	p-values	χ^2	p-values	χ^2	p-values	χ^2	p-values
Dispersion => Performance	29.61	0.00	27.56	0.00	27.36	0.00	15.67	0.00	8.07	0.04	8.45	0.03
Sargan	2.59		2.88		3.55		8.90		2.51		2.49	
Test of first order autocorrelatation	-2.08		-1.66		-1.66		-1.71		-6.22		-6.13	
Test of second order autocorrelation	-0.94		-0.65		-0.89		-0.71		-1.32		-1.25	
Performance => Dispersion	2.72	0.43	24.65	0.00	20.45	0.00	7.77	0.09	39.01	0.00	23.10	0.00
Sargan	1.69*		0.85*		0.90*		3.53*		10.61*		0.81*	
Test of first order autocorrelation	-1.65**		-2.64**		-2.59**		-2.26**		-1.74**		-1.78**	
Test of second order autocorrelation	0.42*		0.04*		0.003*		1.01*		-0.12*		-0.50*	

Table 9: Fund Proliferation
Number of funds and market dispersion

We estimate: $N_{i,c,t} = \alpha + \beta\Gamma_{c,t} + \boldsymbol{\gamma}\mathbf{Controls_{i,t}} + \delta N_{i,c,t-1} + e_{i,c,t}$, where the subscripts "c" and "*t*" refer, respectively, to the *cth* family at time *t*. $N_{i,c,t}$ is the number of funds of the *ith* family in the *cth* category at time *t*. $\mathbf{Controls_{i,t}}$ is a vector of control variables we define below. $\Gamma_{c,t}$ is index of product dispersion of the *cth* category at time *t*. We consider three ways of constructing the index ("non-performance based", "performance-based", "mixed" and two different ways of aggregating its components: weighted average ("weighted index") and simple average ("non-weighted index"). We also consider two alternative standardizations of the index of dispersion: the single components of the index are standardized either using their long-term value ("Performance and changes of market dispersion") or using the average value of the components ("Performance and market dispersion"). The control variables include the average performance of the family and of the competing families in the category, the average total load fees and expense ratios of both the family and the competing families in the category and the family overall performance. Averages are taken across all the funds of the same family over all categories ("overall"). The observations are yearly. The estimations are done using a Dynamic Panel Data estimation, based on GMM with instrumental variables and White's adjusted heteroscedastic consistent least-squares variance-covariance matrix. The model is estimated in Orthogonal Deviations. Lagged levels and differences of the explanatory variables are used as instruments. The Sargan test of over-identifying restrictions as well as tests of First Order Autocorrelation and Second Order Autocorrelation are reported. These tests are based on the standardized average residual covariances which are asymptotically distributed N(0,1) under the null of no autocorrelation. The Sargan test is asymptotically distributed as a chi-square with as many degrees of freedom as the over-identifying restrictions under the null hypothesis of the validity of the instruments. * means *p-value* > 0.1, ** means *p-value* < 0.1

Panel A: Number of funds and market dispersion I

	Specifications											
	Weighted Index						*Non-Weighted Index*					
Explanatory Variables	*Non-performance related Index*		*Performance related Index*		*Mixed Index*		*Non-performance related Index*		*Performance related Index*		*Mixed Index*	
	Value	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*
Γ_{ct}	5.69	2.15	4.50	3.84	5.33	4.11	67.98	2.33	3.88	3.89	4.01	3.89
Control Variables												
Lagged Number of Funds	0.25	1.98	0.08	1.13	-0.003	-0.03	1.10	4.42	0.06	0.86	0.06	0.86
Family Performance in the category	-0.87	-1.97	-0.81	-2.22	-0.77	-2.01	1.71	1.61	-1.78	-4.84	1.78	-4.84
Overall Family Performance	-0.35	-0.02	-36.49	-3.17	-46.46	-3.62	-138.51	-1.92	-40.50	-3.35	-41.04	-3.37
Average Performance of competing families in the category	29.27	0.17	331.80	3.24	361.69	3.36	119.86	0.29	345.61	3.31	345.36	3.31
Family Total Load Fees	-397.84	-1.64	-1171.35	-5.57	-1259.00	5.77	2301.69	1.87	-1286.07	-5.79	-1279.27	-5.78
Total Load Fees of competing families	50.04	0.69	36.55	0.90	38.77	0.87	787.94	0.98	32.73	0.82	32.81	0.82
Family Expense Ratio in the category	-48.25	-0.10	751.33	2.05	797.23	2.07	-913.35	-0.98	1394.36	4.45	1391.82	4.44
Expense Ratio of competing families in the category	0.0002	1.63	0.0001	1.78	0.0001	1.83	0.0001	0.68	0.0001	1.52	0.0001	1.52
Sargan test	4.47*		11.98*		6.41*		7.23*		10.71*		10.73*	
First Order Autocorrelation	-4.22**		-5.19**		-5.02**		-1.66**		-5.27**		-5.27**	
Second Order Autocorrelation	0.90*		0.71*		0.15*		1.49*		0.57*		0.57*	

Panel B: Number of funds and market dispersion II

Explanatory Variables	*Specifications*											
	Weighted Index						*Non-Weighted Index*					
	Non-performance related Index		*Performance related Index*		*Mixed Index*		*Non-performance related Index*		*Performance related Index*		*Mixed Index*	
	Value	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*
Γ_{ct}	5.71	2.11	4.37	3.86	5.13	4.13	55.90	2.37	3.93	3.89	4.07	3.89
Control Variables												
Lagged Number of Funds	0.23	1.81	0.08	1.10	-0.002	-0.03	1.11	4.41	0.06	0.85	0.06	0.85
Family Performance in the category	-0.82	1.88	-0.94	-2.66	-0.93	-2.54	1.86	1.69	-1.77	-4.83	-1.77	-4.83
Overall Family Performance	-0.74	-0.05	-35.32	-3.11	-44.48	-3.55	-152.77	-2.01	-40.35	-3.34	-41.05	-3.38
Average Performance of competing families in the category	30.14	0.18	330.80	3.24	357.28	3.34	139.85	0.33	344.96	3.31	344.68	3.31
Family Total Load Fees	-425.23	-1.80	-1167.96	-5.57	-1248.41	-5.76	2540.81	1.93	-1280.67	-5.79	-1271.69	-5.77
Total Load Fees of competing families	51.21	0.70	35.80	0.88	38.07	0.86	789.05	0.99	32.78	0.82	32.88	0.82
Family Expense Ratio in the category	-50.51	-0.10	746.42	2.04	795.01	2.07	-1189.47	-1.18	1385.95	4.42	1381.69	4.41
Expense Ratio of competing families in the category	0.0002	1.67	0.0001	1.79	0.0001	1.84	0.0001	0.65	0.0001	1.53	0.0001	1.53
Sargan test	4.47*		11.88*		6.54*		6.65*		10.73*		10.75*	
First Order Autocorrelation	-4.21**		-5.19**		-5.01**		-1.78**		-5.27**		-5.27**	
Second Order Autocorrelation	0.79*		0.69*		0.15*		1.54*		0.57*		0.56*	

Table 10: Fund Proliferation
Family coverage and market dispersion

We estimate: $C_{i,t} = \alpha + \beta\Gamma_{i,t} + \boldsymbol{\gamma}\mathbf{Controls_{i,t}} + \delta C_{i,t-1} + e_{i,t}$, where the subscripts "c" and "*t*" refer, respectively, to the *cth* family at time *t*. $C_{i,t}$ is the degree of market coverage, that is number of categories the *ith* family is present at time *t*. $\mathbf{Controls_{i,t}}$ is a vector of control variables we define below. $\Gamma_{i,t}$ is the index that proxies for the average degree of product dispersion of all the categories the families has funds in at time *t*. We consider the three ways of constructing the index ("non-performance based", "performance-based", "mixed"), and two different ways of aggregating its components: weighted average ("weighted index") and simple average ("non-weighted index"). We also consider two alternative standardizations of the index of dispersion: the single components of the index are standardized either using their long-term value ("Performance and changes of market dispersion") or using the average value of the components ("Performance and market dispersion"). The family specific index of product dispersion is constructed by weighting the indexes of all the categories the family is operating in, by the total net assets of the funds it is running in such categories. The control variables include the average overall performance of the family and of the competing families, the average total load fees and expense ratios of both the family and the competing families. Averages are taken across all the funds of the same family over all categories ("overall"). The observations are yearly. The estimations are done using a Dynamic Panel Data estimation, based on GMM with instrumental variables and White's adjusted heteroscedastic consistent least-squares variance-covariance matrix. The model is estimated in Orthogonal Deviations. Lagged levels and differences of the explanatory variables are used as instruments. The Sargan test of over-identifying restrictions as well as tests of First Order Autocorrelation and Second Order Autocorrelation are reported. These tests are based on the standardized average residual covariances which are asymptotically distributed N(0,1) under the null of no autocorrelation. The Sargan test is asymptotically distributed as a chi-square with as many degrees of freedom as the over-identifying restrictions under the null hypothesis of the validity of the instruments. * means *p-value* > 0.1, ** means *p-value* < 0.1

Panel A: Family coverage and market dispersion I

	Specifications											
	Weighted Index						*Non-Weighted Index*					
Explanatory Variables	*Non-performance related Index*		*Performance related Index*		*Mixed Index*		*Non-performance related Index*		*Performance related Index*		*Mixed Index*	
	Value	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*
Γ_{it}	0.17	2.78	0.08	2.54	0.08	2.55	0.0008	4.54	0.0003	4.91	0.0003	4.86
Control Variables												
Lagged Family Coverage	0.44	1.96	-0.01	-0.5	-0.01	-0.08	0.36	1.42	0.39	3.22	0.36	3.17
Family Performance	1.38	4.46	-0.16	-0.77	-0.15	-0.73	-0.04	-0.22	0.36	1.34	0.37	1.34
Average Performance of competing families	1.79	-3.97	-0.18	-0.72	-0.21	-0.80	0.09	0.73	-0.43	-1.44	-0.46	-1.48
Family Total Load Fees	0.15	0.47	-4.82	-3.43	-4.77	-3.46	-2.71	-5.22	0.10	0.17	0.11	0.18
Total Load Fees of competing families	4.24	1.83	-4.03	-1.36	-3.94	-1.35	4.29	2.78	-1.96	-1.92	-1.94	-1.87
Family Expense Ratio	1.62	1.96	1.36	1.09	1.39	1.13	0.43	0.20	2.83	1.57	2.85	1.52
Expense Ratio of competing families	-17.30	-2.67	-9.71	-2.14	-9.99	-2.17	-2.51	-1.58	-0.65	-0.45	0.62	0.43
Sargan test	0.53*		7.97*		8.18*		6.28*		0.67*		0.80*	
First Order Autocorrelation	-6.05**		-4.15**		-4.19**		-5.04**		-1.81**		-1.67**	
Second Order Autocorrelation	1.15*		0.05*		0.02*		1.11*		1.09*		0.97*	

Panel B: Family coverage and market dispersion II

	Specifications											
	Weighted Index						*Non-Weighted Index*					
Explanatory Variables	*Non-performance related Index*		*Performance related Index*		*Mixed Index*		*Non-performance related Index*		*Performance related Index*		*Mixed Index*	
	Value	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*
Γ_{it}	0.11	2.55	0.09	2.82	0.09	2.82	0.0005	4.95	0.0003	4.89	0.0003	4.79
Control Variables												
Lagged Family Coverage	0.48	2.68	0.09	0.49	0.07	0.41	0.37	1.50	0.36	3.20	0.34	3.14
Family Performance	1.37	3.93	-0.15	-0.77	-0.13	-0.70	-0.03	-0.19	0.38	1.35	0.39	1.34
Average Performance of competing families	-1.56	-3.53	-0.24	-1.01	-0.28	-1.11	0.08	0.65	-0.47	-1.50	-0.49	-1.52
Family Total Load Fees	0.08	0.27	-4.37	-3.64	-4.32	-3.66	-2.75	-5.20	0.16	0.24	0.15	0.23
Total Load Fees of competing families	1.01	0.76	-5.73	-1.80	-5.48	-1.77	4.45	2.91	-2.01	-1.91	-1.95	-1.83
Family Expense Ratio	0.95	0.61	1.19	1.05	1.24	1.11	0.34	0.16	2.94	1.57	2.91	1.48
Expense Ratio of competing families	-11.43	-2.04	-12.50	-2.49	-12.94	-2.51	-2.67	-1.67	-0.63	-0.42	-0.61	-0.40
Sargan test	1.29*		8.28*		8.56*		6.25*		0.75*		0.96*	
First Order Autocorrelation	-7.13**		-4.47**		-4.51**		-4.78**		-1.72**		-1.57**	
Second Order Autocorrelation	1.54*		0.30*		0.25*		1.10*		1.00*		0.87*	

Table 11: Fund turnover and market dispersion

We estimate: $T_{c,t} = \alpha + \beta\Gamma_{c,t} + \boldsymbol{\gamma}\mathbf{Controls_{c,t}} + \delta T_{c,t-1} + e_{c,t}$, where the subscripts "c" and "*t*" refer, respectively, to the *cth* category at time *t*. $T_{c,t}$ is the number of new and dead funds in the *cth* category at time *t*. $\mathbf{Controls_{c,t}}$ is a vector of control variables we define below. $\Gamma_{c,t}$ is the index that proxies for the average degree of product dispersion of the *cth* category at time *t*. We consider the three ways of constructing the index ("non-performance based", "performance-based", "mixed") and two different ways of aggregating its components: weighted average ("weighted index") and simple average ("non-weighted index"). We also consider two alternative standardizations of the index of dispersion: the single components of the index are standardized either using their long-term value ("Performance and changes of market dispersion") or using the average value of the components ("Performance and market dispersion"). The control variables include the overall performance of the mutual fund industry, the excess return on the market portfolio, the average risk of the category and the lagged values of the number of fund in the category and of the total net assets of all the funds operating in such a category. The average Risk is constructed as the average of the standard deviations of the twelve monthly returns of all the funds in the category. Averages are taken across all the funds in the category. The observations are *monthly*. The estimations are done using a Dynamic Panel Data estimation, based on GMM with instrumental variables and White's adjusted heteroscedastic consistent least-squares variance-covariance matrix. The model is estimated in Orthogonal Deviations. Lagged levels and differences of the explanatory variables are used as instruments. The Sargan test of over-identifying restrictions as well as tests of First Order Autocorrelation and Second Order Autocorrelation are reported. These tests are based on the standardized average residual covariances which are asymptotically distributed N(0,1) under the null of no autocorrelation. The Sargan test is asymptotically distributed as a chi-square with as many degrees of freedom as the over-identifying restrictions under the null hypothesis of the validity of the instruments. * means *p-value* > 0.1, ** means *p-value* < 0.1

Panel A: Fund turnover and market dispersion I

	Specifications											
	Weighted Index						*Non-Weighted Index*					
Explanatory Variables	*Non-performance related Index*		*Performance related Index*		*Mixed Index*		*Non-performance related Index*		*Performance related Index*		*Mixed Index*	
	Value	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*
Γ_{ct}	47.48	3.47	692.48	3.87	708.35	4.04	9.13	3.10	4390.68	3.09	3913.19	3.03
Control Variables												
Lagged Turnover	-0.27	-3.52	-0.99	-4.98	-0.29	-5.16	-0.38	-5.50	-0.72	-2.99	-0.68	-3.04
Performance	3.07	0.18	13.27	0.38	3.20	0.08	-5.17	-0.36	-336.50	2.77	-309.63	-2.69
Risk	126.52	0.92	-263.50	-1.62	-260.36	-1.60	196.58	1.36	863.01	2.01	796.81	2.02
Overall Performance	117.38	3.13	113.65	3.09	127.68	3.24	117.26	3.49	478.48	2.64	447.17	2.65
Market	-0.38	-3.21	-0.59	-4.63	-0.59	-4.65	-0.34	-2.91	-0.75	1.18	-0.69	-1.22
Lagged Total Assets	-0.002	-2.95	-0.001	-0.66	-0.001	-0.81	-0.002	-3.56	-0.01	-2.47	-0.01	-2.39
Lagged Number of Funds	0.008	1.11	0.02	1.38	0.02	1.36	0.001	2.46	0.11	2.66	0.09	2.55
Sargan test	15.23*		11.88*		12.43*		13.53*		6.10*		7.05*	
First Order Autocorrelation	-3.50**		-3.39**		-3.38**		-3.39**		-3.67**		-3.84**	
Second Order Autocorrelation	-0.63*		-0.69*		-0.77*		-1.57*		0.87*		0.71*	

Panel B: Fund turnover and market dispersion II

	Specifications											
	Value	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*	*Value*	*t-stat*
Explanatory Variables												
Γ_{ct}	43.57	3.71	137.17	2.55	133.54	3.21	9.37	3.43	409.04	2.34	479.07	3.06
Control Variables												
Lagged Turnover	-0.27	-3.58	-0.01	0.13	-0.06	-0.71	-0.36	-5.33	0.11	1.05	0.006	0.64
Performance	5.08	0.29	-12.48	-1.47	-12.53	-1.46	-1.20	-0.08	-15.83	-1.74	-15.92	-1.68
Risk	117.79	0.85	-405.83	-1.95	-389.59	-1.96	179.37	1.23	-193.19	-1.22	-183.95	-1.15
Overall Performance	117.08	3.30	-171.30	-0.59	-170.99	-0.60	112.46	3.41	-136.16	-0.71	-149.85	-0.77
Market	-0.37	-3.13	1.11	0.84	1.11	0.85	-0.33	-2.79	0.95	1.10	1.04	1.17
Lagged Total Assets	-0.002	-2.81	-0.0001	-0.04	-0.0001	-0.08	0.002	-3.43	-0.001	-1.20	-0.0007	-0.92
Lagged Number of Funds	0.008	1.12	0.002	0.35	-0.006	-0.70	0.01	2.30	0.0003	0.02	-0.009	-0.74
Sargan test	15.09*		6.57*		4.74*		13.47*		7.85*		5.38*	
First Order Autocorrelation	-3.55**		-3.26**		-3.18**		-3.32**		-3.32**		-3.43**	
Second Order Autocorrelation	-0.57*		1.38*		1.26*		-1.44*		1.57*		1.45*	