ISSN 0956-8549-638

Organizational Diseconomies in the Mutual Fund Industry

By
Fabian Garavito

DISCUSSION PAPER NO 638

DISCUSSION PAPER SERIES

August 2009

Fabian Garavito is a Finance PhD candidate at the London School of Economics and a research assistant at the Financial Markets Group. His main research interests are in delegated asset management, behavioural finance and empirical asset pricing. He holds a Master of Science in Financial Engineering and a Master of Science in Industrial and Operations Research from the University of Michigan, Ann Arbor. He also holds a Bachelor of Science in Industrial Engineering from Universidad de los Andes, Colombia.

Organizational Diseconomies in the Mutual Fund Industry

Fabian Garavito[1]

First draft: April 2008

This version: July 2008

[1]Garavito: Financial Markets Group (FMG) and Department of Finance, London School of Economics, London WC2A 2AE, UK. Email f.garavito@lse.ac.uk. I am thankful to Antoine Faure-Grimaud, Christopher Polk Dimitri Vayanos and Michela Verardo and seminar participants at Financial Service Authority (FSA) and LSE for their comments and suggestions. I gratefully acknowledge financial support from the Financial Services Authority. The views expressed herein are my own and do not necessarily represent the views of the Financial Services Authority. All errors are my own.

Abstract

I document how the organizational form of a mutual fund affects its investment strategies. I show that centralized funds tilt their portfolios to hard information companies whereas decentralized funds tilt their portfolios to soft information companies. I also show that the investments of decentralized (centralized) mutual funds in soft (hard) information companies outperform those of centralized (decentralized) funds. Moreover, decentralized funds show ability to forecast soft information companies' future returns and a disability at forecasting hard information companies' future returns. On the other hand, centralized funds do not seem to be able to forecast the returns of hard information companies, but they show disability at forecasting hard information companies' future returns. The results corroborate the main predictions of Stein (2002). The results also shed light on the increase in demand for large stocks and the positive relationship between performance of portfolio concentration documented in the literature.

JEL classification: G14, G17, G23, L22

1 Introduction

Information collection is a central part of investing. As suggested by several economists, the collection of information can ameliorate the adverse selection problem faced by investors[2]. It has also been argued that organizational form plays and important role in providing incentives to collect information. Stein (2002) describes how decentralized organizations are better at providing incentives to collect soft information (information that is difficult to put in a numeric score) and centralized organizations are better at incentivizing agents to collect hard information (information that is quantitative in nature). Therefore, while collecting information is a very important part of investing, organizational form may dictate the kind of information that is optimal to collect. The main objective of this study is to explore whether organizational form causes mutual funds to tilt their portfolios towards certain companies based on the type of information (soft/hard information) these companies generate. I also study whether centralized (decentralized) mutual funds are better than decentralized (centralized) funds at investing in companies that mostly generate hard (soft) information. Furthermore, I investigate whether decentralized (centralized) mutual funds are able to forecast returns of soft (hard) information companies held by them.

Stein (2002) develops a theoretical framework to model the effect of organizational design (i.e. hierarchy/centralized, decentralized structure) on the collection of information in firms where division managers compete for internal funds to finance their projects. If research and capital allocation decisions are carried out by different agents (as it is the case in centralized organizations: division managers collect information on investment projects, CEO allocates funds across divisions), division managers will know that the only information a CEO will use in the capital allo-

[2]Stiglitz and Weiss (1980) show that asymmetric information may explain why capital does not flow to firms with positive net present value projects. Lelan and Pyle (1977), Campbell and Kracaw (1980), Diamond (1984), Haubrich (1989) Diamond (1991) describe how large institutional creditors can partially overcome the problem of adverse selection by producing information about firms and using it in their credit decisions

cation process is information that can be credibly transmittable (hard information: quantitative/verifiable information such as sales growth rate over the past 5 years). This, in turn, means that any effort exerted in collecting information that can not be credibly transmittable, or soft information (information that can not be easily agreed upon, i.e. honesty of a CEO), will go to waste. Division managers will know this ex-ante and will re-direct all his efforts to collect hard information. Furthermore, competition amongst division managers for (limited) internal funds will make division managers to collect as much (hard) information as possible on their investment projects, thereby creating vast amounts of (hard) information. The CEO will analyse all (hard) information provided by the division managers and will allocate funds across divisions optimally. In a decentralised organization research and capital allocation will be conducted by the same agent (division managers collect information on projects and decide on a capital allocation strategy). In this case, division managers have all incentives to collect as much information possible on their investment projects. Stein shows that if all information about investment projects is hard centralised organizations have an advantage over decentralized firms. On the other hand, he also shows that is all information about investment opportunities is soft, decentralized organizations will have superior fund allocation across projects than centralized firms.

Actively managed US equity mutual funds provide an ideal environment to analyze the effects of organizational diseconomies in information collection and capital allocation for several reasons. Firstly, investing is a task that is information intensive. Moreover, due to disclosure requirements, it is possible to measure fund organizational characteristics, and the information opaqueness (hard vs. soft information) of funds holdings. Stein's model provides many insights in the way the organizational structure affects the collection of information and the capital allocation process in mutual funds. If the organization structure affects the incentives to collect information, mutual funds with highly hierarchical/centralized organizational forms will tend to tilt their portfolio towards companies that generate a lot of hard information. Conversely, decentralized funds will tend to tilt their portfolios towards soft information companies. Additionally, decentralized funds will be better than centralized funds at investing in soft information companies since they are better equipped to utilize this information in the capital allocation process. Furthermore, if hierarchical funds tend to produce vast amounts of (hard) information, they should be able to profit from the information collected if they can keep it private. If this is the case, hierarchical funds should be better than decentralized funds at investing in companies that produce a lot hard information. I also conjecture that decentralized funds are not only better than centralized funds at investing in soft information companies, but they are also good at it, i.e. they are able to forecast soft information stocks' returns whereas centralized funds are not. A similar argument should also hold for centralized funds: if information collected remains private, centralized funds will be able to forecast hard information stocks' returns while decentralized funds will not.

If the organizational form of funds and the informational "softness" of stocks held by mutual funds were observable, it would be straight forward to test the hypotheses above. The problem is that these characteristics are not observable. In this paper, I construct scores that measure the organizational structure of a mutual fund and the type of information available about publicly traded companies. I measure organizational complexity of a fund by creating a score based on the number of managers in the fund, the fund's assets under management, the number of equity funds and the assets under management of the family to which the fund belongs. The first two variables measure the level of organizational complexity at the fund level and the other two variables measure the level of organizational complexity at the family level. I include organizational complexity at the family level, because fund families have incentives to control the investments of their sibling funds (See Gaspar, Massa and

Matos (2006), Cici Gibson Moussawi (2006)). If this is the case, the portfolio allocation at the fund level would be influenced by the management team of the family to which the fund belongs. This creates the kind of separation between research and decision making that lead managers to steer efforts to collect hard information. In order to measure the type of information available about publicly-traded companies, I construct a proxy using market capitalization, age, number of analyst reports available and institutional ownership. The first two variables measure the nature of the information available to the public. The other two variables measure the extent to which information about a company has been hardened.

My findings support the hypotheses above. I find that the organizational complexity of a fund positively co-varies with the weighted average information score of its portfolio. In other words, the more centralized a fund is, the more it tends to hold hard information companies. In addition to that, decentralized funds are better than centralized funds at investing in soft information companies. For instance, constructing a self-financing trading strategy consisting of a long positions in a portfolio of soft information companies held by decentralized funds and a short position in a portfolio in soft information companies held by hierarchical funds results in a (5-factor) risk adjusted return of 0.55% per month (6.8% per year). On the other hand, centralized funds are also better than decentralized funds at investing in hard information companies, however, to a lesser extent. The weakness could be explained by the fact that hard information is transmittable, and therefore, difficult to contain. Inability to keep information private decreases the chances a manager has to earn abnormal returns. Additionally, decentralized funds are able to forecast soft information stock returns. For instance, an increase of one standard deviation in the average abnormal portfolio tilt of decentralized funds in a soft information stock forecasts an additional 13 bps in the stock monthly return. Centralized funds do not seem to predict the returns of hard information companies. This last piece of evidence seem to support our previous conjecture: hard information is difficult to contain and therefore, it is difficult to earn abnormal profits by trading on it. Moreover, decentralized and centralized funds show an important and statistically significant *disability* when investing in hard and soft information stocks respectively.

The distinction between soft and hard information has been studied before in the banking literature, with particular emphasis on the incorporation of soft and hard information in different lending technologies (i.e. credit scoring, relationship lending) by different organizational forms (large vs. small banks).[3] One of the main conclusions in this strand of the literature is that large banks tend to be at a disadvantage when lending to small businesses. The reason given is that large banks are very centralized and small businesses tend to be informationally opaque (they mostly produce soft information). The disadvantage emerges from the fact that centralized organizations are ill-suited to use soft information.[4] However, Berger Rosen and Udell (2007) argue that past empirical research in this area is inconclusive since some variables of interests were not considered. Therefore the evidence in the banking literature on the effects of organizational form on the collection and usage of information is mixed.

Chen et al (2004) look at the issue of organizational diseconomies in the delegated asset management industry. They examine a particular cross-section of the data, September 1997, and find that small funds are more likely to invest in local stocks and are better at investing in them than large funds. Moreover, they find that the more managers a fund has, the less the fund invests in local stocks (companies whose head quarters are geographically close to the fund's main offices). They present this evidence as an indication that decentralized funds (small funds) are better at collecting and using soft information companies and as evidence in favor of Stein's

[3]For a more detailed discussion on the subject see the papers surveyed in Berger Rosen and Udell (2007).

[4]However, Berger and Udell (2006) have pointed out that large banks (hierarchies) may have developed lending technologies that allow them to lend to opaque businesses (soft information companies). Examples of these lending technologies are small business credit scoring asset-based lending, factoring, fixed assets lending and leasing (See Berger and Udell (2006))

(2002) model. However, they do not look at the other implications of Stein (2002), namely whether hierarchical funds tilt their portfolios to hard information stocks and whether they are better at using hard information. One of the problems with their setup is that they measure organizational complexity with fund size which is a very noisy proxy for organizational complexity and it neglects the effect of fund families in the way sibling fund operates. For instance, Gaspar Massa and Matos (2006) and Cici Gibson and Moussawi (2006), document that fund families have the incentives and mechanisms to influence the capital allocation of its sibling funds. This creates the kind of separation between research and decision making found in centralized organizations. Chen et al (2004) also measure organizational complexity with the number of managers that run a fund. While number of managers may be a better way to measure the separation of research and decision making than fund size, it still neglects the effects of a family in the way a fund operates. Furthermore, Chen et al do not attempt to measure the information opaqueness (soft / hard information) of stocks. As described in Berger Rosen and Udell (2007), not incorporating such information may result in biased results.

The paper is organized as follows. Section 2 describes the hypotheses. Section 3 describes the data, section 4 presents the results and section 5 concludes.

2 Theory and Hypothesis Construction

2.1 Soft and Hard Information

Petersen (2004) presents a detailed characterization of hard and soft information in finance. Hard Information is the kind of information that can be easily reduced to numbers. Examples of hard information in finance are financial statements, credit history, and stock returns. On the other hand, one can think of soft information as information that can not be completely summarized in a numeric score. Examples of soft information in finance can be opinions and rumours. Due to its quantitative nature, hard information can be easily collected, stored and transmitted (these characteristics also make it difficult to contain). A second dimension used by Petersen to characterize information is the way in which it is collected. The collection of hard information need not be personal. Therefore, the collection process can be at arms length, automated and standardized. However, it places restrictions on what can be collected. With soft information, the context under which it is collected and the collector of the information are part of the information itself. For instance, if I say the manager of a firm has great business acumen, the information depends on my definition of business acumen. One of the advantages of hard information is that it can lower production costs through standardization and automatization. Hard information is easy to store as the information does not depend on who collected it. This means that the information remains in an organization even if the agent who collected the information leaves the firm. However, collection of hard information also leads to a loss of information which in some contexts can be quite important (i.e. venture capital). Moreover, the fact that hard information is difficult to contain can keep managers from fully collecting informational rents (i.e. in the case of equity investing, it reduces the ability an investor has to earn abnormal returns).[5]

2.2 Hypotheses construction

Motivated by Stein (2002), I conjecture that the organizational form of a mutual fund affects managers' incentives to collect information. This should be reflected in the kind of stocks that managers pick and in their ability to choose stocks with different

[5]Petersen also notes that soft information can be hardened and cites credit scoring as an example. In addition, he presents examples of hardening of information (Mercantile Agency, R.G.Dun, and Bradstreets in the 1840) and explains how the evolution of financial markets over the last forty years has been in part a replacement of soft information with hard information as the basis for financial transactions

degrees of information "softness".

Hypothesis 1. Centralized mutual funds should tilt their portfolios towards hard information stocks.

Hypothesis 1 follows from the fact that incentives to collect information are affected by organizational form. As stated before, fund managers that operate in centralized funds (where research and decision making are conducted by different agents), will know that they will not be able to credibly transmit soft information. Ex-ante, they decide to only collect hard information. Therefore, one can expect that the portfolios of these funds will be tilted towards companies that produce hard information. On the other hand, we should expect decentralized funds not to have this tilt towards hard information stocks.

Hypothesis 2. Decentralized (centralized) funds are better than centralized (decentralized) funds at investing in soft (hard) information companies.

If decentralized funds are better suited than centralized funds at collecting and using soft information, they should have a superior ability to invest in soft information stocks than centralized funds. On the other hand, centralized funds should be better than decentralized funds at investing in hard information companies since they are better at gathering and incorporating hard information.

Hypothesis 3. Decentralized funds should be able to forecast the future returns of soft information stocks they hold. Centralized funds should be able to predict future stock returns of hard information companies they own. However, decentralized (centralized) funds should exhibit a disability to invest in hard (soft) information stocks

Decentralized funds should not only be better than centralized funds at investing in soft information stocks but should also be able to generate abnormal returns in their soft information stock investments. For instance, high decentralized mutual fund ownership of a soft information stock should forecast higher expected returns than for an average soft information stock. A similar argument could be made for centralized funds: high centralized fund ownership of a hard information stock should predict higher expected returns. However the inability that decentralized (centralized) funds have to gather and utilize hard (soft) information should result in a *disability* to invest in this type of stocks.

3 Dataset Construction and Methodology

3.1 Information Variables and score construction

As suggested by Petersen (2004), one can think of hard and soft information as a continuum of information "softness" rather than as two separate categories. Moreover, Petersen also points out that information can be harden. A good example of information hardening is credit scoring. Nowadays, financial institutions use credit scores to try to quantify the credit worthiness of loan applicants. Without these models, quantifying the credit worthiness of an applicant would be difficult. The score I construct aims at measuring the information "softness" of a firm and the extent to which information about a company has been hardened. The score is based on four variables. The first two variables, market capitalization (SIZE) and age (AGE), measure the information softness of a firm. These variables have been previously used in the banking industry and the main idea is that information available about older and larger firms tends to be harder than information generated by younger and smaller companies. The other two variables, number of analyst forecasts (NUM EST) and institutional ownership (OWN), measure the extent to which information about a company has been hardened. The basic premise is that these two variables measure the level of due diligence on a company. For instance, it is plausible to think that there is more hard information about a company followed by 50 analysts than by a company without analysts coverage.

Table 1 presents some summary statistics on the variables used to construct the information score (Size, Age, Number of Analyst Estimates and Institutional Ownership). It also contains summary statistics on stock return predictors book-to-market (B/M, its natural logarithm), firm-level volatility (VOL), and turnover (TURN). All statistics are calculated cross-sectionally each quarter and are then averaged across time. These statistics are calculated for stocks held by mutual funds. A few points a noteworthy. First, the strong positive correlation between all the information variables indicates that constructing a score based on the ranking of the raw values of these variables does not make the best use of this information. For instance, sorting stocks by NUM EST would be largely similar to sorting them by Log SIZE. Therefore this correlation structure between the variables will have to be taken into consideration when constructing the information score.

3.1.1 Information Score

The construction of the information score for each stock is based on the ranking of its information variables values. However, as noted before, if the raw values of the variables were used, the four rankings for each stock would not differ much. It is also clear that all the variables tend to be correlated with size. For instance, institutional ownership is highly correlated with size. This finding is consistent with other papers (Gompers and Metrick (2001), Nagel (2005)). The same is true for Number of Analyst estimates and Age. To try and purge the size effect from the other information variables, the construction of the information score employs the residuals of AGE NUM EST OWN as sorting variables. The residuals are obtained by regressing the variables on size. This orthogonalization of the information variables is similar to the one used in Hong et al (2000) and Nagel (2005). Since institutional ownership is bounded between 0 and 1, it is necessary to transform the variable so that it maps to the real line. I perform the following logit transformation,

$$Logit(OWN) = \log\left(\frac{OWN}{1-OWN}\right) \quad (1)$$

where the values below 0.0001 and above 0.9999 are replaced with 0.0001 and 0.9999 respectively. The information variable residuals are calculated by regressing the information variables on Log Size and squared Log size. On average, across all quarters, I find the following relations:

$$\textit{Logit(OWN)} = -7.31 + 1.68\ \textit{Log size} - 0.09\ \textit{Log size}^2 + \epsilon \quad (2)$$

$$\textit{NUM EST} = 2.13 - 1.90\ \textit{Log size} + 0.43\ \textit{Log size}^2 + \epsilon$$

$$\textit{Log AGE} = 4.44 - 0.25\ \textit{Log size} + 0.035\ \textit{Log size}^2 + \epsilon$$

Each quarter, I take the universe of NYSE stocks and rank them in 20 groups by size and residual age. I use the NYSE size and residual age rank cut-off points to rank stocks held by mutual funds. I also rank stocks held by mutual funds in 20 groups by residual institutional ownership and residual number of analysts estimates. For each stock, I then calculate an aggregate information variable by summing up the ranks of the four variables for each stock. For instance, if a stock belongs to size group 1, residual age group 2, residual institutional ownership group 2 and residual number of analyst estimates group 10, the aggregate information variable equals 15. Next, I rank stocks by this aggregate information variable each quarter in deciles. The information score will be equal to the aggregate information variable decile.

3.2 Mutual Fund Variables and Hierarchy Score

The main objective of the hierarchy score is to measure the organizational complexity of a fund. As mentioned earlier, there are two dimensions to consider, the actual centralization of tasks within the fund and the actual organizational complexity of the fund family to which the fund belongs. To measure the actual level of organizational complexity within the fund, I use is the number of managers that control the asset allocation (NUM MGRS) and net assets under management (AUM). As far as number of mangers, the premise is that funds with many managers will tend to be team-managed. This, in turn, causes managers to decide on an asset allocation based on consensus, which is the kind of separation between research and decision making process cited in Stein (2002). Regarding net assets under management, the main idea is that larger organizations are more hierarchical, since large organizations tend to centralize activities. This variable has also been used in Chen et al (2004) to measure the organizational complexity of mutual funds. The third and fourth variables I use are the number of funds (NUM FUNDS) and the total assets under management in actively managed US equity funds (FAM SIZE) of the family a fund belongs to. These variables are motivated by papers in fund cross-subsidization and fund proliferation (Massa (2003) and Gaspar Massa Matos (2006)). The idea is that maximizing fee income at the family level is different from maximizing fee income at the fund level. This will lead families to cross-subsidize funds that are the most likely to benefit from the convex relationship between past performance and current net flows documented in Chevalier and Ellison (1997) and Sirri and Tufano (1998). Families will also try to enhance the performance of funds that maximize the positive spill over effect a top-performing fund has on the its sibling fund net flows (Nanda et al (2004)). Therefore fund families have incentives to cross-subsidize fund returns in other to maximize their own income fee. As such, it is reasonable to believe that the asset allocation of a fund will be influenced by its family. This creates the separation between research and fund allocation mentioned in Stein (2002).

Table 2 presents some summary statistics on the variables used to construct the hierarchy score and on other mutual fund variables of interest. We can see that the number of managers is not highly correlated with any other hierarchy variable. However, NUM FUNDS, AUM and MGMT SIZE are highly correlated. For instance, sorting funds on AUM would produce similar results to sorting funds on MGMT SIZE.

3.2.1 Hierarchy Score

As indicated above, the variables Number of Funds, Family Size and AUM are highly and positively correlated. Therefore a hierarchy score based on a function of the individual rankings based on the raw values of the hierarchy variable would not provide much extra information. For instance ranking funds by AUM would produce a very similar ranking if funds are ranked by Family size. I therefore, orthogonalize the number of funds with respect to AUM. I also orthogonalize Family Size with respect to AUM and Number of Funds. On average, across all quarters, I find the following relations:

$$LogNUMFUNDS = 1.64 + 0.074\ Log\ AUM + 0.013\ Log\ AUM^2 + \epsilon \qquad (3)$$

$$\begin{aligned} LogFAMSIZE = {} & 2.04 + 0.40\ Log\ AUM + 0.0004\ Log\ AUM^2 \\ & +1.81\ Log\ NUM\ FUNDS - 0.034 Log\ NUM\ Funds^2 + \epsilon \end{aligned}$$

Each quarter, I rank funds by AUM, residual number family funds, and residual family size in six groups. I calculate an aggregate hierarchy variable by summing up these rankings and the number of managers (which takes values from one to six). For instance if a fund belongs to AUM group 1, residual number of funds group 2, residual family size group 2 and it has 3 managers, the aggregate hierrachy variable equals 8. Each quarter, I rank funds by this aggregate hierarchy variable in deciles each quarter. The hierarchy score will be equal to its aggregate hierarchy variable decile.

3.3 Data

Data on stock returns and prices are from the Center for Research in Security Prices (CRSP) Monthly Stocks File for NYSE, Amex, and NASDAQ stocks. I eliminate closed-end funds, real estate investment trusts (REIT), American Depository Receipts (ADR), foreign companies, primes, and scores. I exclude stocks below the 20th NYSE size percentile from the tests that look at stocks returns due to the well-documented asset-pricing anomalies in small stocks (Griffin and Lemmon (2002)). Market capitalization is defined as the product between share price and shares outstanding. Age is defined as the number of months that a security is present in the CRSP Monthly File. Data on institutional holdings are obtained from the Thomson Financial Institutional Holdings (13F) database. I extract quarterly holdings starting in the first quarter of 1993 and ending in the last quarter of 2006. I calculate the share of institutional ownership by summing the stock holdings of all reporting institutions for each stock in each quarter. Stocks that are on CRSP, but without any reported institutional holdings, are assumed to have zero institutional ownership. Ownership greater than one are omitted as they could be a result of double-reporting by institutional investors. The number of analysts estimates is calculated using I/B/E/S. At the end of a company's fiscal year, I count the maximum number of one-year EPS estimates that were outstanding during the fiscal year in question. NUM EST is the maximum number of one-year EPS estimates in the most recent fiscal year.

As stock return predictors, I use book-to-market (B/M), firm-level volatility (VOL), and turnover (TURN). The book value of equity in the nominator of B/M is taken from the Compustat Database, and it is defined as in Daniel et al (1997). At the end of each quarter t, I calculate B/M as the book value of equity from the most recent fiscal year-end that precedes the previous June divided by the market value of equity at the end of quarter t. Consistent with Fama and French (1993), I exclude firms with negative book values.

Organizational characteristics of mutual funds are taken from the CRSP Survivor-Bias-Free US Mutual Fund Database. I identify all share classes issued by mutual funds using the MFLINKS provided by WRDS and calculate the mutual fund characteristics at the funds level, not at the share class level. The sample starts in the first quarter of 1993 and ends in the last quarter of 2006. I calculate the number of managers by counting the different names in database manager field. The funds in my sample have a maximum of 5 names in the database manager filed. However, for some funds, the manager field is set to "Team Managed". If this is the case, I set the variable "number of managers" equal to six to indicate that team managed funds are the most hierarchical ones in the cross-section. Assets under management equals the total TNA of the fund's share classes. The number of actively managed US equity funds per family is calculated at the end of each quarter. The family size variable is the sum of all actively managed US equity funds TNA offered by a family.

Data on Mutual Fund Holdings is obtained from the Thompson Financial Mutual Fund Database. I also eliminate funds that, on average, have not invested at least 70% of their holdings in or that have less than 20 holdings from the CRSP universe defined above. This leaves funds that mainly invest in US stocks. Furthermore, I eliminate portfolio holdings with extreme portfolio weights (greater that 70%) since they are likely to be errors. I obtain pricing information for holdings from CRSP Monthly Stocks File.

I merge the holdings data and the mutual fund organization data for funds that report on quarter-end dates. I use the MFLINK tables provided by WRDS. This filters out some non-US actively managed equity funds. Additionally, I screen out index funds by looking for the word index and abbreviations in the CRSP fund name variable. I eliminate index funds, since passive investment is not based on information but on minimizing tracking error with respect to a benchmark. I also check that the MFLINK matches between the unique identifiers in the CRSP Mutual Fund Dataset and the unique identifiers in the Thomson Financial Mutual Fund Holdings database correspond to funds managed by the same management company. This eliminates erroneous MFLINK matches. As an additional check on the accuracy of the MFLINK matches, I eliminate funds for which the TNA reported in the TFN database is not between 1/1.3 and 1.3 of the TNA reported in the CRSP database.

4 Results

4.1 Determinants of information score

My first hypothesis indicates that centralized mutual funds will tilt their portfolios towards hard information companies. In order to determine whether centralized funds tend to invest in the higher spectrum of the information score, I calculate value weighted average of the information score of all holdings in a fund's portfolio.

Table 3 Panel A regresses the value weighted average information score contemporaneously on hierarchy score. From the regressions in Panel A, we can see that the mean of the average information score of a fund is above 6. This means that funds, regardless of their hierarchy structure, tend to own companies in the higher side of the information score (the score goes from one to ten and has a median of five). According to Stein's model, this is to be expected as any mutual fund is an organization and as such it has some degree of hierarchy complexity. Model 1 shows that the hierarchy score positively covaries with the information score as predicted by the theory.

It is important to note that the average information score may be determined by other variables. For example, it is sensible to expect that the investment style followed by a fund will influence the type of companies held in its portfolio. Aggressive growth funds tend to hold companies that are younger and smaller. On the other hand, income funds may be interested in companies that generate high dividends. These companies tend to be old and large. To control for style, I create dummies for the self-reported investment style followed by each fund. More specifically, I create dummies for the investment style GROWTH, and GROWTH AND INCOME / BALANCED. The other investment style is AGGRESSIVE GROWTH which is used as the base style in the regressions. Furthermore, net flows can also have an impact in the information score. If a fund experiences high net flows, it will have to allocate the new funds rather quickly as their investment styles may not allow the fund to hold cash reserves above a certain threshold. As pointed out in Pollet and Wilson (2006), in the face of growth, funds tend to scale up their original holdings. The pace with which funds add (or subtract) new holdings seems to be rather slow. Therefore, net flows should affect portfolio holdings of existing positions and thus the value weighted average information score. Models 2 to 4 control for these variables. The first interesting fact is that the effect of HIERARCHY remains virtually unchanged. One can see that the average hierarchy score positively covaries with the average information score after controlling for the other variables of interest. Moreover, the more growth oriented a fund is, the lower the average information score. For instance, the average information score of an aggressive growth fund is 6.30 while the average information score for a growth and income / balanced fund is 7.68. It is also important to note that net flows have an effect on the average information score. For instance, other things being equal, a one standard deviation shock in the FLOW variable in Model 3 leads to a change of 0.20 in the average information score. Past net flows have a similar effect.

One of the potential problems of this specification is the difficulty to establish a causality relationship between hierarchy and weighted average information score of a fund. Hypothesis 1 indicates that the degree of centralization of a funds (hierarchy score) should have an effect on the type of stocks held in its portfolio (i.e. soft vs hard information stocks). However the tests above do not rule out that the effect is the other way around, i.e. the average information "softness" of stocks held in a mutual fund portfolio has an effect on a fund's organizational complexity. For instance, one could argue that funds decide to concentrate in a sub-set of stocks (i.e. soft information companies) and then decide on a hierarchy structure. In other to address these concerns, I regress the contemporaneous average information score on past hierarchy score and on control variables. The results are shown in Table 3, Panel B and C. Panel B lags the explanatory variables by 6 months while Panel C does it for a full year. The results corroborate my previous findings: centralized funds tend tilt their portfolios towards hard information stocks.

4.2 Relative performance of centralized and decentralized mutual funds

Hypothesis 2 states that decentralized (centralized) funds should better at investing in soft (hard) information stocks than centralized (decentralized) funds. In order to test this hypothesis, I construct a self-financing trading strategy that exploits the outperformance of centralized (decentralized) funds in the hard (soft) information universe of stocks. Stein's theory is constructed around the idea that there are soft and hard information only. My approach measures information "softness" as a continuum. In order to test the insights of Stein's model, I define soft information stocks those that belong to the bottom decile of the information score distribution (i.e. information score distribution = 1). Stocks that are in the top decile of the distribution (information score = 10) are labelled as "hard information stocks". I apply a similar strategy with the organizational complexity of funds. Funds in the bottom quintile of the hierarchy score (hierarchy score = 1) are called decentralized funds and funds in the top quintile of the distribution (hierarchy score = 5) are called centralized funds. From the cross-section of stocks held by mutual funds, I identify the set of soft information stocks held by centralized and decentralized mutual funds respectively. The first self-financing trading strategy takes a long position in an equally-weighted portfolio made of soft information companies held by decentralized funds. This purchase is financed by short selling an equally-weighted portfolio composed of soft information stocks held by centralized funds. The return of this self-financing trading strategy is measured every month and rebalanced every quarter. Panel A in Table 4 shows that the average monthly return of this trading strategy is 0.41 % or 5 % a year for the sample. I also regress the return time series on known risk factors. For instance, the 5-factor risk adjusted return is 0.33% a month (4% a year) and is significant at the 10% level.

Similarly, I implement a self-financing trading strategy that aims at exploiting the relative outperformance of centralized funds over decentralized funds in the group of hard information stocks. The trading strategy consists of buying an equally-weighted portfolio of the universe of hard information stocks held by centralized funds financed by short-selling an equally-weighted portfolio made of the universe of hard information companies held by decentralized funds. As before, I keep track of the monthly returns of the self-financing trading strategy, and I rebalance it every month. Panel B shows the results for the second trading strategy. As one can see, the strategy delivers positive returns albeit economically small (0.36% a year). One can infer from these results that decentralized funds tend to be better at picking soft information companies. Centralized funds are better at picking hard information companies but the relative outperformance is not very strong.

The degree by which a funds load on to a holding should indicate the fund managers beliefs about the holding's future performance. For example if a manager strongly believes that IBM stock returns will be high, the manager will overweight this security is its portfolio selection. Likewise, if a manager is not very confident about the future performance of a long position, it is likely that the manager will not overweight it. Therefore, a more direct test of relative performance of centralized and decentralized funds is to incorporate this information in the trading strategy. Therefore, I rebalance the portfolio weights of the trading strategies' to reflect the aggregate beliefs of managers in each organizational structure. The new weights are based on the average "investment intensity" that each organizational structure assigns to each stock. I define the "investment intensity" of a fund on a holding, as the difference between the portfolio weight of that holding and the fund's average portfolio weight. A large and positive difference indicates that the fund is very confident about the positive future performance of the holding. In contrast, a large and negative difference indicates that the fund's management team is not as bullish on a position. I calculate the average "investment intensity" for all the soft and hard information stocks held by decentralized and centralized funds respectively. I then normalize the average "investment intensity" of all the holdings in each of the trading strategies' long and short portfolios. I used the normalized average "investment intensity" as the new portfolio weights. For instance, if all decentralized funds invest only in two soft information stocks, namely stock A and B, and their average "investment intensity" across all decentralized funds are 1.5 and 0.5 respectively, the portfolio of soft information companies held by decentralized funds will have portfolio weights of 75% and 25% in A and B respectively. Furthermore, assume that all centralized funds only invest in stocks A and B as well and that the average "investment intensities" are 1 and 2 respectively. Therefore, the portfolio of soft information companies held by centralized funds has weights of 0.33% in stock A and 0.67% in stock B. In this example, the first self-financing trading strategy buys a portfolio of soft information companies held by decentralized funds (with portfolio weigths of 75% in A and 25% in B) and funds this investment by short-selling a portfolio of soft information stocks held by centralized funds (with portfolio weigths of 33% in A and 67% in B).

Panel A of Table 5 presents the return for the soft information stock trading strategy. This trading strategy achieves an average monthly return of 0.55% or 6.80% a year. This average return is higher than that of the equal-weighted trading strategy presented in Table 4. Moreover, the statistical significance of the results are stronger. For instance the 5-factor risk adjusted average monthly return is 0.51% or 6.30% a year and the result is significant at the one percent level. This reinforces my previous finding: Decentralized funds tend to be better than centralized funds at investing in soft information companies. On the other hand, the second trading strategy shows that centralized funds are not fundamentally better than decentralized funds at investing in hard information stocks. The average return of this trading strategy is 0.01% a month and it is not statistically significant different from zero. One possible explanation to this fact is that it is very difficult to generate abnormal returns when using hard information. The reason lies in the nature of this information. For instance, if a particular piece of information is hardened by an analyst at a research firm, this (hard) information will be shared with many fund managers who, in turn, will act upon it, thus eliminating any mispricings very quickly. Soft information is more likely to remain private as it can not be credibly transmitted.

4.3 Fund-by-fund trading strategy

One of the objections to the trading strategies implemented above could be that they do not control for investment style. For instance, if decentralized funds tend to be growth funds and growth companies tend to be soft information companies, then one could argue that the results above partially show that growth funds tend to be better at picking growth stocks. To address this concern, I implement a strategy based on holdings at the fund level. For decentralized and centralized funds (hierarchy score equal to one and five respectively), I re-calculate a fund-specific information score for the stocks held by each mutual fund (I exclude stocks below the 20th NYSE size percentile). I form a trading strategy for each fund by going long an equally-weighted portfolio composed of its hard information companies (fund-specific information score equal to ten) and short a portfolio made of its soft information companies (fund-specific information score equal to one). Every month, I aggregate the trading strategies by averaging the returns for the universe of decentralized and centralized funds respectively. For every month, I calculate the difference between the average return of the centralized fund trading strategies and the average return of the decentralized fund trading strategies. I rebalance the fund-specific strategies at the end of the quarter. In other to rule out funds that only invest hard or soft information companies (top and bottom deciles of the cross-sectional information score), I exclude funds that do not hold at least $\alpha\%$ of companies that are below or above the median of the cross-sectional information score.

This implementation controls for investment style since the trading strategy is based on the holdings of individual funds as opposed to the holdings of all funds of a particular organizational structure (i.e. goes long growth stocks, shorts growth stocks). Since the trading strategy buys portfolios of hard information companies and short-sells portfolios of soft information companies, the difference between the average return of the centralized fund trading strategies and the average return of the decentralized fund trading strategies should be positive since the average return of the centralized fund trading strategies should be positive and that of the decentralized funds should be negative. The results in Table 6 confirms this conjecture. Each panel in Table 6 presents the results for different levels of alpha. We can see that the difference of these averages goes from 0.22 % a month (2.27% year) to 0.61% a month (7.57% per year). Risk adjusting these differences do not change the results much as the strategies do not load significantly on known risk factors. Therefore, controlling for investment style still reveals evidence of superior investment skill of some organizational structures in stocks with different levels of information "softness".

4.4 Cross-sectional regressions

Hypothesis 3 conjectures that decentralized funds are able to extract informational rents from soft information companies. In other words, since these funds are able to collect and incorporate soft information in their asset allocation process, their superior information should be reflected in the outperformance of their soft information holdings. Likewise, centralized funds should be able to use their superior information to pick hard information companies with high future returns. Hence, the aggregate level of ownership by different organizational structures should tell us something about the future performance of soft and hard information stocks. For instance, if the majority of decentralized funds overweights a particular soft information stock, it should be because the managers of these funds obtained positive (soft information) signals about the future performance of the stock. Therefore, I argue that the average "investment intensity" of decentralized funds should forecast the returns of soft information companies (High average intensity of investment of decentralized funds should forecast high future expected returns for soft information stocks). The same argument works for centralized funds and hard information companies: high average intensity of investment of centralised funds should predict high future expected returns for hard information stocks.

To explore whether ownership of soft and hard information stocks by centralized and decentralized funds predicts the cross-section of stock returns, Table 7 presents a series of cross-sectional regressions of returns of stocks held by mutual funds (I exclude stocks below the 20th NYSE size percentile). Cross-sectional regressions are run every month from April 1993 to March 2007. Dependent variable is the return from month t to t+1, which is regressed on month t stock characteristics. The stock predictors I employ are book-to-market ratio (B/M), the total individual stock return over the previous 12 months (RET12), the monthly trading volume scaled by the number of shares outstanding averaged across the previous three months (TURN), the standard deviation of monthly individual stock returns over the previous 12 months. To test my hypothesis, I construct two variables that measure the average intensity of investment of decentralized funds (DEC) and of centralized funds (CEN). I also include dummies for soft and hard information stocks. Because some of the predictors do not have well-behaved distributions I use their natural logarithm (DEC, CEN, B/M, TURN).

In the first column of Table 7 (Model 1), future returns are regressed on four predictors (B/M, RET12, TURN, and VOL). The results are consistent with previous results in the literature. It is important to remember that return predictability is stronger at long horizons. Therefore, it is not surprising to see some well-known predictors with the right signs, but statistically insignificant (VOL and TURN). We can see that B/M (Value indicator) and RET12 (Momentum indicator) predict future returns with important levels of statistical significance. Model 2, controls for the average return of hard and soft information companies by adding dummies for hard and soft information stocks. We can see that on average hard information stocks (information score equal to 10) do not have a very different average return from other stocks. Soft information stocks tend to have lower returns (15 bps per month) than other companies. However this difference is not statistically significant. Model

3 directly test my hypothesis above. I do this by interacting the hard and soft information dummies with the average intensity of investment of centralized funds (CEN) and decentralized funds (DEC), where decentralized and centralized funds are defined as before. The idea is to see the marginal effect of the average intensity of investment of different organization designs (centralized and decentralized funds) on stocks of different information softness (hard and soft information stocks)

Model 3 tells us that future expected returns of a soft information stocks are higher when the intensity of investment of decentralized funds are higher and lower when the intensity of investment of centralized funds are higher. More specifically, the coefficient on HARD X DEC in Model 3 implies that ownership of hard information stocks by decentralized funds forecast an additional -0.46% a month (- 5.6% per year) for every unit of DEC. However the ownership of soft information stocks by decentralized funds (SOFT X DEC) forecast an additional 0.48 % per month (5.9% per year) for every unit of DEC. This corroborates my initial conjecture, decentralized funds are able to exploit their organizational advantages to process soft information to pick good soft information stocks. Moreover, decentralized funds display a strong disability at investing in hard information companies.

Model 3 also shows that future expected returns of hard information companies are lower when the average intensity of investment of decentralized (DEC) and centralized (CEN) funds are higher. The coefficient on HARD X CEN shows that ownership of hard information stocks by centralized funds tend to forecasts an additional return of -0.30 % per month (-3.65% a year) for each unit of CEN. On the other hand, ownership of soft information stocks by centralized funds forecasts and additional return of -0.59 % per month (-7.3% per year) for each unit of CEN. This shows that centralized funds display an important disability at investing in soft information companies. This is due to the organizational diseconomies described before: centralized organizations can not incorporate soft information in the decision making process. Moreover, centralized funds show no particular ability at investing in hard information companies. This also points out to the fact that it is difficult to earn informational rents when information can not be kept private. In other words, centralized funds can not earn abnormal returns on their hard information stock investments since it is difficult to contain their "superior" (hard) information on investment prospects.

5 Conclusion

I tests the predictions of Stein (2002) in a sample of actively managed US equity funds. I also develop scores that measure the information "softness" of stocks and the organizational complexity of mutual funds. I find that the level of organizational complexity of a fund positively covaries with its average information "softness" of its holdings. I also document that decentralized funds are better than centralized funds at investing in soft information stocks. The ability of centralized fund to outperform decentralized fund in the hard information universe of stocks is less pronounced. Decentralized funds also seem able to pick soft information stocks with high expected returns and show an important disability when investing in hard information stocks. Centralized funds show no particular ability when investing in hard information stocks and show a large and important disability when investing in soft information stocks.

The first set of results indicates that hierarchical funds tilt their portfolios towards hard information stocks. This confirms Stein's insight in that, since the only information that can be transmitted is hard information, hierarchical funds rely more on it and therefore tilt their portfolios to stocks for which most the information available is hard. This relationship between organization and information helps explain the increase in demand for large stocks (usually hard information companies) documented in Gompers and Metrick (2001) and the consequent reversal of the small stock risk premia over the last 30 years. The surge of institutional investors and the growth of the delegated asset management industry, has given rise to complex hierarchical (centralized) organizations as investment vehicles. As explained before, hierarchies tend to be heavy users of hard information which is precisely the kind of information produced by large companies. The rise of these hierarchies as investment vehicles may also help explain the replacement of soft information with hard information as the basis for financial transactions documented in Petersen (2004).

I document that the soft information holdings of decentralized funds perform better than those of centralized funds. Similarly, the hard information holdings of centralized funds outperform those of decentralized funds. However the out performance is not as strong and pronounced. The last set of results show how decentralized funds have a special ability to choose soft information companies while centralized funds show a disability in this respect. Once again, this confirms the inferences made from Stein model in that the collection of soft information in a hierarchy would not be utilized. If this information can not be used in the investment decision process, it is expected to see that funds that do not use this type of information do badly when investing in soft information companies. As far as hard information companies, neither centralized nor decentralized shows an special ability at investing in these stocks. However, centralized funds seem to do better than decentralized funds. Decentralized funds show a much greater disability when it comes to invest in hard information companies. This last fact still shows that Stein model provides a good prediction for hard information stock investment by different hierarchical structures. The fact that centralized funds show no special ability when investing in hard information stocks reflects the difficulty of earning abnormal earnings when information can be easily transmittable.

One last important point is noteworthy. My results help explain why concentrated funds tend to outperform diversified ones [6]. As we have seen, decentralized

[6] Kacperczyk Sialm and Zheng (2005) argue that concentrated managers outperform diversified ones and that the effect is more pronounced amongst managers that hold portfolios concentrated in

funds tilt their portfolios towards soft information stocks. Since soft information is not transferable, collectors of this information have a longer first mover advantage relative to collectors of more transferable information (i.e. hard information). In Van Nieurwerburgh and Veldkamp (2008), investors who can first collect information systematically deviate from holding a diversified portfolio. Therefore, it is optimal for collectors of soft information companies to deviate from holding a diversified portfolio. Instead, they choose to learn extensively about fewer stocks in hope of collecting informational rents in the future. From Table 2 we can see how the average portfolio weight is negatively correlated with the hierarchy score. This indicates that centralized funds tend to be more diversified than decentralized ones. We have also shown that decentralized funds are able to forecast returns of soft information companies and that they do a much better job in soft-information company investment than centralized funds. It is therefore likely that part of the positive relationship between outperformance and portfolio concentration can be explained by the organizational form of funds that choose to hold diversified or concentrated funds. This is a very interesting direction for future research.

few industries. Van Nieuwerburgh and Veldkamp (2008) derive conditions under which deviating and holding a concentrated portfolio is an optimal strategy. Bask Busse and Green (2006) discuss mutual fund performance and managers' willingness to take big bets in a relatively small number of stocks. They document that concentrated managers tend to outperform their diversified counterparts

References

Allen N. Berger, Gregory F. Udell, 2006, A more complete conceptual framework for SME finance, *Journal of Banking and Finance* 30, 2945–2966.

Baks, Klaas P., Jeffrey A. Busse, and T. Clifton Green, 2006, Fund managers who take big bets: Skilled or overconfident, Working paper, Emory University.

Berger, Allen N., Richard J. Rosen, and Gregory F. Udell, 2006, Does Market Size Structure Affect Competition? The Case of Small Business Lending, *Journal of Banking and Finance*, 31, 11–33.

Campbell, Tim and William Kracaw, 1980, Information Production, Market Signaling and the Theory of Financial Intermediation, *Journal of Finance 35*, 863–881.

Chen, Joseph, Harrison Hong, Ming Huang, and Jeffrey Kubik, 2004, Does fund size erode mutual fund performance? The role of liquidity and organization, *American Economic Review* 94, 1276–1302.

Chevalier, Judith, and Glenn Ellison, 1997, Risk taking by mutual funds as a response to incentives, *Journal of Political Economy* 105, 1167–1200.

Cici, Gjergji , George S. Gibson and , Rabih Moussawi, 2006, For Better or Worse? Mutual Funds in Side-By-Side Management Relationships With Hedge Funds, Working Paper, Mason School of Business.

Daniel, Kent, Mark Grinblatt, Sheridan Titman, and Russ Wermers, 1997, Measuring mutual fund performance with characteristic-based benchmarks, *Journal of Finance* 52, 1035–1058.

Diamond, Douglas, 1984, Financial Intermediation and Delegated Monitoring, *Review of Economic Studies* 51, 393-414.

Diamond, Douglas, 1991, Monitoring and Reputation: The Choice between Bank Loans and Directly Placed Debt, *Journal of Political Economy* 99, 689–721.

Fama, Eugene F. and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.

Gaspar, Jose-Miguel, Massimo Massa, and Pedro Matos, 2006, Favoritism in mutual fund families? Evidence of strategic cross-fund subsidization, *Journal of Finance* 61, 73–104.

Gompers, P.A., Metrick, A., 2001. Institutional investors and equity prices. *Quarterly Journal of Economics* 166, 229–260.

Griffin, M.J., Lemmon, M.L., 2002. Book-to-market equity, distress risk, and stock returns. *Journal of Finance* 57, 2317–2336.

Haubrich, Joseph, 1989, Financial Intermediation, Delegated Monitoring, and Long-Term Relationships, *Journal of Banking and Finance* 13, 9-20.

Hong, H., T. Lim, and J. Stein, 2000, Bad news travels slowly: Size, analyst coverage and the profitability of momentum strategies, *Journal of Finance* 55, 265–295.

Kacperczyk, Marcin, Clemens Sialm, and Lu Zheng, 2005, On the industry concentration of actively managed equity mutual funds, *Journal of Finance* 60, 1983–2012.

Leland, H.E., Pyle, D.H., 1977. Informational asymmetries, financial structure, and financial intermediation. *Journal of Finance* 32, 371–387.

Massa, Massimo, 2003, How do family strategies affect fund performance? When performance maximization is not the only game in town, *Journal of Financial Economics* 67, 249–304.

Nagel, Stefan, 2005, Short sales, institutional investors, and the cross-section of stock returns, *Journal of Financial Economics* 78, 277–309

Nanda, Vikram, JayWang, and Lu Zheng, 2003, Family values and the star phenomenon, Working paper, University of Michigan.

Petersen, Mitchell. 2004. Information: Hard and Soft, Working Paper, Northwestern University.

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

Stein, Jeremy, 2002, Information Production and Capital Allocation: Decentralized versus Hierarchical Firms, *The Journal of Finance* 57, 1891–1921.

Stiglitz, Joseph, and Andrew Weiss, 1981, Credit rationing in markets with imperfect information, *American Economic Review* 71, 393–410.

Vannieuwerburgh, Stijn and Veldkamp, Laura, 2008, Information Acquisition and Under-Diversification, Working paper, New York University.

Table 1: **Summary Statistics: Information Variables**

Panel A reports time-series averages of equal-weighted quarterly cross-sectional means and standard deviations of the information variables and the return predictors Log B/M is the natural logarithm of the book-to-market ratio; Log AGE is the natural logarithm of the number of months a stock has been reported in CRSP; NUM EST is the number of analyst one-year EPS estimates outstanding for a stock; RET12 (momentum) is the total individual stock return over the previous 12 months; Log SZ is the log of market capitalization; TURN(turnover) is the monthly trading volume scaled by shares outstanding, averaged over the previous three months and divided by two for Nasdaq stocks; VOL (volatility) is the standard deviation of monthly individual stock returns over the previous 12 months;. Panel B reports time-series averages of cross-sectional correlations. The sample period runs from the first quarter of 1993 to the last quarter of 2006. Only stocks held by mutual funds are included.

	OWN	Log B/M	Log AGE	NUM EST	RET12	Log SZ	TURN	VOL
Panel A: Means and standard deviations								
Mean	0.39	-0.71	4.24	6.08	0.18	5.45	0.07	0.15
Standard Deviation	0.26	0.92	1.04	7.69	0.55	1.90	0.08	0.10
Observations per quarter (average)	5092	4203	5150	5150	5091	5124	5121	5065
Panel B: Contemporaneous correlations								
OWN		-0.16	0.19	0.51	0.06	0.62	0.31	-0.22
Log B/M			0.05	-0.18	-0.41	-0.37	-0.27	-0.12
Log AGE				0.26	0.03	0.21	-0.10	-0.24
NUM EST					-0.03	0.75	0.28	-0.18
RET12						0.15	0.20	0.31
SIZE							0.27	-0.30
TURN								0.31

Table 2: **Summary Statistics: Hierarchy Variables**

Panel A reports time-series averages of equal-weighted quarterly cross-sectional means and standard deviations of the mutual fund variables. Log AUM is the natural logarithm of the total assets under management; NUM MGRS is the total number of managers in charge of the fund; Log FAM SIZE is the log of the total assets under management of all actively managed US equity sibling funds; Lof NUM FUNDS is the number of actively managed US equity sibling funds; Log FLOW is the log flow of new funds and is defined as the difference between log growth rate for the AUM and the log return; Past Log FLOW is the Log Flow in the previous quarter; W is the average portfolio weight. Panel B reports time-series averages of cross-sectional correlations; HIERARCHY is the hierarchy score; INFO is the information score. The sample period runs from the first quarter of 1993 to the last quarter of 2006. Only actively managed US equity mutual funds are included.

	Log AUM	NUM MGRS	Log FAM SIZE	Log NUM FUNDS	Log FLOW	Past Log FLOW	INFO	HIERARCHY	W
Panel A : Means and standard deviations									
Mean	4.95	2.17	7.65	2.11	0.04	0.05	6.70	2.90	0.02
Standard Deviation	1.95	1.58	2.33	1.12	0.25	0.29	1.27	1.40	0.01
Observation per quarter (average)	780	694	694	695	773	761	781	694	781
Panel B : Contemporaneous correlations									
Log AUM		0.04	0.62	0.33	-0.08	-0.07	0.07	0.50	-0.34
NUM MGRS			0.05	0.05	0.00	0.01	0.03	0.48	-0.12
Log FAM SIZE				0.84	0.01	0.01	0.02	0.74	-0.34
Log NUM FUNDS					0.00	0.00	0.04	0.51	-0.30
Log FLOW						0.25	-0.07	0.02	-0.01
Past Log FLOW							-0.08	0.02	-0.01
INFO								0.00	0.02
HIERARCHY									-0.28

Table 3: **Determinants of Information Score**

Dependent variable is the weighted average information score for each mutual fund portfolio in the sample at a quarter -end date. HIERARCHY is the hierarchy score for each mutual fund in my sample at each quarter-end date. GROWTH is a dummy variable for mutual funds that self-report their investment mandate as growth. GR AND INC is a dummy variable for mutual funds that self-report their investment style as growth and income or as balanced. FLOW is the log flow of new funds into a fund and is defined as the difference between the log growth rate for TNA and the log return for the fund in the current quarter. TNA is the fund's total net assets under management. PAST FLOW is the FLOW of the past quarter. Average R^2 is the average R^2 of all Fama-MacBeth regressions. t-statistics are in parentheses.

	Model			
Panel A: Contemporaneous	1	2	3	4
INTERCEPT	7.03	6.30	6.32	6.32
HIERARCHY	0.03	0.03	0.04	0.04
	(3.90)	(4.45)	(4.94)	(5.22)
GROWTH		0.64	0.64	0.65
		(8.44)	(8.56)	(8.60)
GR AND IN		1.38	1.36	1.35
		(13.18)	(13.29)	(13.25)
FLOW			-0.72	-0.43
			-(6.01)	-(3.27)
PAST FLOW				-0.57
				-(4.31)
Average R^2	0.39%	10.55%	11.89%	13.10%
Panel B: 6-Month Lag				
INTERCEPT	7.00	6.31	6.32	6.33
HIERARCHY	0.04	0.04	0.05	0.05
	(4.86)	(5.07)	(5.28)	(5.43)
GROWTH		0.59	0.60	0.60
		(7.76)	(7.83)	(7.86)
GR AND IN		1.32	1.31	1.30
		(12.55)	(12.63)	(12.62)
FLOW			-0.69	-0.39
			-(4.71)	-(2.41)
PAST FLOW				-0.54
				-(3.98)
Average R^2	0.57%	10.98%	12.22%	13.14%
Panel C: 12-Month Lag				
INTERCEPT	7.02	6.38	6.38	6.39
HIERARCHY	0.05	0.05	0.05	0.05
	(6.21)	(6.53)	(6.54)	(6.63)
GROWTH		0.53	0.54	0.54
		(7.56)	(7.69)	(7.70)
GR AND IN		1.24	1.22	1.21
		(12.49)	(12.62)	(12.65)
FLOW			-0.60	-0.35
			-(4.20)	-(1.98)
PAST FLOW				-0.53
				-(3.34)
Average R^2	0.56%	10.52%	11.67%	12.66%

Table 4: **Self-financing trading strategy: Equally-Weighted Monthly Regressions**

Dependent variable is the self-financing trading strategy return each month. Each quarter-end date I identify soft (hard) information companies held by centralised and decentralised mutual funds. I take long positions in soft (hard) information companies held by decentralised (centralised) funds and short positions in soft (hard) information stocks held by centralised (decentralised) funds. The long and short portions of these trading strategies are equally-weighted and are rebalanced every quarter. The three Fama–French factors are zero-investment portfolios representing the excess return of the market, Rm-Rf; the difference between a portfolio of "small" stocks and "big" stocks, SMB; and the difference between a portfolio of "high" book-to-market stocks and "low" book-to-market stocks, HML. The fourth factor, UMD, is the difference between a portfolio of stocks with high past one-year returns minus a portfolio of stocks with low past one-year returns. The fifth factor, LIQ, is the innovations in the aggreagate level of liquidity in Pastor and Stambaugh (2003). N denotes the number of monthly observations, R^2 indicates the regression Adjusted R square and t-statistics are in parentheses.

Mean	Intercept	Rm-Rf	SMB	HML	UMD	LIQ	Adj R^2 / N
Panel A: Soft Information Stocks Self-financing trading strategy							
0.41%							
(2.35)							
	0.37%	0.066					0.89%
	(2.10)	(1.57)					165
	0.33%	0.040	0.193	0.026			9.49%
	(1.91)	(0.83)	(3.87)	(0.41)			165
	0.33%	0.043	0.191	0.028	0.009		8.96%
	(1.82)	(0.87)	(3.76)	(0.43)	(0.25)		165
	0.33%	0.041	0.191	0.027	0.009	0.004	8.40%
	(1.81)	(0.79)	(3.75)	(0.41)	(0.25)	(0.12)	165
Panel B: Hard Information Stocks Self-financing trading strategy							
0.03%							
(1.55)							
	0.03%	-0.001					-0.60%
	(1.55)	-(0.12)					165
	0.02%	0.003	0.016	0.015			3.90%
	(0.88)	(0.60)	(2.97)	(2.20)			165
	0.02%	0.002	0.016	0.014	-0.003		3.71%
	(1.02)	(0.36)	(3.06)	(2.11)	-(0.83)		165
	0.02%	0.001	0.016	0.013	-0.003	0.003	3.54%
	(1.05)	(0.10)	(3.04)	(1.98)	-(0.80)	(0.85)	165

Table 5: **Self-financing trading strategy: Average Abnormal Portfolio Tilts Monthly Regressions**

Dependent variable is the self-financing trading strategy return each month. Each quarter-end date I identify soft (hard) information companies held by centralized and decentralized mutual funds and their portfolio tilts in each mutual fund portfolio. A Mutual portfolio tilt is defined as the portfolio weight of a holding divided by the mean portfolio weight. I take long positions in soft (hard) information companies held by decentralized (centralized) weighted by their normalized average abnormal portfolio tilts and short positions in soft (hard) information stocks held by centralized (decentralized) funds weighted in the same fashion. The self-financing trading strategy is financed each quarter. The other variables are defined as in table **??** N denotes the number of observations, and R^2 indicates the regression Adjusted R square and t-statistics are in parentheses.

Mean	Intercept	Rm-Rf	SMB	HML	UMD	LIQ	Adj R^2 / N
Panel A: Soft Information Stocks Self-financing trading strategy							
0.55%							
(2.80)							
	0.49%	0.102					2.2%
	(2.47)	(2.18)					165
	0.45%	0.087	0.145	0.030			5.3%
	(2.24)	(1.59)	(2.53)	(0.41)			165
	0.51%	0.067	0.158	0.021	-0.056		5.8%
	(2.48)	(1.18)	(2.74)	(0.28)	-(1.39)		165
	0.51%	0.062	0.158	0.018	-0.055	0.011	5.3%
	(2.48)	(1.04)	(2.72)	(0.24)	-(1.38)	(0.28)	165
Panel B: Hard Information Stocks Self-financing trading strategy							
0.01%							
(0.41)							
	0.01%	0.004					-0.4%
	(0.31)	(0.63)					165
	0.05%	-0.012	-0.036	-0.052			16.0%
	(1.80)	-(1.63)	-(4.60)	-(5.28)			165
	0.03%	-0.006	-0.040	-0.049	0.018		20.9%
	(1.17)	-(0.76)	-(5.24)	-(5.12)	(3.34)		165
	0.03%	-0.007	-0.040	-0.050	0.018	0.003	20.6%
	(1.18)	-(0.90)	-(5.24)	-(5.14)	(3.35)	(0.58)	165

Table 6: **Fund-by-fund Self-financing trading strategy**

Each quarter I identify funds that belong to the top and bottom quintile of the hierarchy score (centralized and decentralized funds) that hold at least α% of their assets in stocks that are above or below the information score. For each fund, the information score is recalculated using the fund's holdings only. The self financing trading strategy is constructed by going long hard information companies (information score = 10) and short soft information companies (information score = 1) for each fund at the end of each quarter. I aggregate the return for the decentralized and centralized fund strategies respectively. The dependent variable is the difference between the average return for the centralized fund trading strategies and the the average return for the centralized fund trading strategies. The strategies are held for a quarter and rebalanced at the end of each quarter.

Mean	Intercept	Rm-Rf	SMB	HML	UMD	LIQ	N / Adj R2
Panel A: $\alpha = 20\%$							
0.22%	0.24%	0.01	-0.02	-0.02	-0.03	0.03	165
(1.88)	(1.91)	(0.39)	-(0.56)	-(0.41)	-(1.19)	(1.23)	0.18%
Panel B: $\alpha = 25\%$							
0.28%	0.28%	0.03	0.02	0.00	-0.04	0.03	165
(1.95)	(1.82)	(0.71)	(0.46)	(0.00)	-(1.32)	(0.94)	0.17%
Panel C: $\alpha = 30\%$							
0.37%	0.41%	-0.02	0.07	0.00	-0.06	0.02	165
(2.10)	(2.15)	-(0.35)	(1.34)	(0.01)	-(1.61)	(0.55)	0.00%
Panel D:$\alpha = 35\%$							
0.38%	0.35%	-0.01	0.06	0.10	-0.04	0.06	165
(1.77)	(1.53)	-(0.22)	(0.97)	(1.17)	-(0.91)	(1.33)	0.00%
Panel E:$\alpha = 40\%$							
0.61%	0.58%	-0.08	0.13	0.05	-0.01	0.16	165
(2.00)	(1.80)	-(0.86)	(1.46)	(0.47)	-(0.15)	(2.62)	2.51%

Table 7: **Cross-sectional Regressions**

Cross-sectional regressions are run at the end of each month t from April 1993 to March 2007. The Dependent variable is the return over month t+1. B/M is the log of one plus the book value of equity. TURN (turnover) is the log of one plus monthly trading volume scaled by shares outstanding, averaged over the previous three months. VOL (volatility) is the standard deviation of monthly individual stock returns over the previous 12 months. RET12 (momentum) is the total individual stock returns over the previous 12 months. SOFT is a dummy for soft information companies. HARD is a dummy is a dummy for hard information companies. DEC is the average "abnormal" portfolio tilt of decentralized funds, where abnormal tilt of a mutual fund holding is defined as the holding's portfolio weight divided by the mean weight in the portfolio. CEN is the average "abnormal" portfolio tilt of decentralized funds.

	Model		
Predictor Variable	1	2	3
B/M	0.0037	0.0037	0.0036
	(3.29)	(3.21)	(3.15)
RET12	0.0057	0.0058	0.0059
	(1.79)	(1.81)	(1.83)
TURN	-0.0002	-0.0003	-0.0002
	-(0.12)	-(0.20)	-(0.16)
VOL	-0.0148	-0.0133	-0.0141
	-(0.68)	-(0.63)	-(0.67)
HARD		0.0004	0.0045
		(0.26)	(1.92)
SOFT		-0.0015	-0.0015
		-(1.27)	-(1.02)
HARD X DEC			-0.0046
			-(2.03)
SOFT X DEC			0.0048
			(2.11)
HARD X CEN			-0.0030
			-(1.29)
SOFT X CEN			-0.0059
			-(2.21)