Two Essays on Equity Mutual Funds

Puneet Kumar Jaiprakash

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
In
Business, Finance

Raman Kumar, Committee Chair
Gregory B. Kadlec
Arthur J. Keown
Yong Chen

August 5, 2011
Blacksburg, VA

Keywords: Fund Flows, Return Predictability, Macroeconomic Variables, Flow-performance Relationship, Semiparametric Regression, Conflicts of Interest

Two Essays on Equity Mutual Funds

Puneet Kumar Jaiprakash

ABSTRACT

This dissertation consists of two essays. The first essay studies the investment decisions of fund investors at the aggregate level. Previous research has shown that expected market returns vary over time and that this variation can be predicted by variables such as dividend yields and book-to-market ratios (Fama and French (1989); Campbell and Thompson (2008)). Further, macroeconomic variables affect asset returns (Flannery and Protopapadikas (2002)). We investigate whether the investment decisions of mutual fund investors incorporate information about future stock returns contained in predictive and macroeconomic variables. If investors incorporate this information, then variation in flows should be related to that in predictive variables and macroeconomic variables. Using quarterly flow data from 1951Q4 to 2007Q4, we find that both predictive and macroeconomic variables have a relatively small impact on flows. Our results suggest that fund investors, as a group, fail to adequately incorporate the information contained in these variables.

The second essay studies the investment decisions of fund investors in a cross-sectional setting. Existing literature documents that (i) an asymmetric flow-performance relationship creates an incentive for managers to extract rents from shareholders, and (ii) managers respond to such incentives by strategically altering portfolio risk. Using the semiparametric regression model proposed by Chevalier and Ellison (1997), we show that the flow-performance relationship has become linear in recent years (2000-2009) and fund managers no longer respond to such incentives. Fund managers, however, change portfolio risk in response to past performance; such changes have a positive impact on fund performance and are indicative of a better alignment of interests between managers and shareholders.

Table of Contents

List of Tables

List of Figures

Do Aggregate Fund Flows Incorporate Information contained in Predictive and Macroeconomic Variables?

Abstract

Previous research has shown that expected market returns vary over time and that this variation can be predicted by variables such as dividend yields and book-to-market ratios (Fama and French (1989); Campbell and Thompson (2008)). Further, macroeconomic variables affect asset returns (Flannery and Protopapadikas (2002)). We investigate whether the investment decisions of mutual fund investors incorporate information about future stock returns contained in predictive and macroeconomic variables. If investors incorporate this information, then variation in flows should be related to that in predictive variables and macroeconomic variables. Using quarterly flow data from 1951Q4 to 2007Q4, we find that both predictive and macroeconomic variables have a relatively small impact on flows. Our results suggest that fund investors, as a group, fail to adequately incorporate the information contained in these variables.

1. Introduction

The investment-consumption decisions of investors are of great importance to financial economists. In this paper, we study, at the aggregate level, the investment decisions of an important[1] segment of investors–open end mutual fund investors. In particular, we investigate whether fund investors account for time-varying expected market returns by incorporating information contained in predictive and macroeconomic variables.

To answer the above question, we draw upon results from three strands of literature: (i) *aggregate fund flow literature* which documents a positive contemporaneous relationship between market returns and unexpected flows[2] (Warther (1995)), (ii) *return predictability literature* (Fama and French (1989); Campbell and Thompson (2008)), which shows that variables (hereafter predictive variables) such as dividend yields and book-to-market ratios predict market returns, and (iii) *macroeconomic variables literature* which documents that market returns are related to innovations of macroeconomic variables such as industrial production and inflation (Chen, Roll, and Ross (1986); Flannery and Protopapadikas (2002)). Figure 1 shows the interaction among these variables.

If fund investors incorporate such information, then the variation in unexpected flows should be related to that in expected returns. This reasoning has two testable implications. First, unexpected flows should be related to the predictive variables, several

[1]At the end of 2008, open-end mutual funds held assets worth 9.6 trillion dollars (2009 ICI Fact Book, available at http://ici.org/pdf/2009_factbook.pdf) and 45% of the households in the U.S. held mutual funds (Bass, Sabelhaus, and Schrass (2009)).
[2]Expected flows are obtained from an AR model. Unexpected flows are the difference between actual flows and expected flows.

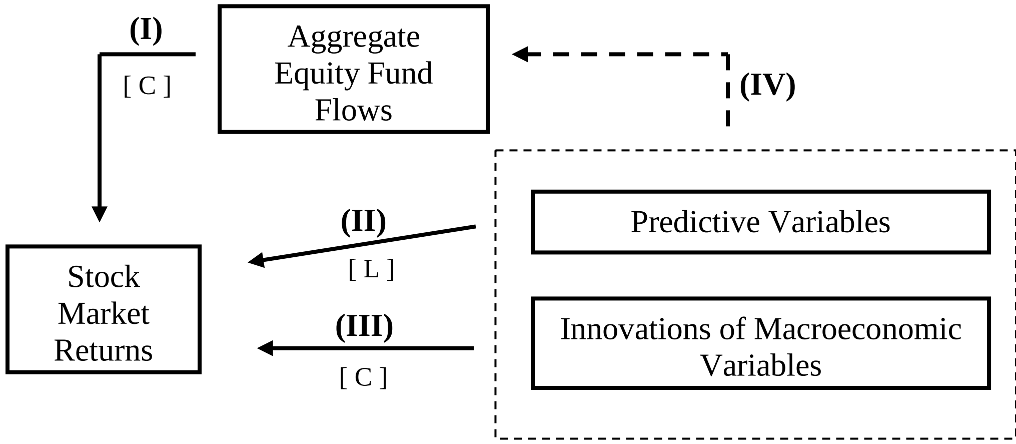

Figure 1: Interaction among Returns, Flows, Predictive and Macroeconomic Variables

The figure illustrates the hypothesis being tested in this paper. (I) indicates the positive contemporaneous relationship between stock market returns and aggregate equity mutual fund flows documented by Warther (1995). (II) refers to the return predictability literature documenting that variables such as dividend yields, book-to-market ratios, term spread, and default spread, *cay*, etc. predict returns (Fama and French (1989); Campbell and Thompson (2008)). (III) refers to the literature which shows that innovations of macroeconomic variables affect returns (Chen, Roll, and Ross (1986); Flannery and Protopapadikas (2002)). [C] indicates a contemporaneous relationship and [L] indicates a lagged relationship. If fund investors incorporate information contained in predictive & macroeconomic variables, then (IV) should exist; consequently, controlling for predictive and macroeconomic variables, (I) should weaken.

of which capture investors' expectations about future returns, and to factors such as macroeconomic variables that affect the investors' opportunity set. Second, if the observed return-unexpected flow relation is a manifestation of investment decisions that incorporate information contained in the predictive and macroeconomic variables, then controlling for these variables, the return-unexpected flow relationship should weaken.

For our empirical analysis, we use quarterly data on aggregate fund flows from 1951Q4 to 2007Q4. Since the existing fund literature uses monthly data for a shorter period (1984-95), we first confirm that the positive return-unexpected flow relation obtains in our longer sample period. Next, we examine the relation between unexpected flows, and predictive and macroeconomic variables. Among the predictive variables only *cay* predicts unexpected flows, and among the macroeconomic variables only money supply and labor income affect unexpected flows. In addition, the macroeconomic variables appear to not have any incremental power after controlling for predictive variables. The ability of predictive and macroeconomic variables to explain only a small variation in unexpected flows indicates the unsophisticated nature of fund investors, and is consistent with existing results (Goetzmann and Peles (1997); Alexander, Jones, and Nigro (1998); Wilcox (2003)) in the cross-sectional fund literature.

Finally, we examine the relation between returns, unexpected flows, predictive, and macroeconomic variables. In the presence of predictive and macroeconomic variables, the significance of the return-unexpected flow relation decreases only marginally. Moreover, the variation in returns explained together by unexpected flows, predictive and macroeconomic variables is marginally lower than the sum of (i) variation in returns explained by only unexpected flows and (ii) variation in returns explained together by predictive and macroeconomic variables. This suggests that the return-unexpected flow relationship is *largely independent* of the predictive and macroeconomic variables. We also test the hypothesis in a Vector Autoregression (VAR) framework by implementing Geweke's measure of conditional linear dependence and find that the results remain unchanged.

The rest of the paper is organized as follows. Section 2 briefly discusses the literature in each strand mentioned above. Section 3 discusses the construction of various fund flow, predictive and macroeconomic variables. Section 4 documents results from the existing literature. Section 5 contains new results on unexpected flows and predictive and/or macroeconomic variables, and tests of the above hypothesis. Section 6 examines the hypothesis in a VAR framework by using Geweke's measure of conditional linear dependence. Section 7 concludes.

2. Literature Review

The literature on aggregate fund flows examines the relation between aggregate flows and market returns and tests two hypotheses. First, given the stylized fact from the cross-sectional fund literature that investors chase past performance, do fund investors invest based on recent performance of the market (Feedback Trader Hypothesis)? Second, do flows affect prices contemporaneously? If fund investors possess information, then flows will reveal this information. Prices will *permanently* move in the direction of flows to incorporate this information thereby causing prices and flows to be positively correlated (Information Revelation Hypothesis). Alternatively, flows might exert *temporary* pressure on prices or flows might proxy for investor sentiment and thus affect prices positively (Price-Pressure/Investor Sentiment Hypothesis). Empirically, the feedback trader hypothesis predicts a positive relation between flows and lagged market returns. The information revelation and price-pressure/investor sentiment hypothesis both predict a positive relation between contemporaneous flows and returns; however, the information revelation hypothesis predicts no relation between returns and lagged flows since prices will incorporate information quickly whereas the price-pressure/investor sentiment hypothesis predicts a negative relation between returns and lagged flows since prices will reverse once the pressure/sentiment disappears.

Using monthly data from 1984-1992 for net sales of mutual funds, Warther (1995) documents a strong positive contemporaneous relation between return and unexpected flows, no relation between returns and lagged flows, and inconsistent with the feedback trader hypothesis, a negative relation between flows and lagged returns. Fant (1999) splits net sales into its constituents and estimates a VAR to show that Warther's anomalous finding is driven by the exchanges-out component[3] and conditioned on exchanges-in and exchanges-out, there is no relation (contemporaneous or lagged) between new sales and redemptions and returns.

Two other strands of literature have documented a relation between returns and other variables. The first is the return predictability literature. Starting with Keim and Stambaugh (1986), numerous papers have documented the predictive ability of aggregate variables such as dividend yields (Keim and Stambaugh (1986), Fama and French (1989)), book-to-market ratio (Kothari and Shaken (1997) and Pontiff and Schall (1998)), term spread (Fama and French (1989)), default spread (Fama and French (1989)), payout ratio (Lamont (1998)), short term interest rates (Campbell (1987)), share of equity issues in total new debt and equity issues (Baker and Wurgler (2000)), and *cay* (Lettau and Ludvigson (2001)). The common theme underlying several of these variables (excluding share of equity issues) is that they capture investors' expectations about future returns. For example, scaling dividends by lagged prices gives a proxy for expected returns. The second strand of literature establishes a link between returns and innovations of macroeconomic variables. Chen, Roll and Ross (1986) find that monthly growth in Industrial Production, unexpected inflation, and default spread influence returns[4].

3. Data Collection and Description of Variables

This section describes the construction of the variables used in the empirical analysis. Given the number of variables involved, we group them into 3 groups:

---

[3]Net Sales are defined as New Sales–Redemptions+Exchanges-In–Exchanges-Out. For equity funds, exchanges in (out) represent flow to (from) equity funds in a fund complex from (to) non-equity funds in the same complex.

[4]There are numerous papers in each of these strands of literature. Welch and Goyal (2008), Campbell and Thompson (2008) and Flannery and Protopapadakis (2002) are examples of recent representative papers.

(i) Mutual Fund Flow Variables, (ii) Predictive Variables, and (iii) Macroeconomic Variables. In the discussion that follows, unless specified otherwise, respective variables are in millions of dollars, are seasonally adjusted at annual rates, are in real terms, and are available for the entire sample period of 1951Q4-2007Q4 (see Table 1 for a brief description of all the variables).

3.1 Mutual Fund Flow Variables

Prior literature (Warther (1995) and Fant (1999)) investigating aggregate mutual fund flow uses data from the Investment Company Institute (ICI), a national association of U.S. investment companies. The ICI provides monthly data on cash flows into mutual funds and cash balances for different investment objectives such as aggressive growth, corporate bonds, and precious metals. For each investment objective, cash flows are further divided into four categories: New Sales, Redemptions, Exchanges-in, and Exchanges-out. New sales and redemptions reflect transactions between the mutual fund and the investor, whereas exchanges-in and exchanges-out reflect transfers between funds within the same fund complex. Flows/Net Sales are defined as the sum of new sales and exchanges-in minus the sum of redemptions and exchanges-out. Net purchases are defined as net sales minus change in cash balances. Data are available for several investment objectives starting in 1984, and for only three equity fund objectives from 1976-1984. Since several of the variables discussed below are available at a quarterly frequency, we use an alternative source–the Federal Reserve Board's quarterly Z.1 statistical release *Flow of Funds Accounts of the United States*–that has the advantage of providing a longer time series.

The *Flow of Funds Accounts* (FFA hereafter) is organized by sector and by instruments. The mutual fund sector in the FFA includes all open-end investment companies (including Unit Investment Trusts), but excludes money-market mutual funds, limited-maturity municipal bond funds, variable annuities and hedge funds. *Table F.122 Mutual Funds* reports Gross Saving, Net Acquisition of Financial Assets, and Net Share Issues (Liabilities) for the mutual fund sector on an annual basis from 1946 and on a quarterly basis from 1951Q4. The *Net Acquisition of Financial Assets* series is further divided into the following asset categories: Security Repurchases, Credit Market Instruments, Corporate Equities and Miscellaneous Assets. Credit Market Instruments are further divided into Open Market Paper, U.S. Government Securities, which is composed of Treasury and Agency securities, Municipal securities, and Corporate and Foreign Bonds. Since our scope is limited to equity funds, we use the Corporate Equities series which is available in three versions: Seasonally Adjusted (SA) Flow, Not Seasonally Adjusted (NSA) Flow, and Not Seasonally Adjusted Level. We model the time series of Corporate Equities in a later section (Section 4.1); hence, we choose the NSA Flow version for the empirical analysis[5]. Thus the series we use as flows represents the dollar amount of stocks purchased by the mutual fund sector during a quarter[6]. Since this series shows an increasing trend over the sample period, we normalize the flows during quarter *t* by the total value of the stock market at the end of quarter *(t-1)*. We use the CRSP series "USDVal", which is defined as the total market value of all securities used in the CRSP

[5]For a series whose levels are reported at market values, the first differences are composed of flows and change in market values (capital gain or capital loss). The description of the Corporate Equities NSA Flow series in the *Guide to the Flow of Funds Accounts, Volume 1* (pg. 600) indicates that it represents flows. A plot of the first differences of the NSA Level series and NSA Flows series also indicates that these two series are not identical.

[6]Warther (1995) uses Net Sales in his analysis, but mentions that Net Purchases and Net Sales are highly correlated (0.89) and that his results are insensitive to the series used.

NYSE/AMEX/NASDAQ Index, for this purpose. In Section 4.1 we show that the normalized series exhibits all findings of earlier literature[7].

3.2 Predictive Variables

This group consists of variables that have been shown to predict returns on the market.

3.2.1 Term Spread (term): We follow Fama and French (1989) and define term spread as the difference between the yield of a portfolio of AAA-rated corporate bonds and the yield of a one month T-Bill. The portfolio yield is *Moody's Seasoned Aaa Corporate Bond Yield* series available at the Federal Reserve Economic Data (FRED hereafter). The T-Bill data is obtained from the CRSP Risk-Free Rates file and is the average of yields calculated using bid and ask prices.

3.2.2 Default Spread (def): The default spread is the difference between the yield of a portfolio of BAA-rated corporate bonds and that on a portfolio of AAA-rated corporate bonds. Both series are obtained from the FRED.

3.2.3 T-Bill Rate (rrel): Campbell (1987) shows that T-Bill rates predict stock returns. We use the *3-month Treasury Bill: Secondary Market Rate* series from the FRED and follow Campbell (1991) and Hodrick (1992) to stochastically detrend this series by subtracting the average yield over the past 12-months from each observation.

3.2.4 Dividend Yield (d/y): We follow the literature (see Keim and Stambaugh (1986) and Fama and French (1989)) in constructing this variable. The dividend yield is the dividend

[7]In the discussion, unless specified otherwise, "flows" refers to the normalized flows series. All empirical analysis uses the normalized flows series.

on the S&P 500 Index divided by *lagged* level of the index. Dividend on the S&P 500 is a moving sum of dividends paid by the index during the past 4-quarters (ending in the existing quarter). Thus the dividend yield for quarter *t* is sum of dividends paid by the index during quarters *(t-3)* through *t* divided by the index level at the end of quarter *(t-4)*. We obtain the dividend data till 2005 from Amit Goyal's website and that for 2006 and 2007 from the S&P Corporation.

3.2.5 Payout Ratio (d/e): The payout ratio (Lamont (1998)) is defined as the ratio of dividend to earnings of the S&P 500 Index. Earnings of the index are defined analogously to the dividend – moving sum of earnings paid by the index during the past 4-quarters (ending in the existing quarter).

3.2.6 Book to Market Ratio (b/m): Pontiff and Schall (1998) and Kothari and Shanken (1997) demonstrate the predictive ability of the book to market ratio of the Dow Jones Industrial Average. We obtain this data from Amit Goyal's website for 1951Q4 to 2005Q4, and follow Pontiff and Schall (1998) to construct this ratio for 2006Q1 to 2007Q4. In particular, the B/M ratio for quarter *t* is defined as the book value at the previous fiscal year-end divided by the level of the index at the end of quarter *t*. Since the DJIA is a price-weighted index, its fiscal year-end book value is the sum of the book values of stocks in the index divided by the year-end index divisor. For 2005, this value is obtained from Value Line's publication *A Long-Term Perspective, Dow Jones Industrial Average 1920-2005*. For 2006, we first calculate the year-end (31^{st} December 2006) divisor by dividing the sum of year-end stock prices for firms in the DJIA by the year-end index level. The index book value is then calculated as the sum of book values of stocks in the DJIA for the fiscal year 2006 divided by the index divisor[8].

3.2.7 Share of Equity Issues (eqis): Baker and Wurgler (2000) show that the share of equity issues in total new debt and new equity issues predicts returns. They construct this variable using data from the *Federal Reserve Bulletin* and work at the annual frequency, since monthly data for debt and equity issues is not available for the full sample period of 1927-1996. However, for our sample period (1951Q4-2007Q4) this data is available at the monthly frequency. Consequently, we construct a monthly series of new debt and new equity issues to obtain a quarterly series for their variable (see Appendix A for details).

3.2.8 Consumption, Asset Holdings and Labor Income (cay):

We follow Lettau and Ludvigson (2001) to construct their *cay* variable. First, we estimate the equation below using all observations from 1951Q4 to 2007Q4.

$$c_t = \alpha + \beta_a a_t + \beta_y y_t + \sum_{i=-k}^{k} \beta_{a,i} \cdot \Delta a_{t+i} + \sum_{i=-k}^{k} \beta_{y,i} \cdot \Delta y_{t+i} + \varepsilon_t, \qquad t = (k+1) \ldots (T-k)$$

In this equation, c_t refers to log per-capita real consumption, a_t refers to log per-capita real asset holdings, y_t refers to log per-capita real labor income, and Δ indicates the first-difference operator. Consumption and labor income are as defined below (see Section 3.3.2 and 3.3.3), and asset holdings are the household net-worth series from the Federal Reserve Board of Governors. We obtain nominal aggregate figures from the sources and then convert them to real per-capita values. To convert nominal figures to real, we use the series *Consumer Price Index for All Urban Consumers: All Items*, available from the

[8]During 2006, no firms were added to or deleted from the DJIA. Amongst the firms that were included in the DJIA during 2006, 24 had fiscal year-end of December, 2 (Wal-Mart Stores Inc. and Home Depot Inc.) had fiscal year-end of January, 2 (Proctor & Gamble Co. and Microsoft Corp.) had fiscal year-end of June, and the remaining has fiscal year-end of September (Disney Co.) and October (Hewlett-Packard Co.).

FRED, and use the year 2000 as the base year. To obtain per-capita figures, we divide by the mid-period population. The population data is obtained from *Table 2.1: Personal Income and its Disposition* (Row 38). Next, the estimates of *cay* are obtained as $cay_{t} \equiv (c_t - \hat{\beta}_a a_t - \hat{\beta}_y y_t)$, t=1…T. For our sample, this becomes $cay_{t} = (c_t - 0.4727 a_t - 0.74 y_t)$, t=1…T.

3.3 Macroeconomic Variables

This group consists of macroeconomic variables whose innovations have been shown to influence market returns. We first discuss the construction of each series and then the procedure used to obtain the innovations.

3.3.1 Gross Domestic Product: We obtain this data from *Table 1.1.5: Gross Domestic Product*, published by the U.S. Department of Commerce, Bureau of Economic Analysis.

3.3.2 Consumption: We obtain this data from *Table 2.3.5: Personal Consumption Expenditures by Major Type of Product*, published by the U.S. Department of Commerce, Bureau of Economic Analysis. We follow Lettau and Ludvigson (2001) and define consumption as the sum of non-durables and services excluding shoes and clothing.

3.3.3 Labor Income: We follow Lettau and Ludvigson (2001) and define after-tax labor income as the sum of Wages and Salaries, Transfer Payments, and Employer contributions for employee pensions and insurance, minus the sum of Employee contributions for social insurance and Taxes. Employee contributions for social insurance are defined as Contributions for government social insurance minus Employer contributions for government social insurance. Taxes are defined as [wages and salaries / (wages and salaries + proprietors' income with IVA and Ccadj + rental income + personal dividends + personal interest income)] * personal current taxes, where IVA is Inventory Valuation and Ccadj is capital consumption adjustments. This definition differs from the one provided in the Appendix of Lettau and Ludvigson (2001), is based on a revised version on Martin Lettau's website, and accounts for the revision of the NIPA accounts in 2003 by the BEA. See Martin Lettau's website for further details. All items in the above definition are obtained from *Table 2.1: Personal Income and Its Disposition*, published by the U.S. Department of Commerce, Bureau of Economic Analysis.

3.3.4 Industrial Production: We use the *Industrial Production Index* series from the FRED. The index currently has the year 2002 as the base year.

3.3.5 Money Supply: We use the series *M1 Money Stock* available at the FRED for data from January 1958 to December 2007. We use the *Banking and Monetary Statistics 1941-1970* publication available through the Federal Reserve Archival System for Economic Research (FRASER hereafter) to obtain the data from October 1951 to December 1958.

3.3.6 Unemployment: We use the series *Unemployed* available at the FRED for this variable. This data is provided by the U.S. Department of Labor, Bureau of Labor Statistics (BLS) and is obtained through the Current Population Survey (CPS), which is a monthly survey of households conducted by the Bureau of Census for the BLS.

We obtain the innovations by estimating a Vector Autoregression (VAR) consisting of log real aggregate Consumption, log real aggregate Gross Domestic Product, log real Money Supply (M1), log real aggregate Labor Income, and log Unemployment. We first check whether each time series is stationary by using the Phillips-Perron unit root tests and find that all series are non-stationary. All series are integrated of order one–they are non-stationary but the first differences are stationary. Next, we determine the lag length to be used in estimating the VAR using the Akaike Information Criteria (AIC). The minimum value of the criterion is -44.8431 and obtained when three lags are used. We then check whether the VAR(3) is stationary by examining the roots of the characteristic polynomial. The modulus of the eigen value of the first root was equal to 1 whereas that of the other 14 roots were less than one, indicating that the VAR(3) is non-stationary. Hence, we investigate the number of co-integrating vectors using the Trace test and find evidence of one co-integrating vector (Trace statistic=37.7138 and 5% Critical Value=47.21 for the null (alternative) hypothesis that Rank=1 (greater than 1)). Finally, we estimate an Error Correction Model and use the residuals as innovations for the macroeconomic variables. Since all macroeconomic variables are announced with a lag, we use residuals from quarter *(t-1)* as residuals for quarter *t*.

Finally, we use the returns on the VW CRSP NYSE/NASDAQ/AMEX Index that include dividends as a proxy for returns on the market portfolio.

4. Results from Existing Literature

This section examines the relationship between (i) returns and flows (Section 4.1), and (ii) returns and predictive variables and/or macroeconomic variables (Section 4.2). These have already been examined in earlier literature. However, we document them since they are either required given our measure of flows or form the building blocks for results in later sections.

4.1 Relationship between Returns and Flows

Since flows are highly autocorrelated (see Table 2), the relation between returns and the expected component may be different from that between returns and the unexpected component. Hence, we first split flows into expected and unexpected components using the Box-Jenkins methodology[9].

Table 4 reports regressions of flows on its lagged values. The second last row shows the test statistic for the general LaGrange multiplier test of Godfrey (1978) and Breusch (1978) which is used to test for the presence of autocorrelated residuals. The null hypothesis is that the residuals are not autocorrelated and the alternative is that they are AR(1). The last row gives the p-values. The first lag explains about 56% of the variation in flows and the coefficient is significant and positive. Inclusion of the second lag results in a significant coefficient and the adjusted R^2 increases to 0.58. However, adding the third lag makes the second lag insignificant. When the first four lags (Regression [4]) are included, only the first and fourth are significant and the adjusted R^2 increases to 0.61. Additional lags (fifth through seventh) are not significant and do not affect the adjusted R^2 much. When the eighth lag (Regression [8]) is included, the fourth lag becomes insignificant and only the first and eighth lags are significant. Given these patterns, Regression [9] includes only the first and fourth lag. The adjusted R^2 for this model is almost identical to that of Regression [4]. Similarly, Regression [10] includes only the first, fourth and eighth lags, each of which are significant. Again, its adjusted R^2 is almost identical to that of Regression [8]. Thus, both Regression [9] and Regression [10] seem to model the flow series appropriately. Since their adjusted R^2s differ by only 0.0142, we

[9]A prerequisite for this procedure is that the series be stationary. Phillips-Perron unit root tests strongly reject the null hypothesis of a unit root (p-value 0.001) in flows.

choose the parsimonious model–Regression [9]. The fitted values from this regression are the expected component of flows and the residuals are the unexpected component.

Regression [1] in Table 5 indicates a contemporaneous positive relation between returns and flows. The inclusion of lagged flows (Regression [2]) increases the coefficient estimate of contemporaneous flows by a factor of 2.4 and adjusted R^2 from 0.048 to 0.122. In addition, the lagged flows are significant and negatively related to returns. Inclusion of the second and third lags of flows (Regression [3] and [4]) is not helpful; neither are their coefficients significant nor do the adjusted R^2s increase. The findings of Regression [2] could be consistent with the Price Pressure Hypothesis or the fact that lagged flows are removing the expected part from contemporaneous flows. To distinguish between these two, when lagged flows are included as the only regressor in the model (Regression [5]), they are insignificant. Regression [6], which includes only the expected and unexpected components of contemporaneous flows, shows that only the unexpected component is (positively) related to returns, whereas the expected part is not. When lagged flows are included along with the expected and unexpected components of contemporaneous flows (Regression [7]), they are insignificant. This suggests that in Regression [2] lagged flows are proxying for the expected part of contemporaneous flows. Regression [8] includes lags of the expected and unexpected components, but only unexpected component of contemporaneous flows is significant. Given the results of Regression [6] through [8], Regression [9] includes only the unexpected component of contemporaneous flows. The adjusted R^2 of this specification is 0.124 and is comparable to that of earlier specifications.

Regression [10]-[14] investigates whether returns affect unexpected flows. There is a positive relation between unexpected flows and lagged returns. The second and third lags of returns are not significant and only change the adjusted R^2 by a small amount. When the first four lags are included, only the first and fourth are significant and the adjusted R^2 increases to 0.034.When only the first lag of returns is included as the dependent variable, it is positive and significant; however, the adjusted R^2 is 0.015. This finding provides support for the feedback-trader hypothesis.

Taken together, the results in Table 5 show the following facts: (i) The positive contemporaneous relationship between returns and flows is driven by the unexpected component of contemporaneous flows; the expected part is not related to returns, (ii) No support for the Price-Pressure Hypothesis since lagged flows do not affect returns negatively, and (iii) Support for the Feedback-Trader Hypothesis since lagged returns *positively* affect the unexpected component of flows. (i) and (ii) above are consistent with Warther (1995); however, (iii) is inconsistent since he finds evidence of negative feedback-trading.

4.2 Relationship between Returns, Predictive and Macroeconomic Variables

4.2.1 Returns and Predictive Variables

We begin by including all predictive variables (Table 6, Regression [1]). In this case, only *eqis* and *cay* are significant. Regression [2] therefore retains only *eqis* and *cay*. Table 3 shows that returns are correlated positively (0.15) to *term* and negatively (-0.19) to *rrel*. Hence, in Regression [3] *rrel* is included along with *eqis* and *cay*. The adjusted R^2 increases marginally from 0.09 to 0.098, but *rrel* is not significant. Next, *term* is included along with *eqis* and *cay* (Regression [4]). It is not significant and the adjusted R^2 decreases to 0.086. From the remaining variables, we find that when d/y is included along with *cay* and *eqis*, it is significant and the adjusted R^2 slightly increases to 0.102. The increase in adjusted R^2 from Regression [2] to [5] indicates that *eqis* and *cay* together explain a greater variation in return than does d/y. In unreported results, we find that when d/y and b/m are included along with *eqis* and *cay*, both become insignificant since they are highly correlated (0.86). Their VIFs increase by a factor of 4 compared to when only one of them is included along with *eqis* and *cay*. Note that the adjusted R^2 in regression [5] is typical of the return predictability literature (see for example Table 3 in Lettau and Ludvigson (2001))

4.2.2 Returns and Macroeconomic Variables

Table 7 investigates the relation between quarterly real returns on the CRSP VW NYSE/NASDAQ/AMEX Index and innovations of several macroeconomic variables. When all macroeconomic variables are included (Regression [1]), none are significant at the 5% level, and *y* is significant at the 10% level. Hence in Regression [2], only *y* is included; however, it loses its significance and the adjusted R^2 drops from 0.008 to 0.001. Table 3 indicates that no other macroeconomic variable is correlated with returns. Hence, we sequentially add a variable to Regression [2], beginning with the one that is least correlated with *y–unemp* (Regression [3]). When *m1* (Regression [4]) is included along with *y*, the adjusted R^2 increases from 0.001 to 0.011, but neither variable is significant. To this specification (Regression [4]), we add *unemp* since it is uncorrelated with *m1* and its correlation with *y* is low, and find that it lowers the adjusted R^2 to 0.007. Finally, when we add *gdp* to Regression [4], *y* becomes significant at the 5% level and the adjusted R^2 increases to 0.012.

4.2.3 Returns, Predictive Variables, and Macroeconomic Variables

Table 8 looks at the relation between returns, predictive variables, and macroeconomic variables. Again, we begin with both sets of variables (Regression [1]) and find that only *eqis* is significant; none of the macroeconomic variables are significant. Hence, Regression [2] includes only *eqis*. Although *eqis* is significant, the adjusted R^2 decreases from 0.09 to 0.022. Since the results of Section 3.2.1 indicate that *cay* has explanatory power for returns, Regression [3] includes it. *cay* is significant and the adjusted R^2 now equals that of Regression [1], in which all variables were included. Regression [4] includes *m1* since is uncorrelated to both *eqis* and *cay* (Table 3). *m1* is not significant but increases the adjusted R^2 to 0.092. When d/y is included along with *eqis* and *cay* (Regression [5]), it is significant and increases the adjusted R^2 to 0.102. Regressions [6] to [10] investigate whether any macroeconomic variable can explain variation in returns, beyond that explained by *eqis*, *cay*, and d/y. None of the macroeconomic variables are significant and the adjusted R^2s increase or decrease by a small amount. In essence, the macroeconomic variables are unimportant when predictive variables are included.

5. New Results

We now investigate the relation between unexpected flows, predictive and macroeconomic variables. Table 9 reports regressions of unexpected flows on predictive variables. In this table Regressions [1], [5] and [7] use Newey-West heteroskedasticity and autocorrelation consistent standard errors to obtain the *t*-statistics. The remaining specifications use OLS standard errors to calculate the *t*-statistics because specification tests indicated that the errors are homoskedastic (White's Test) and not autocorrelated (LM test of Godfrey (1978) and Breusch (1978)). Regression [1] shows that when all the predictive variables are included only *cay* is significant. In addition, the variance inflation factors (not reported in the tables) of d/y and b/m are relatively high (5.0904 and 4.8272 respectively) compared to that of other variables. Hence, Regression [2] excludes d/y and b/m. The correlation matrix (Table 3) shows that (i) unexpected flows are positively correlated (0.171) with *cay* and (ii) *cay* is most correlated with *term* (0.422), followed by *default* (0.34), and *rrel* (-0.237), and (iii) *cay* is uncorrelated with d/e and *eqis*. Given this, Regressions [3] through [5] sequentially drop a variable starting with the one that is most correlated with *cay–term*. Neither of the remaining variables becomes significant, and *cay* retains its significance. Regression [6] and [7] drop d/e and *eqis* sequentially. Regression [8] shows that when *cay* is the only independent variable, it is significant and the adjusted R^2 is higher than that of previous specifications (Regressions [1] through [7]).

Table 10 reports regressions of unexpected flows on macroeconomic variables. In this table, all regressions use OLS standard errors to calculate the *t*-statistics because specification tests indicated that the errors were homoskedastic and not autocorrelated. When all the variables are included, none are significant (Regression [1]). Since the *t*-statistics of *y* and *m1* are close to the 10% significance level, Regressions [2] to [4] include them together and separately. None of them is significant, but when both are included (Regression [2]), the adjusted R^2 (0.01) is higher than that of Regression [1] (0.004). Since, *unemp* is uncorrelated to *m1* and its correlation with *y* is relatively low, Regression [5] includes *unemp*. Labor income is significant and negatively related to unexpected flows and money supply is significant at the 10% level. The adjusted R^2 also increases from 0.01 to 0.013. Given the results in Table 7, this weak explanatory power of macroeconomic variables is not surprising.

Table 11 examines the relation between unexpected flows and both sets of variables. In this table, Regressions [1], [4] and [5] use Newey-West heteroskedasticity and autocorrelation consistent standard errors to calculate the *t*-statistics. Regression [1] includes both sets of variables, and only *cay* is significant. Regression [2] therefore includes only the significant variable–*cay*. It is significant and the adjusted R^2 increases from 0.019 to 0.025. Since *m1* is uncorrelated to *cay*, Regression [3] includes *m1* as an independent variable along with *cay*; it is not significant and marginally increases the adjusted R^2. Regression [4] adds b/m to the variables in Regression [3]; while the adjusted R^2 increases to 0.033, *cay* still remains the only variable that is significant. This result remains unchanged when *y* is included (Regression [5]).

Finally, Table 12 reports regressions of returns on unexpected flows, predictive and macroeconomic variables. Regression [1] is identical to Regression [9] in Table 4 and is included for purposes of comparison. We begin by including both sets of variables (Regression [2]). Amongst the predictive variable, only *eqis* is significant and none of the macroeconomic variables are significant. Together, these variables increase the adjusted R^2 from 0.124 to 0.185. When only unexpected flows and *eqis* are included as the independent variables (Regression [3]), both are significant but the adjusted R^2 decreases to 0.141 from 0.185. Regressions [4] and [5] include variables (*cay* and d/y) that were found to be significant in Section 3.2.1.These variables are once again significant and increase the adjusted R^2 to 0.20. Regressions [6] and [7] investigate the explanatory power of macroeconomic variables that were significant in Section 4.2.2. Both *y* and *m1* are insignificant and reduce the adjusted R^2s by a small amount. Thus, Regression [5] is the best model.

If the contemporaneous relation between unexpected flows and returns was a manifestation of the relation between predictive and macroeconomic variables, then adding both these sets of variables to Regression [1] should reduce the significance of unexpected flows. Compared to Regression [1], the coefficient on unexpected flows in Regression [8] is slightly lower, but is still significant. In addition, the adjusted R^2 of the regression of returns on unexpected flows (Regression [1]) is 0.124, and that of returns on predictive and macroeconomic variables (Regression [5] in Table 8) is 0.102. The adjusted R^2 of the regression of returns on unexpected flows, return predictor and macroeconomic variables (Regression [8]) is 0.2. This suggests that the two effects are largely independent.

This conclusion remains unchanged when we use a naïve method–log growth rate–to obtain the innovations of macroeconomic variables or use nominal figures as opposed to real. The statistical significance of variables that were found to be significant (e.g. *cay*, *d/y*, *eqis*, etc) increases, but none of the other variables become statistically significant.

6. Geweke's Decomposition

We now test the above hypothesis in a Vector Autoregression framework. To this end, we implement the results in Geweke (1982,1984) which measure the linear dependence between two vectors of time series, say X and Y, and decompose it into three parts: (i) Contemporaneous Dependence, (ii) Feedback from Y to X, and (iii) Feedback from X to Y. The next paragraph discusses only the implementation of the unconditional version of these tests; the extension to the conditional version is straightforward. For a more thorough discussion, the reader is referred to Geweke (1982, 1984).

Let W={$\mathbf{w}_t$, t=integer} be a stationary multiple time series that has been partitioned into 2 sub-vectors $\mathbf{x}_t$ and $\mathbf{y}_t$ with dimensions k x 1 and ℓ x 1 respectively. Consider the following Vector Autoregressions (VARs):

$$\mathbf{x}_t = \mathbf{a}_1 + \sum_{i=1}^{p} A_{1i}\mathbf{x}_{t-i} + \mathbf{u}_{1t}, \qquad \text{var}(\mathbf{u}_{1t}) = \Sigma_1 \qquad \ldots (5.1)$$

$$\mathbf{x}_t = \mathbf{a}_2 + \sum_{i=1}^{p} A_{2i}\mathbf{x}_{t-i} + \sum_{i=1}^{p} B_{2i}\mathbf{y}_{t-i} + \mathbf{u}_{2t}, \qquad \text{var}(\mathbf{u}_{2t}) = \Sigma_2, \qquad \ldots (5.2)$$

$$\mathbf{y}_t = \mathbf{b}_1 + \sum_{i=1}^{p} B_{1i}\mathbf{y}_{t-i} + \mathbf{v}_{1t}, \qquad \text{var}(\mathbf{v}_{1t}) = \Omega_1, \qquad \ldots (5.3)$$

$$\mathbf{y}_t = \mathbf{b}_2 + \sum_{i=1}^{p} A_{3i}\mathbf{x}_{t-i} + \sum_{i=1}^{p} B_{3i}\mathbf{y}_{t-i} + \mathbf{v}_{2t}, \qquad \text{var}(\mathbf{v}_{2t}) = \Omega_2 \qquad \ldots (5.4)$$

In these VARs, p is the lag length, $\mathbf{a}_i$ ($\mathbf{b}_i$) is a k x 1 (ℓx1) vector of constants, Σ_i is a k x k variance-covariance matrix of the residuals from (1) or (2) and Ω_i is an ℓ x ℓ variance-covariance matrix of the residuals from (3) or (4). In addition, let γ be an $(k+\ell)$ x $(k+\ell)$ variance-covariance matrix of the residuals from a VAR of $\mathbf{w}_t$ on its lags. Given the above VARs, γ can be written as $\begin{bmatrix} \Sigma_2 & C \\ C' & \Omega_2 \end{bmatrix}$, where C= $\mathbf{u}_{2t}\mathbf{v}'_{2t}$.Then the feedback from Y to X is defined as $\ell n(|\Sigma_1|) - \ell n(|\Sigma_2|)$, feedback from X to Y is defined as $\ell n(|\Omega_1|) - \ell n(|\Omega_2|)$, contemporaneous dependence between X and Y is defined as $\ell n(|\Sigma_2| \cdot |\Omega_2|) - \ell n(|\gamma|)$, and the total linear dependence, which is the sum of the above three components, is defined as $\ell n(|\Sigma_1| \cdot |\Omega_1|) - \ell n(|\gamma|)$. Since the time series are stationary, the above VARs can be estimated using OLS and estimates of the variance-covariance matrices are obtained as $\hat{\Sigma}_i = (1/T)\sum_{t=1}^{T} u_{it}u'_{it}$, $\hat{\Omega}_i = (1/T)\sum_{t=1}^{T} v_{it}v'_{it}$, i=1,2 and $\hat{C}$ = $(1/T)\sum_{t=1}^{T} u_{2t}v'_{2t}$ .Geweke (1982) showed that under the null hypothesis of:

(i) no feedback from Y to X, the LR test statistic $T \cdot \{\ell n(|\hat{\Sigma}_1|) - \ell n(|\hat{\Sigma}_2|)\}$ has χ^2 distribution with $k\ell p$ degrees of freedom,

(ii) no feedback from X to Y, the LR test statistic $T \cdot \{\ell n(|\hat{\Omega}_1|) - \ell n(|\hat{\Omega}_2|)\}$ has χ^2 distribution with $k\ell p$ degrees of freedom,

(iii) no contemporaneous dependence between X and Y, the LR test statistic $T \cdot \{\ell n(|\hat{\Sigma}_2| \cdot |\hat{\Omega}_2|) - \ell n(|\hat{\gamma}|)\}$ has χ^2 distribution with $k\ell$ degrees of freedom, and

(iv) no relation between X and Y, the LR test statistic $T \cdot \{\ell n(|\hat{\Sigma}_1| \cdot |\hat{\Omega}_1|) - \ell n(|\hat{\gamma}|)\}$ has χ^2 distribution with $k\ell\,(2p+1)$ degrees of freedom.

The conditional version is identical to the unconditional one except that the lags of the conditioning variables appear on the right hand side of equations (5.1) to (5.4) along with the existing variables.

The results of implementing this procedure are in Table 13. Panel A and B report the unconditional and conditional versions of the tests, respectively. Given the results of the earlier sections, we use d/y, *eqis*, and *cay* as the conditioning variables. Consistent with the results of earlier sections, the test statistic for contemporaneous dependence in a VAR(1) model decreases from 30.4261 to 26.2121 but the *p*-value remains unchanged.

7. Conclusion

We examine whether the investment decisions of fund investors incorporate the information contained in predictive and macroeconomic variables. If this is so, then aggregate flows should reflect the time series variation in expected market returns. Thus, flows should be related to predictive and macroeconomic variables, and controlling for predictive and macroeconomic variables, the contemporaneous relation between returns and flows should weaken. Our empirical analysis suggests that at the quarterly frequency, predictive and macroeconomic variables explain a small variation in flows. Consequently, the return-flow relationship is largely unaffected by predictive and macroeconomic variables–while the strength of the relationship weakens marginally, it does not become statistically insignificant. These results suggest that fund investors, as a group, do not adequately account for information in the predictive and macroeconomic variables while making their investment decisions.

Appendix A

This appendix describes the methodology used to construct the *eqis* variable, originally proposed by Baker and Wurgler (2000) (BW hereafter), at the quarterly frequency. The data source and the variables are identical to those used by BW; however, the methodology used to collect the data is slightly different.

We obtain the data from the *Financial and Business Statistics* section of the *Federal Reserve Bulletin*, a monthly bulletin published by the Federal Reserve Board. After 2004, this section is called *Statistical Supplement to the Federal Reserve Bulletin* and is published separately[10]. From this section, we use the information provided in the New Security Issues/Total New Issues/New Security Issues U.S. Corporations[11] table to construct BW's variable. For the time period January 1951–December 1970 and January 2004–April 2008, the Bulletin is available online at the FRASER and Board of Governors of the Federal Reserve System website respectively. For the remaining period, hard copies of the Bulletin were used to obtain the relevant information.

The New Security Issues table provides nominal dollar amounts of debt and equity issues in millions. Typically, in each issue of the Bulletin (say month *t*) the table reports monthly data for months (*t-4)* through (*t-10)* and annual data for 3 years for time period before month *(t-10)*. The month *(t+1)* issue of the Bulletin drops the month (t-10) observation of the month *t* issue of the Bulletin and adds the month (t-3) observation. The annual observations are updated in a similar way. Typically, the monthly (annual) figures remain in the Bulletin for one (three) year(s). Sometimes, the data reported in earlier issues are revised. These are indicated by a superscript "r" and are more frequent in the

[10]For simplicity, in the discussion we use the term Bulletin
[11]The title of the table changes over time. In the discussion we use the term New Security Issues.

late 1980s and 1990s.

We first divide the issues of the Bulletin from January 1953–April 2008 into smaller sub-periods: (i) January 1953–December 1970, (ii) January 1971–December 1980, (iii) January 1981–September 1987, (iv) October 1987–July 1999, and (v) August 1999–April 2008. For the earlier part of the sample the breakpoints are chosen arbitrarily (periods (i) and (ii)) and for the latter part of the sample they are governed by changes in the variables of interest reported in the Bulletin. For example, the September 1987 issue of the Bulletin reports two subcategories–Common Stock and Preferred Stock–for Stocks and two subcategories–Public and Private Placement–for Bonds. The October 1987 issue reports three subcategories (Common, Preferred, and Private Placement) for Stocks and three subcategories (Public, domestic; Private Placement, domestic; Sold Abroad) for Bonds.

Next we use issues of the Bulletin to obtain the relevant variables for each sub-period. In particular, within each sub-period we start with the last issue of the Bulletin and record the relevant variables for all the months and years. We then use the second to last issue of the Bulletin to obtain the variables for the month (and in some cases year) that was not present in the last issue. In calendar time of the Bulletin issues, this month corresponds to the one that is dropped from the Bulletin. But by filling the data in reverse calendar time within a sub-period, we ensure that the most updated value for a particular month is used. We repeat this process until the beginning of the sub-period is reached[12].

[12]In the early part of the sample, BW use the first occurrence of the data. In the later part of the sample, they account for revisions by using data from later issues of the Bulletin. However, they don't use the last occurrence of the data. This difference accounts for the imperfect correlation (0.99746) between their annual series and the annual series implied by our monthly series.

Consider for example, sub-period (iii). The September 1987 issue of the Bulletin reports annual data from 1984 to 1986, and monthly data from October 1986 to May 1987. We record the relevant variables for all the years and months published in this issue. The previous issue (August 1987) reports annual data from 1984 to 1986, and monthly data from September 1986 to April 1987. Thus, the data for September 1986 were not included in the September 1987 issue. We pick the relevant variables for September 1986 from the August 1987 issue of the Bulletin, thereby picking up the last occurrence of the data for the month September 1986. This process is repeated until the January 1981 issue of the Bulletin is reached.

The above procedure is first implemented on sub-period (i), then on sub-period (ii) and so on. The last occurrence of the monthly/annual data does not correspond to the breakpoints chosen above. Hence, while filling in the data for a particular sub-period, we update some of the data that was recorded while filling up the previous sub-period. In the above example, while working with sub-period (iii), the relevant variables for all years (1984-1986) and months (October 1986-May 1987) were recorded from the September 1987 issue. While working with sub-period (iv), each of these monthly and annual figures is updated. For example, the last occurrence of the data for May 1987 is the April 1988 issue of the Bulletin and that for the year 1986 is the March 1990 issue of the Bulletin.

Using the above procedure and the definitions of Debt Issues (*d*) and Equity Issues (*e*) used by BW, the following formulas are used to get the monthly of *d* and *e*:

$$d_t = Public + \text{Private} \qquad \ldots \qquad t \le Dec.1982$$

$$= Public + \text{Private} \qquad \ldots \qquad Jan.1983 \le t \le Aug.1998$$

$$= \text{Sold in U.S.} \qquad \ldots \qquad t \geq Sep.1998$$

$$e_t = \text{Preferred } Stock + Common\ Stock \qquad \ldots \qquad t \leq Aug.1987$$

$$= Public \qquad \ldots \qquad t \geq Sept.1987$$

Note that the monthly numbers for “Private Placement of Bonds” for January 1983 to August 1998 were obtained from the annual “Private Placement” numbers for 1983 to 1998 assuming that they were uniformly distributed over the 12 months. The quarterly series of *eqis* is obtained by using the following formula: $S_t = \frac{\sum_{i=0}^{2} e_{t-i}}{\sum_{i=0}^{2}(e_{t-i} + d_{t-i})}$, where t = 3, 6, 9, or 12. To gauge the accuracy of our methodology, we compare the annual *eqis* implied by our monthly series and that provided by BW on their website. At the time of this writing, their website provided annual *eqis* from 1927 to 2004. For the overlapping sample period of 1952 to 2004, the mean and standard deviation of BW’s series is 0.19171 and 0.07877 respectively. The annual *eqis* implied by our monthly series has a mean and standard deviation of 0.19329 and 0.07786 and the correlation between the two series is 0.99746.

REFERENCES

Akaike, H, 1974, A new look at the statistical model identification, *IEEE Transactions on Automatic Control* 19, 716-723.

Alexander, Gordon J., Jonathan D. Jones, and Peter J. Nigro, 1998, Mutual fund shareholders: Characteristics, investor knowledge, and sources of information, *Financial Services Review* 7, 301-316.

Baker, Malcolm, and Jeffrey Wurgler, 2000, The equity share in new issues and aggregate stock returns, *The Journal of Finance* 55, 2219-2257.

Bass, Steven, John Sabelhaus, and Daniel Schrass, 2009, Characteristics of mutual fund investors, 2008, *Research Fundamentals, Investment Company Institute* 18.

Breusch, T. S., 1978, Testing for autocorrelation in dynamic linear models, *Australian Economic Papers* 17, 334-355.

Campbell, John Y., 1987, Stock returns and the term structure, *Journal of Financial Economics* 18, 373-399.

Campbell, John Y., 1991, A variance decomposition for stock returns, *The Economic Journal* 101, 157-179.

Campbell, John Y., and Samuel B. Thompson, 2008, Predicting excess stock returns out of sample: Can anything beat the historical average?, *Review of Financial. Studies* 21, 1509-1532.

Chen, Nai-Fu, Richard Roll, and Stephen A. Ross, 1986, Economic forces and the stock market, *The Journal of Business* 59, 383-403.

Fama, Eugene F., and Kenneth R. French, 1989, Business conditions and expected returns on stocks and bonds, *Journal of Financial Economics* 25, 23-49.

Fant, L. Franklin, 1999, Investment behavior of mutual fund shareholders: The evidence from aggregate fund flows, *Journal of Financial Markets* 2, 391-402.

Flannery, Mark J., and Aris A. Protopapadakis, 2002, Macroeconomic factors do influence aggregate stock returns, *Review of Financial Studies* 15, 751-782.

Geweke, John F., 1982, Measurement of linear dependence and feedback between multiple time series, *Journal of the American Statistical Association* 77, 304-313.

Geweke, John F., 1984, Measures of conditional linear dependence and feedback between time series, *Journal of the American Statistical Association* 79, 907-915.

Godfrey, L. G., 1978, Testing for higher order serial correlation in regression equations when the regressors include lagged dependent variables, *Econometrica* 46, 1303-1310.

Goetzmann, William, and Nadav Peles, 1997, Cognitive dissonance and mutual fund investors, *Journal of Financial Research* 20, 145-158.

Hodrick, R. J., 1992, Dividend yields and expected stock returns: Alternative procedures for inference and measurement, *Rev. Financ. Stud.* 5, 357-386.

Keim, Donald B., and Robert F. Stambaugh, 1986, Predicting returns in the stock and bond markets, *Journal of Financial Economics* 17, 357-390.

Kothari, S. P., and Jay Shanken, 1997, Book-to-market, dividend yield, and expected market returns: A time-series analysis, *Journal of Financial Economics* 44, 169-203.

Lamont, Owen, 1998, Earnings and expected returns, *The Journal of Finance* 53, 1563-1587.

Lettau, Martin, and Sydney Ludvigson, 2001, Consumption, aggregate wealth, and expected stock returns, *The Journal of Finance* 56, 815-849.

Pontiff, Jeffrey, and Lawrence D. Schall, 1998, Book-to-market ratios as predictors of market returns, *Journal of Financial Economics* 49, 141-160.

Warther, Vincent A., 1995, Aggregate mutual fund flows and security returns, *Journal of Financial Economics* 39, 209-235.

Welch, Ivo, and Amit Goyal, 2008, A comprehensive look at the empirical performance of equity premium prediction, *Review of Financial. Studies* 21, 1455-1508.

Wilcox, Ronald T., 2003, Bargain hunting or star gazing? Investors' preferences for stock mutual funds, *The Journal of Business* 76, 645-663.

Table 1: Glossary and Definitions of Variables

Unless specified otherwise, respective variables are in millions of dollars, are seasonally adjusted at annual rates, are in real terms, and are available for the entire sample period 1951Q4-2007Q4. We use the series, *Consumer Price Index for All Urban Consumers: All Items*, available from the FRED, to convert nominal figures to real, and use the year 2000 as the base year. To obtain innovations of the macroeconomic variables, we estimate a VAR(3) consisting of log real aggregate Consumption, log real aggregate Gross Domestic Product, log real Money Supply (M1), log real aggregate Labor Income, and log Unemployment. We use the residuals from the VAR as innovations. For quarter *t*, we use residuals from quarter *(t-1)* since all macroeconomic variables are announced with a lag. The abbreviations are: Bureau of Economic Analysis (BEA), Federal Reserve Archival System for Economic Research (FRASER), Federal Reserve Economic Data (FRED), and National Income and Product Account (NIPA). In the "Variable" column, the first letter in the square brackets indicates the type (Flow/Level) and the second letter indicates the frequency (monthly or quarterly) of the series available from the data source. A "." indicates Not Applicable.

SYMBOL	VARIABLE	DEFINITION/SOURCE
Return		
ret_t	Return [F, Q]	Quarterly VW real return on CRSP NYSE/NASDAQ/AMEX Index that includes dividends.
Mutual Fund Flow Variables		
$flow_t$	Fund Flows [F, Q]	Dollar amount of Corporate Equities (seasonally unadjusted) bought by the mutual fund sector during quarter *t*. Data are obtained from the Federal Reserve Board's quarterly Z.1 statistical release *Flow of Funds Accounts of the United States*. The mutual fund sector covers all open-end investment companies (including Unit Investment Trusts), but excludes money-market mutual funds, limited-maturity municipal bond funds, variable annuities and hedge funds. Normalized flows are obtained by dividing flows during quarter *t* by market value of NYSE, NASDAQ, and AMEX at the end of quarter (*t-1)*.
$nflow_t$	Normalized Fund Flows [F, .]	
Predictive Variables		
$(d/y)_t$	Dividend-Yield [F, Q]	Log of the ratio of dividend on the S&P 500 Index to lagged level of the index (Fama and French (1987)). Dividends are defined as the moving sum of dividends paid by the index during the past 4 quarters (ending in the existing quarter) and the lagged level corresponds to that at the end of quarter (*t-4)*. We obtain the dividend data till 2005 from Amit Goyal's website and that for 2006 and 2007 from the S&P Corporation.
$(d/e)_t$	Payout Ratio [F, Q]	Log of the ratio of dividend to earnings for the S&P 500 Index (Lamont (1998)). Earnings of the index are defined analogously to the dividend – moving sum of earnings paid by the index during the past 4 quarters (ending in the existing quarter).
$(b/m)_t$	Book-to-Market [L, .]	Log of the ratio of book value to market value for the DJIA (Pontiff and Schall (1998) and Kothari and Shanken (1997)).Data until 2005 are from Amit Goyal's website, and for 2006-2007 are calculated following Pontiff and Schall (1998).
$term_t$	Term Spread [L, M]	Difference between the yield of a portfolio of AAA-rated corporate bonds and the yield on a one month T-Bill. The portfolio yield is *Moody's Seasoned Aaa Corporate Bond Yield* series from the FRED. The T-Bill data is obtained from the CRSP Risk-Free Rates file and is the average of yields calculated using bid and ask prices.

Table 1 (contd.) : Glossary and Definitions of Variables

SYMBOL	VARIABLE	DEFINITION/SOURCE
Predictive Variables (contd.)		
$default_t$	Default Spread [L, M]	Difference between the yield of a portfolio of BAA-rated corporate bonds and that on portfolio of AAA-rated corporate bonds. Both series are obtained from the FRED.
$rrel_t$	Short-term T-Bill [L, M]	The *3-month Treasury Bill: Secondary Market Rate* series from the FRED. It is detrended by subtracting the average yield over the past 12-months from each observation (Campbell (1991) and Hodrick (1992)).
$eqis_t$	Equity Issue [F, M]	Share of equity issues in total new debt and equity issues (Baker and Wurgler (2000)). A monthly series of new debt and new equity issues, constructed from data published in the Federal Reserve Bulletin (see Appendix A for details), is used to obtain the quarterly series.
cay_t	Consumption, asset holdings and labor income [., Q]	Variable is constructed following the instructions in the Appendix of Lettau and Ludvigson (2001), revised definition of Labor Income on the authors' website, and using log and per capita values for consumption, labor income and household net-worth. The population data is obtained from *Table 2.1: Personal Income and its Disposition* (Row 38).
Macroeconomic Variables		
c_t	Consumption [F, Q]	Innovation in log aggregate consumption in quarter *t*. Consumption is defined as the sum of non-durables and services excluding shoes and clothing (Lettau and Ludvigson (2001)). The innovations are the residuals from the Consumption equation of the VAR(3). Data for nominal consumption are from *Table 2.3.5: Personal Consumption Expenditures by Major Product Type*, published by the BEA.
gdp_t	Gross Domestic Product [F, Q]	Innovation in log Gross Domestic Product in quarter *t*. The innovations are the residuals from the Gross Domestic Product equation of the VAR(3). Data for nominal GDP are obtained from *Table 1.1.5: Gross Domestic Product*, published by the BEA.
$m1_t$	Money Supply [L, M]	Innovation in log Money Supply (M1) at the end of quarter *t*. The innovations are the residuals from the Money Supply equation of the VAR(3). Data for October 1951-December 1958 are obtained from the *Banking and Monetary Statistics 1941-1970* publication available through FRASER and that for January 1958-December 2007 are from the *M1 Money Stock* series available at the FRED.
$unemp_t$	Unemployment [L, M]	Innovation in log Unemployment at the end of quarter *t*. The innovations are the residuals from the Unemployment equation in the VAR(3). The series *Unemployed* available at the FRED is used.
y_t	Labor Income [F, Q]	Innovation in log Labor Income in quarter *t*. The innovations are the residuals from the Labor Income equation in the VAR(3). The after-tax labor income is calculated following the revised definition on Martin Lettau's website; it accounts for the revision of the NIPAs in 2003 by the BEA. All items used in the definition are obtained from *Table 2.1: Personal Income and Its Disposition*, published by the BEA.

Table 2: Descriptive Statistics for Fund Flows, Predictive and Macroeconomic Variables

This table shows the descriptive statistics for fund flows, predictive variables and innovations of macroeconomic variables used in the text. The sample period is from 1951Q4 to 2007Q4 for all variables except Unflow (Panel B), and Innovations in Macroeconomic Variables (Panel D). For these the sample period is 1952Q4 -2007Q4. See Table 1 for description and definitions of these variables. Data for the following variables are reported in percentages: returns, normalized flows, term spread, default spread, and short-term T-bill. In the "Autocorrelation" column, bold face numbers indicate statistically significant first-order autocorrelation. The unexpected flows are defined as the difference between the actual normalized flows and expected normalized flows.

	Min	Q1	Median	Q3	Max	Mean	Std.Dev	Skewness	Kurtosis	Autocorrelation
Panel A: Real Returns										
Ret	-26.927	-2.244	2.953	6.911	22.354	2.080	8.147	-0.524	1.005	0.0585
Panel B: Mutual Fund Flows										
flow	-37101	56	347	11507	108306	9398.790	18109.730	1.981	5.158	**0.7052**
nflow	-0.281	0.015	0.090	0.214	0.909	0.134	0.199	1.247	1.916	**0.7496**
Unflow	-0.444	-0.053	-0.007	0.041	0.703	0.000	0.126	0.746	5.760	-0.0639
Panel C: Predictive Variables										
d/y	-4.542	-3.521	-3.375	-3.127	-2.671	-3.403	0.420	-0.703	0.165	**0.9835**
d/e	-1.190	-0.875	-0.674	-0.582	-0.269	-0.727	0.201	-0.442	-0.465	**0.9658**
b/m	-2.078	-1.100	-0.633	-0.336	0.184	-0.717	0.528	-0.636	-0.234	**0.9850**
term	-2.091	1.175	2.156	3.396	6.423	2.310	1.450	0.240	-0.360	**0.7691**
default	0.340	0.670	0.810	1.130	2.690	0.937	0.412	1.471	2.453	**0.9095**
rrel	-4.220	-0.404	0.075	0.610	4.615	0.030	1.075	-0.215	4.121	**0.5243**
eqis	0.041	0.119	0.171	0.238	0.471	0.188	0.095	0.776	0.257	**0.7689**
cay	-2.902	-2.834	-2.811	-2.790	-2.754	-2.815	0.032	-0.468	-0.508	**0.9306**
Panel D: Innovations in Macroeconomic Variables										
c	-0.019	-0.003	0.000	0.003	0.021	0.000	0.005	-0.027	1.377	-0.03551
y	-0.038	-0.006	0.001	0.007	0.031	0.000	0.010	-0.301	0.843	-0.02056
m1	-0.036	-0.008	0.000	0.007	0.053	0.000	0.012	0.470	1.948	-0.03545
gdp	-0.032	-0.005	0.000	0.006	0.033	0.000	0.009	-0.033	1.104	0.00132
unemp	-0.178	-0.029	-0.009	0.026	0.340	0.000	0.057	1.011	6.284	0.00566

Table 3: Correlation Matrix of Fund Flows, Predictive and Macroeconomic Variables

This table shows the correlation among the fund flows, predictive variables and innovations in macroeconomic variables used in regressions. Correlations are calculated using observations from 1952Q4 to 2007Q4 since the first four observations are lost in calculating unexpected flows. See Table 1 for description and definitions of these variables. A “.” indicates a statistically insignificant correlation.

	Ret_t	$Unflow_t$	d/y_{t-1}	d/e_{t-1}	b/m_{t-1}	$term_{t-1}$	def_{t-1}	$rrel_{t-1}$	$eqis_{t-1}$	cay_{t-1}	c_t	y_t	$m1_t$	gdp_t	$unemp_t$
Ret_t	1														
$Unflow_t$	0.358	1													
d/y_{t-1}	.	.	1												
d/e_{t-1}	.	.	0.352	1											
b/m_{t-1}	.	.	0.858	0.284	1										
$term_{t-1}$	0.152	.	-0.187	0.199	-0.218	1									
def_{t-1}	.	.	0.264	.	0.380	0.344	1								
$rrel_{t-1}$	-0.191	.	.	-0.284	.	-0.564	-0.336	1							
$eqis_{t-1}$	-0.164	.	0.514	0.212	0.517	-0.180	0.273	0.150	1						
cay_{t-1}	0.276	0.171	0.323	.	0.191	0.422	0.340	-0.237	.	1					
c_t	.	.	.	.	.	.	.	.	.	.	1				
y_t	.	.	.	.	.	.	.	.	.	-0.151	0.535	1			
$m1_t$	.	.	.	0.199	.	0.168	.	-0.341	.	.	0.440	0.435	1		
gdp_t	.	.	.	.	.	.	.	.	.	.	0.583	0.627	0.309	1	
$unemp_t$	.	.	.	.	.	0.155	0.165	-0.225	.	0.147	-0.230	-0.391	.	-0.497	1

Table 4: Splitting Flows into Expected Flows and Unexpected Flows

The table reports estimates from OLS regressions of normalized flows on its lagged values. Normalized flows are split into expected and unexpected components using Regression [9]. Expected flows are the fitted values from Regression [9]. Unexpected flows are the difference between the actual flows and expected flows. The sample period is 1951Q4 to 2007Q4. For each regression, the first line reports the coefficient estimate and the second line reports *t*-statistics in parentheses, calculated using Newey-West heteroskedasticity and autocorrelation consistent standard errors. *t*-statistics significant at the 5% (10%) level are indicated in bold face (italics). The second to last row reports test statistics for the general LaGrange multiplier test of Godfrey (1978) and Breusch (1978) which is used to test for the presence of autocorrelated residuals. The null hypothesis is that the residuals are not autocorrelated and the alternative is that they are AR(1). *p*-values, in parentheses, are indicated in the last row.

Independent Variables	*Dependent Variable: nflow*									
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]
Intercept	0.032	0.0248	0.020	0.016	0.016	0.017	0.014	0.012	0.019	0.013
	(**3.445**)	(**3.107**)	(**2.655**)	(**1.959**)	(**2.076**)	(**2.130**)	(1.839)	(1.493)	(**2.214**)	(1.473)
Lag 1	0.752	0.588	0.545	0.502	0.506	0.505	0.513	0.481	0.571	0.532
	(**14.810**)	(**8.366**)	(**6.709**)	(**5.257**)	(**5.503**)	(**5.437**)	(**5.704**)	(**4.869**)	(**6.432**)	(**5.388**)
Lag 2		0.220	0.108	0.083	0.085	0.096	0.096	0.120		
		(**3.121**)	(1.265)	(1.114)	(1.188)	(1.291)	(1.428)	(1.656)		
Lag 3			0.192	0.081	0.083	0.087	0.0540	0.057		
			(1.888)	(0.768)	(0.760)	(0.803)	(0.489)	(0.491)		
Lag 4				0.211	0.223	0.229	0.214	0.180	0.283	0.199
				(**2.504**)	(**2.216**)	(**2.224**)	(**2.018**)	(1.807)	(**3.140**)	(**2.540**)
Lag 5					-0.022	0.003	-0.016	-0.028		
					(-0.308)	(0.034)	(-0.188)	(-0.343)		
Lag 6						-0.052	-0.128	-0.153		
						(-0.727)	(-1.305)	(-1.558)		
Lag 7							0.157	0.064		
							(1.792)	(0.722)		
Lag 8								0.188		0.170
								(**2.339**)		(**2.523**)
R^2	0.562	0.583	0.598	0.616	0.616	0.617	0.626	0.639	0.609	0.624
Adj. R^2	0.560	0.580	0.593	0.609	0.607	0.606	0.614	0.625	0.605	0.619
LM Test	10.673	7.894	9.495	0.092	0.050	4.854	7.343	2.765	2.061	0.357
p-value	(0.001)	(0.005)	(0.002)	(0.762)	(0.824)	(0.028)	(0.007)	(0.096)	(0.151)	(0.550)

Table 5: Evidence on the Price Pressure and Feedback Trader Hypotheses

The table reports estimates from OLS regressions of the variable in the "Dependent Variable" column on those appearing under the "Independent Variables" column. CRSP VW Returns are returns (including dividends) to the CRSP NYSE/NASDAQ/AMEX Index. *nflow* is the net purchases of stocks by mutual funds divided by lagged total value of stock market (NYSE+AMEX+NASDAQ). *Enflow* (*Unflow*) is the fitted value (residuals) obtained from Regression [9] in Table 4. Regressions 1-5 are estimated using data from 1951Q4 to 2007Q4 and regressions 6-14 are estimated using data from 1952Q4 to 2007Q4. For each regression, the first line reports the coefficient estimate and the second line reports *t*-statistics in parentheses. *t*-statistics significant at the 5% (10%) level are indicated in bold face (italics). Regression [5] uses Newey-West standard errors and the remaining regressions use OLS standard errors because specification tests indicated that the errors were homoskedastic and not autocorrelated.

	Dependent Variable													
Independent Variables	*CRSP VW Returns*									*Unflow*				
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[13]	[14]
Intercept	0.830	1.448	1.508	1.549	2.163	2.061	1.759	1.687	2.100	-0.005	-0.002	-0.004	0.000	-0.001
	(1.300)	(**2.290**)	(**2.340**)	(**2.370**)	(**3.162**)	(**3.020**)	(**2.390**)	(**2.270**)	(**4.060**)	(-0.530)	(-0.210)	(-0.490)	(-0.020)	(-0.060)
nflow														
Lag 0	9.319	22.238	22.884	23.216										
	(**3.490**)	(**5.740**)	(**5.740**)	(**5.690**)										
Lag 1		-17.279	-15.760	-15.573	-0.549		-10.271							
		(**-4.440**)	(**-3.460**)	(**-3.390**)	(-0.228)		(-1.090)							
Lag 2			-2.702	-1.741										
			(-0.670)	(-0.380)										
Lag 3				-1.752										
				(-0.430)										
Enflow														
Lag 0						0.287	12.896	8.076						
						(0.090)	(1.070)	(0.570)						
Lag 1								-5.182						
								(-0.410)						
Unflow														
Lag 0						23.401	23.401	23.309	23.401					
						(**5.660**)	(**5.670**)	(**5.610**)	(**5.680**)					
Lag 1								-8.915						
								(-0.930)						
CRSP VW Return														
Lag 1										0.002	0.002	0.002	0.002	0.002
										(**2.110**)	(**2.190**)	(**2.240**)	(**2.220**)	(**2.070**)
Lag 2											-0.001	-0.001	-0.002	
											(-1.320)	(-1.390)	(-1.490)	
Lag 3												0.001	0.001	
												(1.210)	(1.340)	
Lag 4													-0.002	-0.002
													(**-1.980**)	(-1.840)
R^2	0.052	0.130	0.132	0.133	0.000	0.128	0.133	0.136	0.128	0.020	0.028	0.034	0.051	0.035
Adj. R^2	0.048	0.122	0.120	0.117	-0.004	0.120	0.121	0.120	0.124	0.015	0.019	0.021	0.034	0.026

Table 6: Returns and Predictive Variables

The table reports estimates from OLS regressions of returns (including dividends) to the CRSP NYSE/NASDAQ/AMEX Index on predictive variables. The sample period is 1952Q4 to 2007Q4. See Table 1 for definitions of the variables. For each regression, the first line reports the coefficient estimate, and the second line reports *t*-statistics in parentheses, calculated using OLS standard errors because specification tests indicated that the errors were homoskedastic and not autocorrelated. *t*-statistics significant at the 5% (10%) level are indicated in bold face (italics).

Independent Variables		*Dependent Variable: CRSP VW Returns*				
		[1]	[2]	[3]	[4]	[5]
Intercept		140.171	194.441	175.728	189.650	166.600
		(**2.460**)	(**4.270**)	(**3.760**)	(**3.730**)	(**3.520**)
d-y	Lag 1	4.472				3.223
		(1.560)				(**1.990**)
d-e	Lag 1	-1.301				
		(-0.390)				
b-m	Lag 1	-0.774				
		(-0.350)				
term	Lag 1	0.011			0.088	
		(0.020)			(0.220)	
def	Lag 1	0.195				
		(0.110)				
rrel	Lag 1	-0.984		-0.863		
		(-1.530)		(-1.700)		
eqis	Lag 1	-19.091	-12.951	-11.592	-12.737	-20.405
		(**-2.660**)	(**-2.330**)	(**-2.070**)	(**-2.250**)	(**-3.060**)
cay	Lag 1	42.960	67.468	60.900	65.853	53.177
		(**2.020**)	(**4.160**)	(**3.670**)	(**3.680**)	(**3.020**)
R^2		0.130	0.099	0.110	0.099	0.115
Adj. R^2		0.097	0.090	0.098	0.086	0.102

Table 7: Returns and Macroeconomic Variables

The table reports estimates from OLS regressions of returns (including dividends) to the CRSP NYSE/NASDAQ/AMEX Index on innovations of macroeconomic variables. The sample period is 1951Q4 to 2007Q4. See Table 1 for definitions of the variables and for the method used to obtain the innovations. For each regression, the first line reports the coefficient estimate and the second line reports *t*-statistics in parentheses, calculated using OLS standard errors because specification tests indicated that the errors were homoskedastic and not autocorrelated. *t*-statistics significant at the 5% (10%) level are indicated in bold face (italics).

Independent Variables		*Dependent Variable: CRSP VW Returns*					
		[1]	[2]	[3]	[4]	[5]	[6]
Intercept		2.097	2.103	2.103	2.100	2.100	2.097
		(**3.810**)	(**3.810**)	(**3.800**)	(**3.820**)	(**3.810**)	(**3.820**)
c	Lag 0	26.555					
		(0.200)					
y	Lag 0	-151.372	-59.650	-41.806	-108.730	-95.019	-161.401
		(*-1.900*)	(-1.050)	(-0.680)	(-1.730)	(-1.380)	(**-2.100**)
m1	Lag 0	78.072			93.282	89.536	90.087
		(1.420)			(1.790)	(1.700)	(1.730)
gdp	Lag 0	118.529					94.827
		(1.280)					(1.190)
unemp	Lag 0	11.045		7.800		5.132	
		(0.960)		(0.740)		(0.480)	
R^2		0.030	0.005	0.008	0.020	0.021	0.026
Adj. R^2		0.008	0.001	-0.002	0.011	0.007	0.012

Table 8: Returns, Predictive Variables and Macroeconomic Variables

The table reports estimates from OLS regressions of returns (including dividends) to the CRSP NYSE/NASDAQ/AMEX Index on predictive variables and innovations of macroeconomic variables. The sample period is 1951Q4 to 2007Q4. See Table 1 for definitions of the variables and for the method used to obtain the innovations. For each regression, the first line reports the coefficient estimate and the second line reports *t*-statistics in parentheses, calculated using OLS standard errors because specification tests indicated that the errors were homoskedastic and not autocorrelated. *t*-statistics significant at the 5% (10%) level are indicated in bold face (italics).

Independent Variables		*Dependent Variable: CRSP VW Returns*									
		[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]
Intercept		128.008	4.747	194.441	195.554	166.600	169.139	164.012	168.419	170.608	162.955
		(**2.180**)	(**3.930**)	(**4.270**)	(**4.290**)	(**3.520**)	(**3.560**)	(**3.420**)	(**3.550**)	(**3.580**)	(**3.400**)
Predictive Variables											
d-y	Lag 1	4.381				3.223	3.206	3.214	3.129	3.164	3.236
		(1.490)				(**1.990**)	(**1.980**)	(**1.980**)	(*1.930*)	(*1.950*)	(**1.990**)
d-e	Lag 1	-1.403									
		(-0.420)									
b-m	Lag 1	-0.825									
		(-0.370)									
term	Lag 1	0.104									
		(0.190)									
def	Lag 1	0.481									
		(0.260)									
rrel	Lag 1	-0.827									
		(-1.220)									
eqis	Lag 1	-19.123	-14.149	-12.951	-12.549	-20.405	-20.324	-20.345	-19.825	-20.490	-20.405
		(**-2.640**)	(**-2.460**)	(**-2.330**)	(**-2.250**)	(**-3.060**)	(**-3.040**)	(**-3.040**)	(**-2.960**)	(**-3.070**)	(**-3.050**)
cay	Lag 1	38.961		67.468	67.891	53.177	54.106	52.273	53.976	54.667	51.866
		(1.780)		(**4.160**)	(**4.190**)	(**3.020**)	(**3.060**)	(**2.940**)	(**3.060**)	(**3.080**)	(**2.910**)
Macroeconomic Variables											
c	Lag 0	13.367					68.316				
		(0.100)					(0.710)				
y	Lag 0	-115.767						-21.146			
		(-1.480)						(-0.390)			
m1	Lag 0	32.871			51.569				46.809		
		(0.590)			(1.150)				(1.050)		
gdp	Lag 0	122.808								47.808	
		(1.370)								(0.800)	
unemp	Lag 0	4.995									4.885
		(0.440)									(0.520)
R^2		0.144	0.027	0.099	0.104	0.115	0.117	0.115	0.119	0.117	0.116
Adj. R^2		0.090	0.022	0.090	0.092	0.102	0.100	0.099	0.103	0.101	0.099

Table 9: Unexpected Flows and Predictive Variables

The table reports estimates from OLS regressions of unexpected flows on predictive variables. Flows are normalized by the lagged total stock market value. Unexpected flows are the difference between the actual flows and expected flows. Expected flows are the fitted values from Regression [9] in Table 4. The sample period is 1952Q4 to 2007Q4. See Table 1 for definitions of the predictive variables. For each regression, the first line reports the coefficient estimate and the second line reports *t*-statistics in parentheses. *t*-statistics significant at the 5% (10%) level are indicated in bold face (italics). Regressions [1], [5] and [7] use Newey-West heteroskedasticity and autocorrelation consistent standard errors, but the remaining regressions use OLS standard errors because specification tests indicated that the errors were homoskedastic and not autocorrelated.

Independent Variables		*Dependent Variable: Unexpected Normalized Flows*							
		[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
Intercept		2.330	2.001	1.837	1.629	1.771	1.806	1.842	1.858
		(**2.431**)	(**2.390**)	(**2.310**)	(**2.190**)	(**2.550**)	(**2.490**)	(**2.560**)	(**2.570**)
d-y	Lag 1	0.011							
		(0.221)							
d-e	Lag 1	0.058	0.023	0.019	0.031	0.043	0.034		
		(1.206)	(0.470)	(0.390)	(0.670)	(1.175)	(0.800)		
b-m	Lag 1	-0.048							
		(-1.650)							
term	Lag 1	-0.011	-0.005						
		(-1.172)	(-0.640)						
def	Lag 1	0.009	-0.016	-0.019					
		(0.385)	(-0.600)	(-0.740)					
rrel	Lag 1	-0.012	-0.013	-0.011	-0.008				
		(-1.189)	(-1.280)	(-1.120)	(-0.890)				
eqis	Lag 1	0.004	-0.048	-0.034	-0.068	-0.086		-0.066	
		(0.037)	(-0.450)	(-0.330)	(-0.740)	(-1.117)		(-0.903)	
cay	Lag 1	0.806	0.692	0.639	0.566	0.613	0.633	0.650	0.660
		(**2.267**)	(**2.340**)	(**2.260**)	(**2.130**)	(**2.519**)	(**2.440**)	(**2.561**)	(**2.570**)
R^2		0.059	0.044	0.042	0.040	0.036	0.032	0.032	0.029
Adj. R^2		0.023	0.017	0.020	0.022	0.023	0.023	0.023	0.025

Table 10: Unexpected Flows and Macroeconomic Variables

The table reports estimates from OLS regressions of unexpected flows on innovations of macroeconomic variables. Flows are normalized by the lagged total stock market value. Unexpected flows are the difference between the actual flows and expected flows. Expected flows are the fitted values from Regression [9] in Table 4. The sample period is 1952Q4 to 2007Q4. See Table 1 for definitions of the macroeconomic variables and for the method used to obtain the innovations. For each regression, the first line reports the coefficient estimate and the second line reports *t*-statistics in parentheses. *t*-statistics significant at the 5% (10%) level are indicated in bold face (italics).All regressions use OLS standard errors to calculate the *t*-statistics because specification tests indicated that the errors were homoskedastic and not autocorrelated.

Independent Variables		*Dependent Variable: Unexpected Normalized Flows*				
		[1]	[2]	[3]	[4]	[5]
Intercept		0.000	0.000	0.000	0.000	0.000
		(0.000)	(0.000)	(0.010)	(-0.010)	(0.000)
c	Lag 0	-0.564				
		(-0.270)				
y	Lag 0	-2.249	-1.642	-0.911		-2.196
		(-1.840)	(-1.710)	(-1.050)		(**-2.080**)
m1	Lag 0	1.582	1.389		0.798	1.541
		(1.880)	(1.740)		(1.110)	(*1.920*)
gdp	Lag 0	0.419				
		(0.300)				
unemp	Lag 0	-0.191				-0.208
		(-1.080)				(-1.280)
R^2		0.027	0.019	0.005	0.006	0.026
Adj. R^2		0.004	0.010	0.000	0.001	0.013

Table 11: Unexpected Flows, Predictive Variables and Macroeconomic Variables

The table reports estimates from OLS regressions of unexpected flows on predictive variables and innovations of macroeconomic variables. Flows are normalized by the lagged total stock market value. Unexpected flows are the difference between the actual flows and expected flows. Expected flows are the fitted values from Regression [9] in Table 4. The sample period is 1952Q4 to 2007Q4. See Table 1 for definitions of the variables. For each regression, the first line reports the coefficient estimate and the second line reports *t*-statistics in parentheses. *t*-statistics significant at the 5% (10%) level are indicated in bold face (italics).Regression [1], [4] and [5] use Newey-West heteroskedasticity and autocorrelation consistent standard errors to calculate the *t*-statistics. The remaining regressions use OLS standard errors to calculate the *t*-statistics because specification tests indicated that the errors were homoskedastic and not autocorrelated.

Independent Variables		*Dependent Variable: Unexpected Normalized Flows*				
		[1]	[2]	[3]	[4]	[5]
Intercept		2.175	1.858	1.875	2.076	1.922
		(**2.438**)	(**2.570**)	(**2.600**)	(**2.873**)	(**2.791**)
Predictive Variables						
d-y	Lag 1	0.004				
		(0.081)				
d-e	Lag 1	0.053				
		(1.172)				
b-m	Lag 1	-0.042			-0.025	-0.026
		(-1.379)			(-1.605)	(-1.649)
term	Lag 1	-0.009				
		(-0.921)				
def	Lag 1	0.009				
		(0.418)				
rrel	Lag 1	-0.011				
		(-1.035)				
eqis	Lag 1	0.016				
		(0.150)				
cay	Lag 1	0.761	0.660	0.666	0.744	0.689
		(**2.291**)	(**2.570**)	(**2.600**)	(**2.891**)	(**2.812**)
Macroeconomic Variables						
c	Lag 0	-0.683				
		(-0.473)				
y	Lag 0	-1.840				-1.352
		(-1.364)				(-1.423)
m1	Lag 0	1.101		0.835	0.811	1.294
		(1.143)		(1.170)	(1.127)	(1.508)
gdp	Lag 0	0.427				
		(0.376)				
unemp	Lag 0	-0.229				
		(-1.784)				
R^2		0.077	0.029	0.036	0.046	0.055
Adj. R^2		0.019	0.025	0.027	0.033	0.037

Table 12: Returns, Unexpected Flows, Predictive Variables and Macroeconomic Variables

The table reports estimates from OLS regressions of returns (including dividends) to the CRSP NYSE/NASDAQ/AMEX Index on unexpected flows, predictive variables and innovations of macroeconomic variables. The sample period is 1952Q4 to 2007Q4. Flows are normalized by the lagged total stock market value. Unexpected flows are the difference between the actual flows and expected flows. Expected flows are the fitted values from Regression [9] in Table 4. See Table 1 for definitions of other variables. For each regression, the first line reports the coefficient estimate and the second line reports *t*-statistics in parentheses. *t*-statistics significant at the 5% (10%) level are indicated in bold face (italics).With the exception of Regression [3] which uses Newey-West heteroskedasticity and autocorrelation consistent standard errors, all regressions use OLS standard errors to calculate the *t*-statistics because specification tests indicated that the errors were homoskedastic and not autocorrelated.

Independent Variables		*Dependent Variable: CRSP VW Returns*						
		[1]	[2]	[3]	[4]	[5]	[6]	[7]
Intercept		2.100	82.809	4.416	156.713	124.530	123.511	126.121
		(**4.060**)	(1.470)	(**4.310**)	(**3.580**)	(**2.740**)	(**2.690**)	(**2.770**)
Unflow	Lag 0	23.401	20.785	22.853	20.477	20.953	20.921	20.731
		(**5.680**	(**5.020**)	(**4.011**)	(**5.060**)	(**5.220**)	(**5.200**)	(**5.150**)
<u>Predictive Variables</u>								
d-y	Lag 1		4.300			3.624	3.620	3.561
			(1.550)			(**2.370**)	(**2.360**)	(**2.320**)
d-e	Lag 1		-2.495					
			(-0.780)					
b-m	Lag 1		0.041					
			(0.020)					
term	Lag 1		0.286					
			(0.540)					
def	Lag 1		0.302					
			(0.170)					
rrel	Lag 1		-0.599					
			(-0.930)					
eqis	Lag 1		-19.457	-12.378	-11.600	-19.951	-19.926	-19.590
			(**-2.840**)	(**-2.327**)	(**-2.200**)	(**-3.160**)	(**-3.150**)	(**-3.090**)
cay	Lag 1		23.134		54.155	37.776	37.421	38.442
			(1.100)		(**3.480**)	(**2.240**)	(**2.190**)	(**2.270**)
<u>Macroeconomic Variables</u>								
c	Lag 0		27.560					
			(0.220)					
y	Lag 0		-77.530				-8.843	
			(-1.040)				(-0.170)	
m1	Lag 0		9.985					29.479
			(0.190)					(0.690)
gdp	Lag 0		113.929					
			(1.340)					
unemp	Lag 0		9.762					
			(0.900)					
R^2		0.128	0.237	0.149	0.194	0.214	0.214	0.216
Adj. R^2		0.124	0.185	0.141	0.183	0.200	0.196	0.198

Table 13: Geweke's Measure of Linear Dependence

The table reports results from implementing Geweke's (1982, 1984) measure of linear dependence. The measure decomposes the linear dependence between two vector of time series, say X and Y, into three parts: (i) Contemporaneous Dependence, (ii) Feedback from Y to X, and (iii) Feedback from X to Y. "ret" denotes returns (including dividends) to the CRSP NYSE/NASDAQ/AMEX Index and "nflow" denotes the dollar amount of stocks bought by mutual funds in quarter *t* divided by the total value of the stock market at the end of quarter *(t-1)*. The sample period is 1951Q4 to 2007Q4. Panel A reports results for the unconditional version of the tests. Test statistics are based on the estimates of the variance-covariance matrix of residuals from a combination of VARs (5.1)-(5.4) in the text and are Chi-Square distributed under the null hypothesis. Panel B shows the results of the conditional version of the test. The conditioning variables are d/y, *eqis*, and *cay* and are selected based on results in earlier sections. These tests identical to the unconditional one except that the lags of the conditioning variables appear on the right hand side of VARs (5.1)-(5.4) along with the existing variables.

Panel A: Unconditional Decomposition of Geweke (1982)

	Chi-Square Test Statistic	p-value	df
Total Linear Dependence	32.4551	**0.0000**	3
Contemporaneous Dependence	30.4261	**0.0000**	1
$nflow_t \leftarrow ret_{t-1}$	1.86259	0.1723	1
$ret_t \leftarrow nflow_{t-1}$	0.16647	0.6833	1

	Chi-Square Test Statistic	p-value	df
Total Linear Dependence	44.5542	**0.0000**	9
Contemporaneous Dependence	28.9159	**0.0000**	1
$nflow_t \leftarrow ret_{t-i}$, i=1,…4	13.8293	**0.0079**	4
$ret_t \leftarrow nflow_{t-i}$, i=1,…4	1.80905	0.77083	4

Panel B: Conditional Decomposition of Geweke (1984)

	Chi-Square Test Statistic	p-value	df
Total Linear Dependence	32.7450	**0.0000**	3
Contemporaneous Dependence	26.2121	**0.0000**	1
$nflow_t \leftarrow ret_{t-1}$	5.1219	**0.0236**	1
$ret_t \leftarrow nflow_{t-1}$	1.4110	0.2349	1

	Chi-Square Test Statistic	p-value	df
Total Linear Dependence	34.9416	**0.0000**	9
Contemporaneous Dependence	27.9635	**0.0000**	1
$nflow_t \leftarrow ret_{t-i}$, i=1,…4	4.17063	0.38341	4
$ret_t \leftarrow nflow_{t-i}$, i=1,…4	2.80741	0.59055	4

New Evidence on Mutual Fund Managerial Risk-Shifting

Abstract

Existing literature documents that (i) an asymmetric flow-performance relationship creates an incentive for managers to extract rents from shareholders, and (ii) managers respond to such incentives by strategically altering portfolio risk. Using the semiparametric regression model proposed by Chevalier and Ellison (1997), we show that the flow-performance relationship has become linear in recent years (2000-2009) and fund managers no longer respond to such incentives. Fund managers, however, change portfolio risk in response to past performance; such changes have a positive impact on fund performance and are indicative of a better alignment of interests between managers and shareholders.

1. Introduction

Investors who invest in mutual funds receive shares that represent a contingent claim on the underlying portfolio. Their monies are invested in accordance with the fund's objective by the fund manager. While the manager *should* act in the best interest of the fund's investors, various factors might reduce his incentive to do so. For example, investor flows might respond more to positive past performance than to negative past performance (Sirri and Tufano (1998)), thus generating a call option type payoff for the fund manager, whose compensation is typically a fraction of assets under management. A segment of the mutual fund literature, beginning with Chevalier and Ellison (1997), documents that managers who have performed poorly in the first half of the year increase the risk of their portfolio in the second half of the year to take advantage of the asymmetry in investor flows. Thus, this manager-investor relationship is no exception to the agency problems proposed by Jensen and Meckling (1976).

We make three contributions to the mutual fund literature. First, we document that the flow-performance relationship is linear for the time period 2000-2009 (see Figure 2-4). This result holds for funds with age between 2-5 years (Young funds), for funds with age between 6-10 years (Medium funds), and for funds with age more than 11 years (Old funds). However, the behavior of investor flows varies across fund categories. In particular, investors in Medium and Old funds invest their monies only if these funds beat the benchmark by more than 11%. In contrast, investors invest in Young funds as long as the underperformance is less than 2%.

Second, we document that the strategic managerial-behavior mentioned above has changed in recent years. Based on the arguments above, if managers act in their self interest (as opposed to that of shareholders) we would expect a positive relation between portfolio risk changes and expected incentives/managerial benefits arising from the flow-performance relationship. In contrast, we find that managers no longer respond to these incentives. Amongst the fund categories, managers of Young funds subsequently decrease risk whereas managers of Medium and Old funds do not change portfolio risk in response to these incentives. In addition, these risk changes do not affect fund performance.

Finally, we show that managers change portfolio risk in response to past performance and that these risk changes benefit fund investors thereby indicating a better alignment of interests between managers and investors. The direction of risk changes varies with the level of funds' performance: managers of funds that have over/under performed the benchmark by a large (small) amount subsequently increase (decrease) risk. Amongst the fund categories, managers of Young funds respond to all levels of performance whereas managers of Medium and Old funds increase risk only if they have outperformed the benchmark by a large amount.

The rest of the paper is structured as follows. We begin by briefly discussing the literature on managerial risk-shifting (Section 2) and then describe the construction of our sample (Section 3). Section 4 discusses the semiparametric flow-performance relationship model of Chevalier and Ellison (1997) and the corresponding measure of risk-shifting incentive. We examine managerial risk-shifting and the impact on fund performance using this measure in Section 5. Section 6 repeats the analysis using returns and Section 7 concludes.

2. Literature Review

A fund manager's compensation is typically a fraction of assets under management (AUM). This motivates the manager to maximize the value of AUM and thus aligns the interests of managers and investors. However, the value of AUM at any point in time depends not only on the value of existing assets but also on the new money received by the fund. To maximize compensation the manager might focus on maximizing flows instead of maximizing the value of existing assets. If investor flows are more responsive to positive past performance than to negative past performance, a manager who has performed poorly (loser hereafter) in the first half of the year has an incentive to increase portfolio risk in the second half of the year. On the other hand, a manager who has performed well (winner hereafter) has an incentive to index. This fact was documented by Chevalier and Ellison (1997)[1,2]. The subsequent literature has developed along three segments.

The first segment has documented another factor–employment risk–that can cause the manager to alter the portfolio's risk (Chevalier and Ellison, 1999) and studied the impact of *both* compensation incentive and employment risk on managerial risk-shifting. Kempf, Ruenzi, and Thiele (2009) find that the behavior of losers depends on which incentive dominates at the end of the first half. When compensation incentive dominates, losers increase risk (relative to winners) in the second half since it offers a chance to catch up, and when employment risk dominates they decrease risk to prevent job loss. Hu, Kale, Pagani, and Subramanian (2009) argue that the interaction between compensation incentive and employment risk gives rise to a U-shaped relationship between performance and subsequent risk. For winners, employment risk is not a concern; hence, they increase risk to exploit the convexity in flow-performance relationship. For

---

[1] Specifically, they document that (i) the flow-performance relationship is non-linear and convex for young funds, and (ii) young loser funds increase risk whereas young winner funds index.

[2] A similar result has been documented by Brown, Harlow, and Starks (1996). In particular, the ratio of risk in the second half and risk in the first half is greater for losers than for winners: $(\sigma_{2nd}/\sigma_{1st})_{Losers} > (\sigma_{2nd}/\sigma_{1st})_{Winners}$. However, unlike Chevalier and Ellison (1997), Brown, Harlow, and Starks (1996) use contingency tables for their analysis and define risk as the standard deviation of monthly returns (as opposed to excess monthly returns). Recent papers based on Brown, Harlow, and Starks (1996) are Busse (2001), Goriaev, Nijman and Werker (2005), and Chen and Pennacchi (2009). Busse (2001) finds no evidence of tournaments when daily data is used to obtain volatility estimates, and argues that the autocorrelation of daily returns biases the volatility estimates based on monthly returns. Goriaev, Nijman and Werker (2005) show that monthly data are more robust to autocorrelation effects than are daily data. Chen and Pennacchi (2009) show that losers increase volatility of returns in excess of a benchmark and not volatility of returns.

losers, employment risk is a concern and increasing risk offers a possibility of not getting fired; hence, they also increase risk. Funds with intermediate performance, decrease risk, since they want to reduce employment risk.

The second segment examines the relation between risk-shifting and economic states. If the shape of the flow-performance relationship varies with economic activity, compensation incentive and therefore risk-shifting will also vary with economic activity. For a sample of no-load funds, Olivier and Tsay (2009) show that when economic activity is strong, the flow-performance relationship is convex and when activity is weak, the relationship is concave. Consequently, risk-shifting occurs only when economic activity is strong. Wang (2009) extends their finding to show that risk-shifting during good economic states contributes to the counter cyclical fund performance (Glode (2010) and Kosowski (2006)).

The third segment examines the mechanisms used for altering portfolio risk and the consequences of such behavior. Huang, Sialm, and Zhang (2009) find that funds that risk-shift perform poorly than funds that do not. In addition, the idiosyncratic risk channel is more important in explaining subsequent poor performance than other channels–changing systematic risk and switching between equity and cash.

3. Data

3.1 Sample Construction

The CRSP Mutual Fund database provides share class-level data for open-end mutual funds based in the U.S. The information is contained in three datasets depending on the frequency of the underlying variables. The first dataset (characteristic file hereafter) contains variables such as management fee, expense ratio, front-end load, back-end load, and 12b-1 fee at quarterly frequency starting in 1999. The second dataset (monthly file hereafter) contains net (of management fees and expenses but excluding loads) returns, total net assets, and per share net asset value at monthly frequency starting in 1990. The last dataset (daily file hereafter) contains net (of management fees and expenses but excluding loads) returns and per share net asset value at daily frequency starting in 1998.

We focus on equity mutual funds and restrict the sample period from January 2000 to December 2009. The characteristic file provides objective codes from vendors such as Lipper, Strategic Insight, and Wiesenberger for different time periods. For the time period under consideration, objective codes from only Lipper and Strategic Insight are available. Hence, we begin by extracting share classes with Strategic Insight objective codes 'AGG', 'GRI', 'GRO', 'ING', 'SCG', 'CPF', 'EPR', and 'GMC', and Lipper objective codes 'CA', 'EI', 'G', 'GI', 'MC', 'MR', and 'SG'. Next, we exclude (i) funds with more than 30% in bonds, and (ii) funds with the words ‘Index’, ‘Sector’, ‘S&P’, ‘Global’, ‘International’, ‘Metal’, ‘Estate’, ‘Money’ in their names. The resulting panel of share classes (share class–y–q) has 261797 observations and 14306 unique share classes from January 2000 to December 2009. We then remove funds that have assets less than $15 million and funds that are less than 2 years old, thus resulting in 206544 observations (11374 unique share classes). Since the flow-performance regressions use lagged variables, we extract data for the year 1999 for the unique share classes that exist in our sample in the year 2000. This increases the panel of share classes to 216910 but leaves the number of unique share classes unchanged (sample characteristic file hereafter). We now merge the sample characteristic file with the monthly file to obtain monthly data for all unique share classes for only the quarters that they exist in our sample. The resulting file (sample monthly file hereafter) has 611552 observations (11374 unique share classes). Finally, we merge the sample characteristic file with the daily file to obtain daily data for each share class for only the quarters that it exists in our sample. The resulting file (sample daily file hereafter) has 12175808 observations (11374 unique share classes). Thus, we now have three files–sample characteristic, monthly, and daily–that will be used to calculate the variables needed in the empirical analysis. In each of the following sections, we discuss the methodology used to aggregate share class-level data to the fund-level. This aggregation is required since many funds have multiple share classes (Nanda,Wang, and Zheng (2009)) and the empirical analysis is at the fund-level. Note that the sample files provide data at a higher frequency (quarterly, monthly or daily) than is required for the analysis, which is carried out at the annual frequency to maintain consistency with the literature. Consequently, we also describe the methods used to calculate annual values for the required variables.

3.2 Sample characteristic file

3.2.1 Aggregation to fund-level: The file consists of share class-level and fund-level variables that are reported at a quarterly frequency. We convert share class-level variables to fund-level variables by taking a TNA-weighted sum across all share classes of a fund in the current quarter. These variables include expense ratio (ER), front-end load (*FEL*), back-end load (*BEL*), management fee (*MFee*), and 12b-1 fee (*Fee12b1*). For fund-level variables such as turnover (*Turn*), percentage of the portfolio invested in domestic common and preferred stock (*Stock*), and percentage of the portfolio invested in cash (*Cash*) we retain only one observation for each fund. We calculate two additional variables from the information provided in this file. First, the age of each share class (*Age*) is calculated as the number of years since its First-Offer-Date. The corresponding fund-level variable is the age of the oldest share class. Second, the number of share classes for a fund (*NShc*) is the number of share classes that have the same fund name.

3.2.2 Classification of Funds: We follow Chevalier and Ellison (1997) by defining funds with ages between 2 and 5 years as Young funds. However, we classify the remaining funds into two categories as opposed to one: funds with ages between 6 and 10 years as Medium funds and funds with age more than 10 years as Old funds. We do so because the number of observations is too large and we run out of memory while estimating the kernel regressions (see *Section 4.1.1*).

3.3 Sample monthly file

3.3.1 Excess Returns (*r*): The file contains share class-level returns that are net of expenses and management fees. To obtain monthly return for a fund, we take a TNA-weighted sum across all share classes of a fund. This fund-level monthly return is then compounded to obtain annual return for year *t*. The excess return during year *t* is the difference between the compounded fund-level return and compounded monthly returns on the CRSP NYSE/NASDAQ/AMEX Index. The variable is defined in a similar manner for time periods shorter than one year–6 months.

3.3.2 Risk-Adjusted Returns: We use different factor models to obtain risk-adjusted returns. In particular, the *monthly* risk-adjusted returns for fund *i* are obtained as ($\alpha_i + e_{i,m}$), where α_i and $e_{i,m}$ are obtained from one of the following time-series regressions:

$$(R_{i,m} - R^f_m) = \alpha_i + \beta_{i,RMRF} RMRF_m + e_{i,m}$$

$$(R_{i,m} - R^f_m) = \alpha_i + \beta_{i,RMRF} RMRF_m + \beta_{i,SMB} SMB_m + \beta_{i,HML} HML_m + e_{i,m}$$

$$(R_{i,m} - R^f_m) = \alpha_i + \beta_{i,RMRF} RMRF_m + \beta_{i,SMB} SMB_m + \beta_{i,HML} HML_m + \beta_{i,UMD} UMD_m + e_{i,m}$$

In these equations, $R_{i,m}$ is the monthly return for fund *i* during month *m* obtained in 3.*3.1* above, R^f_m is the return on a 1-month T-Bill. RMRF is the difference between the return on a value-weighted CRSP stock portfolio and 1-month T-Bill yield, SMB is the return on the mimicking portfolio for the common size factor in stock returns, HML is the return on the mimicking portfolio for the common book-to-market equity factor in stock returns, UMD is the return on a factor-mimicking portfolio for one-year return momentum, and $\beta_{i,.}$ is the factor loading of the corresponding factor for fund *i*. SMB and HML are constructed according to the descriptions in Fama and French (1993) and Carhart (1997) contains details on the construction of the UMD factor. The factor data are available from Ken French's website. The annual risk-adjusted returns are then obtained by (i) compounding the variation in ($R_{i,m}$ - R^{f}_{m}) attributable to the risk-factor over the months of year *t*, and (ii) then subtracting it from the excess (of the risk free rate) compounded fund return.

3.3.3 Standard Deviation of Monthly Returns (σ): The monthly returns for a fund in year *t* obtained in *Section 3.3.1* above are used to calculate the sample standard deviation. The variable is defined in a similar manner for time periods shorter than one year–6 months.

3.3.4 Investor Flows: Since the file does not report the new money received by a share class/fund, we follow the literature (Sirri and Tufano (1998) and Zheng (1999)) by inferring this information from the total net assets and return data for that share class. To that end, we first calculate the fund-level total net assets (*TNA*) by adding the total net assets of all share classes of a fund in the current quarter. Next, we calculate the dollar amount of new money received by a fund in month *m* as the difference between assets of a fund at the end of month *m* and the appreciated value of its assets at the end of month *m-1*: $TNA_m - TNA_{m-1} \cdot (1 + r_{m-1,m})$, where $r_{m-1,m}$ is the monthly fund-level return obtained in *Section 3.3.1* above[3]. The new money

[3]The new money received by a fund during a time period is, by definition, a flow variable. However, the above definition, which is typical in the literature, suggests that the new money is received/invested at the end of the time period.

received by a fund during year *t* (*NewM*) is the sum of new money received by the fund in each month of year *t*. Finally, we calculate monthly fund-level flows, which express the dollar amount of money as a fraction of assets, by dividing new money received in month *m* by assets at the end of month *m-1*. The flows received by a fund during year *t* (*Flow*) are the sum of new money received by the fund in each month of year *t* divided by the average assets of the fund during year *t*. These variables (annual fund-level new money and flows) are defined in a similar manner for shorter time periods such as 6 months.

3.3.5 Number of Shares Outstanding (*NShares*): The share class-level number of shares outstanding are obtained by dividing the share class-level total net assets by the share class-level (per share) net asset value. The fund-level variable (*NShares*) is obtained by adding corresponding share class-level variable across a fund's share classes in the current quarter.

3.4 Sample daily file

We use the daily file to calculate the change in total risk ($\Delta Risk$) of a fund during year *t*, defined as the difference between the total risk of a fund in the first half of year *t* and that in the second half of year *t*. The total risk of a fund in the first/second half of a year is defined as the sample standard deviation of weekly excess (of CRSP NYSE/NASDAQ/AMEX Index) fund returns in that half. Since the daily file contains share class-level returns that are net of expenses and management fees, we obtain daily return for a fund by taking a TNA-weighted sum across all share classes of a fund. Next, we compound these to obtain weekly fund returns.

Finally, each of the fund-level variables obtained in Sections 3*.2.1*, 3*.3*, and 3*.4* above are winsorized at the 1st and 99th percentile to reduce the impact of outliers in the empirical analysis. After all the above variables have been calculated, we merge the resulting files into a one file that is used in the subsequent analysis. The unit of observation in this file is “fund” and the frequency is ‘annual’; it has 13905 fund-year observations, 2030 of which are for Young Funds, 5218 of which are for Medium Funds, and 6657 of which are for the Old Funds. The large reduction in the number of observations, relative to the numbers mentioned earlier, is due to (i) conversion to the fund-level, and (ii) converting from the quarterly/monthly frequency to the annual frequency.

3.5 Descriptive Statistics

3.5.1 Background

Table 1 contains definitions of all the variables. Table 2 reports fund-level descriptive statistics for all funds (All Funds), for funds with ages between 2 to 5 years (Young Funds), for funds with ages between 6 to 10 years (Medium Funds), and for funds with age more than 11 years (Old Funds). Panel A reports various fund characteristics. Since the variables in this panel are available at a higher frequency (monthly or quarterly), they are averaged over each year before computing the statistics below. The exception is *Age*, which is calculated at the end of each year. Panel B reports annual investor flows. Panel C reports annual returns in excess of CRSP NYSE/NASDAQ/AMEX Index and annual risk-adjusted returns calculated using 1-, 3-, and 4-factor models as described in *Section 3.3.2*. Panel D reports risk measures.

The columns labeled “AV”, “MD” and “SD” report the sample mean, median and standard deviation for the variable appearing in the column “Variable”. The column labeled “Tests of Differences” tests (for the variable in the column “Variable”) whether the difference between the two types of funds is statistically significant. Under this column, the column labeled “AV” tests the null hypothesis that the difference between the means of the two types of funds is zero, against a 2-sided alternative. Under the null hypothesis of equal means, the test statistic is $\frac{\overline{X_1}-\overline{X_2}}{se}$, where $\overline{X_1}$ and $\overline{X_2}$ are the sample means, and *se* is calculated based on a test of equality of variances for the two types of funds. The column labeled "MD" reports results from the Mann-Whitney-Wilcoxon Test to test the null hypothesis that the distributions for the two types of funds are identical against a 2-sided alternative. For brevity, only *p*-values significant at the 5% level are shown; a "." indicates an insignificant *p*-value.

3.5.2 Discussion of Descriptive Statistics

The average fund in our sample (column labeled All Funds) held assets worth $1455 million, had 69 million shares outstanding and was 15 years old. The fund invested 95% of its assets in stocks and 4% in cash, and traded 87% of its assets each year. Consistent with the trend documented by Nanda, Wang, and Zheng (2009), the median fund has 3 share classes. Investors in the fund incurred total expenses of 1.25% each year, 0.74% of which was paid to the fund manager, 0.20% was attributable to marketing & distribution efforts to attract new investors, and the remainder was attributable to administrative and other expenses. Existing investors incurred back-end load of 0.65% if they withdrew their funds, whereas new investors incurred front-end loads of 1.35%. As a group, existing and new investors withdrew $7.8 million or 3.11% of the fund's assets from the fund each year. Consistent with existing evidence on fund performance, the average fund in our sample does not provide superior performance on a risk-adjusted basis.

We now discuss the differences between the three categories of funds. On average, young funds in our sample held assets worth $396 million and 11 million shares outstanding, whereas medium and old funds are much larger. Each fund category has multiple share classes, with Old funds having more share classes than Medium and Young funds. Young funds trade 107% of their assets each year, followed by Medium funds which trade a significantly lower 89.28%, and by Old funds which trade the lowest amount–79.58%. On average, Young funds invest 94.93% of its portfolio in domestic common and preferred stock. Medium and Old funds invest a significantly higher amount in equities–approximately 95%. Young funds hold the highest amount of cash (4.19%) and both Medium and Old funds hold a significantly lower amount of 3.54% and 3.68% respectively. Investors in Young funds incur the highest expense ratio of 1.36% and those in Medium funds incur a significantly lower expense ratio of 1.28%. Investors in Old funds incur the lowest expense ratio of 1.19%. The management fees component of expense ratio is highest for Medium funds and lowest for Old funds. Compared to investors in an Old fund, investors in a Young/Medium fund incur lower loads while investing their monies but incur higher loads while withdrawing their monies. Investor flows differ across the fund categories. A typical Young fund received *inflows* of $2.19 million each year, whereas a typical fund in the Medium and Old category experienced *outflows* of $5.85 and $21 million respectively. When expressed as a percentage of the fund's assets, the flows to each of the three types of funds show a similar trend. None of the fund categories provide superior risk-adjusted performance; however, Old funds have the best inferior risk-adjusted performance of 0.71% per year.

This discussion suggests that Young funds are small, have the highest operating costs; investors in such funds incur the lowest front-end but highest back-end loads. Compared to Medium and Old funds, Young funds hold more cash and less stock but have higher turnover. For the time period under consideration, investors invested monies into Young funds whereas they withdrew monies from Medium and Old funds.

4. Estimating the Risk-Shifting Incentive

To calculate the risk-shifting incentive arising from a fund's flow-performance relationship we use the measure developed in Chevalier and Ellison (1997). They begin with the following semiparametric model for the flow-performance relationship:

$$Flow_{i,t} = g(r_{i,t-1}) + \beta_1 r_{i,t} + \beta_2 r_{i,t-2} + \beta_3 r_{i,t-3} + \beta_4 IndustryGrowth_{i,t} + \beta_4 ln\left(\frac{TNA_{i,t-1}}{\bar{A}_{t-1}}\right) + e_{i,t} \quad \ldots(4.1)$$

where $r_{i,t-1}$ is the excess (of CRSP NYSE/AMEX/NASDAQ/ Index) return for fund *i* during year *t-1*, $Flow_{i,t}$ is the investors' flow to fund *i* during year *t*, $IndustryGrowth_{i,t}$ is growth in the assets of the equity mutual fund industry during year *t*, and $ln\left(\frac{TNA_{i,t-1}}{\bar{A}_{t-1}}\right)$ is the log of total net assets of fund *i* at the end of year *t-1* scaled by the average total net assets of funds for which the equation is being estimated. For compactness, Equation 4.1 can be written as $Flow_{i,t} = g(r_{i,t-1}) + \mathbf{Z}_i'\boldsymbol{\beta} + e_{i,t}$, where $\mathbf{Z}$ is a vector representing the parametric component. If the manager does not respond to the incentives generated from the flow-performance relationship, the expected (at the end of the first half of year *t-1)* flows for year *t*, are

$$\mathrm{E}_{t-1,1st}(\mathrm{Flow}_{i,t}^{noRS}) = \mathrm{E}_{t-1,1st}(g(r_{i,t-1,1st} + u_{i,t-1,2nd}) + \mathbf{Z}_i'\boldsymbol{\beta}) \quad \ldots(4.2)$$

In this equation, $r_{i,t-1,1st}$ is the excess return for fund *i* during the first half of year *t-1*. $u_{i,t-1,2nd}$ is a normally distributed random variable with zero mean and standard deviation $\sigma_{i,t-1,1st}$ and represents the fund's excess return during the second half of year *t-1*. When the manager responds to the incentives and strategically alters portfolio risk, excess returns ($v_{i,t-1,2nd}$) for the second half of year *t-1* are generated from a distribution with zero mean and standard deviation $\sigma_{i,t-1,1st} + \Delta \cdot \sigma_{i,t-1,1st}$ and the expected flows can be written as

$$\mathrm{E}_{\mathrm{t-1,1st}}\left(\mathrm{Flow}_{\mathrm{i,t}}^{\mathrm{RS}}\right) = \mathrm{E}_{\mathrm{t-1,1st}}\left(g\left(r_{i,t-1,\mathrm{1st}} + \mathrm{v}_{\mathrm{i,t-1,2nd}}\right) + \mathbf{Z}_{\mathrm{i}}'\boldsymbol{\beta}\right) \qquad \ldots(4.3)$$

The difference between Equation 4.3 and 4.2, represents the benefit to the fund manager from responding to such incentives. i.e.

$$RSI_{i,t-1,\mathrm{1st}} = \mathrm{E}_{\mathrm{t-1,1st}}\left(\mathrm{Flow}_{\mathrm{i,t}}^{\mathrm{RS}}\right) - \mathrm{E}_{\mathrm{t-1,1st}}\left(\mathrm{Flow}_{\mathrm{i,t}}^{\mathrm{noRS}}\right) \qquad \ldots(4.4)$$

$$= \mathrm{E}_{\mathrm{t-1,1st}}\left[g\left(r_{i,t-1,\mathrm{1st}} + \mathrm{v}_{\mathrm{i,t-1,2nd}}\right) - g\left(r_{i,t-1,\mathrm{1st}} + \mathrm{u}_{\mathrm{i,t-1,2nd}}\right)\right] \ldots(4.5)$$

Thus, $RSI_{i,t-1,\mathrm{1st}}$ represents the incremental expected flows during year *t* if the manager alters portfolio risk during the second half of year *t-1*. The manager's compensation is typically linked to assets under management, which in turn depend on flows and value of existing assets. Thus, the measure links volatility of fund returns to the manager's compensation.

To implement this measure, we need the (i) volatility of fund returns for the first half, and (ii) estimates of the term in square brackets in Equation 4.5. We obtain (i) above by using the sample standard deviation of weekly excess fund returns from January to June, and (ii) above by separately estimating $g(r_{i,t-1,\mathrm{1st}} + \mathrm{v}_{\mathrm{i,t-1,2nd}})$ and $g(r_{i,t-1,\mathrm{1st}} + \mathrm{u}_{\mathrm{i,t-1,2nd}})$ via kernel regressions that use $(Flow_{i,t} - \mathbf{Z}_{\mathrm{i}}'\widehat{\boldsymbol{\beta}})$ and $r_{i,t-1}$. *Section 4.1* discusses (ii) above in greater detail. *Section 4.2* tests whether the flow-performance relationship is linear and *Section 4.3* presents descriptive statistics for $RSI_{i,t-1,\mathrm{1st}}$.

4.1 Estimating the flow-performance relationship

4.1.1 Methodology

To estimate Equation 4.1, Chevalier and Ellison (1997) follow Robinson (1988) which develops a procedure to obtain $\sqrt{n}$-consistent estimates of the parametric component. The first step involves estimating the parametric component. Taking expectations with respect to $r_{i,t-1}$ in

Equation 4.1, and subtracting the resulting equation from Equation 4.1, gives the following equation which does not have the nonparametric component:

$$Flow_{i,t} - \mathrm{E}(Flow_{i,t}|r_{i,t-1}) = (\mathbf{Z}_{\mathrm{i}} - \mathrm{E}(\mathbf{Z}_{\mathrm{i}}|r_{i,t-1}))' \ \boldsymbol{\beta} + e_{i,t} \qquad \ldots(4.6)$$

$\mathrm{E}(Flow_{i,t}|r_{i,t-1})$ is estimated by a kernel regression of $Flow_{i,t}$ on $r_{i,t-1}$; similarly, $\mathrm{E}(\mathbf{Z}_{\mathrm{i}}|r_{i,t-1})$ is estimated by a kernel regression of each element of $\mathbf{Z}_{\mathrm{i}}$ on $r_{i,t-1}$. After replacing $\mathrm{E}(\cdot\,|r_{i,t-1})$ by their estimates in Equation 4.6, an OLS regression is estimated to obtain $\widehat{\boldsymbol{\beta}}$. The nonparametric component $g(.)$ can now be estimated by a kernel regression of $(Flow_{i,t} - \mathbf{Z}'_{\mathrm{i}}\widehat{\boldsymbol{\beta}})$ on $r_{i,t-1}$. All kernel regressions use the Nadaraya-Watson estimator with the Epanechnikov kernel and the bandwidths are chosen using the Least Squares Cross Validation method.

We estimate the Equation 4.1 separately for each fund category since the behavior of flows differs across Young funds and Medium/Old funds (see Panel B in Table 2). In the OLS regression discussed above, the residuals (for a particular fund) may be correlated over time[4], thereby causing standard errors to be biased (Petersen (2009)). To correct for these correlations, we cluster the standard errors by fund.

4.1.2 Discussion of Results

Estimates of the parametric component of Equation 4.1 for the sample period 2000-2009 are reported in Table 3. Panel A, B and C report results for Young, Medium and Old funds respectively. The coefficient on each return component is significant and the magnitude decreases with the lag. For example, flows to Young funds in year *t* increase by 0.56% in response to an additional 1% excess return in the current year *t* whereas they increase by only 0.18% in response to an additional 1% excess return in the year *t-2*. The coefficient on the

[4] We assume that the correlation (at a point in time) across residuals of funds is accounted for by the *IndustryGrowth* variable

IndustryGrowth variable is insignificant for Young funds but significant for Medium and Old funds. Chen, Hong, Huang, and Kubik (2004) find that larger funds have lower returns than smaller funds; thus larger funds might receive lower flows. Consistent with this reasoning, the coefficient on $ln\left(\frac{TNA_{i,t-1}}{\bar{\mathrm{A}}_{\mathrm{t-1}}}\right)$ is negative and significant at the 10% level for Young funds. The coefficient for Medium and Old funds is insignificant.

Estimates of the nonparametric component $g(.)$ obtained from a kernel regression of $(Flow_{i,t} - \mathbf{Z}_{\mathrm{i}}'\widehat{\boldsymbol{\beta}})$ on $r_{i,t-1}$ are plotted in Figures 1 to 3 for each fund category and the R^2 is reported in Table 3. These plots show the expected flows in year *t* in response to performance in year *t-1*, after controlling for the effect of variables in vector **Z**. The relation appears smooth for Medium and Old funds but appears wavy for Young funds and is indicative of oversmoothing–too small bandwidth. Using twice the optimum bandwidth for Young funds resulted in a smooth curve (Figure 4). The relation is very similar for Medium and Old funds–investors invest their monies only if these funds beat the benchmark by more than 11% and existing withdraw their monies if overperformance is less than 11%. In contrast, existing investors in Young funds withdraw their monies only if the underperformance is more than 2%.

The natural question is whether these flow-performance relationships are non-linear. Figures 2, 3 and 4 suggest that these relationships are linear. The next section describes and implements a statistical test developed by Härdle and Mammen (1993) and confirms the inference obtained from the plots.

4.2 Tests for Linearity of flow-performance relationship

Chevalier and Ellison (1997) use the specification test developed in Ellison and Ellison (2000) to test departures from linearity. Miles and Mora (2003) compare the performance of several nonparametric specification tests using simulations and find that the test of Härdle and Mammen (1993) performs similar to that of Ellison and Ellison (2000). Specifically, for different choice of bandwidths the size (row labeled c=0 in Table 1) ranges from 0.05 to 0.064 for the test of Härdle and Mammen (1993) and from 0.031 to 0.082 for the test of Ellison and Ellison (2000). Similarly, the power (rows with c≠0 in Table 1) ranges from 0.159 to 0.993 for the test of Härdle and Mammen (1993) and from 0.109 to 0.995 for the test of Ellison and Ellison (2000). Hence we use the test of Härdle and Mammen (1993) to test departures from linearity.

The test of Härdle and Mammen (1993) tests the null hypothesis $E(Flow1_{i,t}|r_{i,t-1}) = \gamma_0 + \gamma_1 r_{i,t-1}$, where $Flow1_{i,t} \equiv (Flow_{i,t} - \mathbf{Z}_{\mathrm{i}}'\widehat{\boldsymbol{\beta}})$ versus a nonparametric alternative. The test statistic $T = \sqrt{h} \cdot \sum_{i,t}\left(\left\{\hat{g}_{NP}(r_{i,t-1}) - \hat{g}_P(r_{i,t-1})\right\}^2 \cdot w(r_{i,t-1})\right)$, where $\hat{g}_{NP}(\cdot)$ and $\hat{g}_P(\cdot)$ are the nonparametric and parametric estimators of $g(\cdot)$ respectively, $w(\cdot)$ is a weighting function and *h* is the bandwidth. To account for the different convergence rates of the parametric and nonparametric estimators, Härdle and Mammen (1993) replace $\hat{g}_P(\cdot)$ with a kernel-smoothed parametric estimator. Thus the test examines whether (weighted squared) deviations between a parametric and nonparametric model are large enough. Following their recommendation, the critical values for *T* are obtained using the wild bootstrap procedure with 1000 replications.

These results are reported in Table 3 in the row labeled "Tests for Linearity Test". Regardless for the fund category, the test statistic is less than the 95% critical value. Thus, we fail to reject the null hypothesis of a parametric model for the flow-performance relationship for each fund category[5].

4.3 Descriptive Statistics for RSI

[5] Using data from 2000-2008, Kim (2010) finds that the relationship is concave on the positive side and linear on the negative side.

Table 4 presents various statistics for the *RSI* obtained by estimating Equation 4.1 separately for Young, Medium and Old funds. The column labeled "All Funds" reports statistics for the pooled sample of estimates of *RSI*. For the columns labeled "All Funds", "Young Funds", "Medium Funds", and "Old Funds", the number in the first row reports the statistic appearing in the "Statistic" column. Under the "Statistic" column, the second row, if present, reports the *p*-values (in parentheses) to test the null hypothesis that the statistic is equal to zero. Under the null hypothesis of zero mean, the test statistic is $\frac{\bar{X}}{SD/\sqrt{N}}$, where $\bar{X}$ is the sample mean, SD is the sample standard deviation and N is the number of (non-missing) observations, and follows a *T*-distribution with *N*-1 degrees of freedom. The test for the median is based on the Wilcoxon Signed Rank Test.

For "All Funds", while the median *RSI* is statistically indifferent from zero (*p*-value 0.912), the mean *RSI* is -0.0161 and significant (*p*-value 0.002). Both Young and Medium funds show a similar pattern. For Old funds, however, both the mean and median *RSI* are insignificant. Thus, in contrast to earlier time periods, the expected benefits for each fund category are not positive i.e. fund managers cannot expect to gain additional flows by altering risk. The column labeled "Tests of Differences", tests whether the difference in Mean/Median *RSI* between any two funds types is significant. The results indicate that Young fund managers stand to lose the most if they increase the volatility of their portfolio, followed by Medium fund managers, and Old fund managers.

5. Tests Using the Risk-Shifting Incentive

The previous section discussed and presented descriptive statistics for an ex-ante measure of benefits arising from a fund's flow-performance relationship to its manager. This section examines whether managers change the risk of their portfolio in response to such incentives (Section 5.1 and Table 5), and the impact of such changes on the fund's performance (Section 5.2 and Table 6).

5.1 Managerial Response to Incentives

5.1.1 Regression Specification

To examine whether managers alter the risk of their portfolio in response to incentives, we regress changes in (total) portfolio risk on the risk-shifting incentive and a set of control variables, which account for other factors that can cause portfolio risk to change. In particular, we estimate the following model:

$$\begin{aligned}\Delta Risk_{i,t-1} = {} & \gamma_1 RSI^{+}_{i,t-1,1st} + \gamma_2 RSI^{-}_{i,t-1,1st} + \gamma_3 RSI^{+}_{i,t-1,1st} \cdot D^{-}_{i,t-1,1st} \\ & + \gamma_4 RSI^{-}_{i,t-1,1st} \cdot D^{+}_{i,t-1,1st} + \gamma_5 RSI^{+}_{i,t-1,1st} \cdot lnTNA_{i,t-1,1st} \\ & + \gamma_6\, RSI^{-}_{i,t-1,1st} \cdot lnTNA_{i,t-1,1st} + \boldsymbol{\gamma' Controls}_{i,t-1,1st} + v_{i,t-1} \qquad \ldots (5.1)\end{aligned}$$

The justification for the choice of this specification is as follows. The discussion in *Section 2* suggests that managers that have a positive *RSI* will increase risk, whereas managers that have a negative *RSI* will decrease risk. Hence, we split *RSI* into its positive ($RSI^{+}_{i,t-1} \equiv \max(0, RSI_{i,t-1})$) and negative component ($RSI^{-}_{i,t-1} \equiv \min(0, RSI_{i,t-1})$). Further, funds within each category (positive/negative *RSI*) might respond differently based on their performance during the first half of year *t-1*. For example, funds that have underperformed the CRSP Index by a small amount might subsequently increase risk, but those that have underperformed the CRSP

Index by a large amount might not change risk. To allow for this possibility, we interact RSI^{+} (RSI^{-}) with a dummy variable $D^{-}_{t-1,1st}$ ($D^{+}_{t-1,1st}$) that takes value 1 if the fund's excess return in the first half of year *t-1* is below (above) the median negative (positive) excess returns of funds for which the equation is being estimated. In addition, the ease with which the manager can alter portfolio risk will depend on the size of the fund. For example, the manager of a small fund and positive *RSI* might increase risk, whereas the manager of a large fund with the same *RSI* might not respond to the incentive, since he incurs greater transaction costs due to the fund's large positions. To model this, we interact RSI^{+} and RSI^{-} with the natural logarithm of the average size of the fund during the first half of year *t-1* ($lnTNA_{t-1,1st}$). Finally, to control for other factors that can cause changes in portfolio risk, we include a set of control variables ($\boldsymbol{Controls}_{i,t-1,1st}$) that includes the fund's (i) existing risk ($\text{Risk}_{t\text{-}1,1st}$) to account for mean reversion in portfolio risk, and (ii) turnover ($\text{Turn}_{t\text{-}1,1st}$) to account for the fact that managers that trade more often will have higher changes in risk.

Thus, this specification distinguishes between the response of managers that face positive and negative incentives (γ_1 & γ_2). Within each of these groups, it allows the response to vary based on the (i) fund's performance in the first half of the year (γ_3 for RSI^{+} and γ_4 for RSI^{-}) and (ii) size of the fund (γ_5 for RSI^{+} and γ_6 for RSI^{-}).

5.1.2 Discussion of Results

The results of estimating Equation 5.1 are reported in Table 5. Since this is a panel dataset, the residuals (for a particular fund) may be correlated over time and/or residuals (at a point in time) may be correlated across funds, thereby causing OLS standard errors to be biased. To correct for these correlations, we need to cluster the standard errors by fund and year. However, two-way clustering works only if there is large number of clusters in *each* dimension (see Figure 7 on pg. 460 in Petersen (2009)). Consequently, we use time dummies to account for cross-sectional dependence and cluster the standard errors by fund.

We begin by discussing results for all funds (Panel A). Regression [1] includes the control variables mentioned above, the RSI and an interaction term with *lnTNA*. The coefficient on $Risk_{t-1,1st}$ is negative and highly significant suggesting mean reversion in portfolio risk and the coefficient on $Turn_{t-1,1st}$ is positive and significant suggesting that funds with higher turnover have higher changes in risk. More importantly, the coefficient on RSI is negative but *insignificant* suggesting that managers do not respond to such incentives. Since managers respond differently to positive and negative incentives, Regression [2] splits RSI into its components, RSI^{+} *and* RSI^{-}, along with interaction terms with *lnTNA*. One would expect the coefficient on RSI^{+} and RSI^{-} to be positive; in contrast, the coefficients are negative and statistically insignificant. Similarly, the coefficients on the interaction terms with *lnTNA* are insignificant; thus managers of both small and large funds do not respond to such incentives. The last specification extends Regression [2] by distinguishing between managers that are above and those that are below the median positive/negative performance at the end of the first half of the year. The coefficient on each of the interacted dummy variables $D^{-}_{t-1,1st}$ and $D^{+}_{t-1,1st}$ is negative and insignificant. This implies that both type of funds–funds that underperform (beat) the benchmark by a large amount and funds that underperform (beat) the benchmark by a small amount–do not respond to such incentives.

Panel B reports results for Young funds. Similar to that in Panel A, the coefficient on the control variables $Risk_{t-1,1st}$ and $Turn_{t-1,1st}$ is significant. However, the coefficient on RSI is *negative* and significant (*t*-statistic of 1.97) suggesting that managers decrease risk in response to such incentives. Regression [2] splits RSI into positive and negative components. While the coefficient on each component is negative (-0.07 and -0.045 respectively), it is insignificant. Similarly, the interaction terms with *lnTNA* are insignificant. Next, Regression [3] controls for performance in the first half of the year. The coefficient on RSI^{+}, RSI^{-}, and the interaction terms with *lnTNA* do not change much and they are still insignificant. The interaction terms with the dummy variables are also insignificant.

Panel C and D report results for Medium and Old funds respectively. The results in this panel are identical to those in Panel A. The coefficients on the control variables are significant and those on *RSI* and the corresponding interaction terms are insignificant. Note that the R^2 within each panel is very similar across the regression specifications. In addition, the R^2s for Medium and Old funds are similar to those for "All Funds".

Overall, these results indicate that managers of Medium and Old funds do not respond to incentives. Managers of Young funds, on the other hand, change portfolio risk in response to incentives, but they do so in a manner inconsistent with the risk-shifting hypothesis.

5.2 Impact of Managerial Response on Fund Performance

The results of the previous section indicate that managers of Young funds decrease risk when they stand to gain. This behavior is opposite to that predicted by the risk-shifting hypothesis. Could it be that these fund managers alter portfolio risk in a manner that benefits shareholders? To investigate this possibility, we examine the impact of risk-changes arising from *RSI* on fund performance during the second half of the year. In particular, we estimate the following panel regression:

$$r_{i,t-1,2nd} = \theta_1 \cdot \widehat{\Delta Risk}^{RSI}_{i,t-1} + \theta_2 \cdot \widehat{\Delta Risk}^{Other}_{i,t-1} + \boldsymbol{\theta}' \cdot \boldsymbol{Controls}_{i,t-1} + u_{i,t-1} \qquad \ldots(5.2)$$

In this expression, $r_{i,t-1,2nd}$ is the excess (of CRSP Index) return for fund *i* during the second half of year *t-1*, $\widehat{\Delta Risk}_{i,t-1}$ is the predicted value for $\Delta Risk_{i,t-1}$ for fund *i* obtained from Model 1 in Table 5, $\widehat{\Delta Risk}^{RSI}_{i,t-1}$ is the predicted value for $\Delta Risk_{i,t-1}$ for fund *i* attributable to the risk-shifting incentive terms, $\widehat{\Delta Risk}^{Other}_{i,t-1}$ is the predicted value for $\Delta Risk_{i,t-1}$ for fund *i* attributable to all other terms (coefficients in the vector $\boldsymbol{\gamma}$ and time-fixed effects in Equation 5.1), and $\boldsymbol{Controls}_{i,t-1}$ includes variables that control for other factors that might affect returns. In particular, we include the following variables:

(i) Contemporaneous investor flows ($Flow_{i,t-1,2nd}$) since funds with higher returns receive higher flows,

(ii) Lagged investor flows ($Flow_{i,t-1,1st}$) since funds that received high flows in the first half will have lower returns due to decreasing returns to scale (Berk and Green (2004)),

(iii) Contemporaneous turnover ($\mathrm{Turn}_{i,t-1,2nd}$) since funds that trade more often incur greater transaction costs, which in turn lower returns,

(iv) Natural logarithm of average monthly TNA during the second half of year *t-1*($lnTNA_{i,t-1,2nd}$) since larger funds have lower returns (Chen, Hong, Huang, and Kubik (2004)), and

(v) Contemporaneous expense ratio ($\mathrm{ER}_{i,t-1,2nd}$) since funds with higher expense ratio might have lower returns.

Since managers of only Young funds respond to incentives, investigating the impact of risk changes is meaningful for only these funds. However for completeness, Table 7 presents results from estimating Equation 5.2 above for each of the fund categories–separately and combined. The method used to estimate this equation is identical to that for Equation 5.1 (see *Section 5.1.2* for details).

We begin by discussing results for all funds (Panel A). Regression [1] begins by including only $\Delta Risk_{i,t-1}$ and all of the control variables mentioned above, and indicates an insignificant negative contemporaneous relationship between returns and changes in risk–the coefficient on $\Delta Risk_{i,t-1}$is negative (-0.154) and insignificant (*t*-statistic -0.569). In addition, each of the control variables is significant and the sign of the coefficient is consistent with the reasoning described above. Regression [2] replaces $\Delta Risk_{i,t-1}$by its predicted value ($\widehat{\Delta Risk}_{i,t-1}$) obtained from Regression [1] in Panel A of Table 5 and retains all control variables. The coefficient on $\widehat{\Delta Risk}_{i,t-1}$ is positive with a *t*-statistic of 0.275 and the R^2 remains unchanged. Each of the control variables is still significant and has the same sign. Finally, Regression [3] decomposes the predicted risk changes into the two parts mentioned above. The coefficient on $\widehat{\Delta Risk}_{i,t-1}^{RSI}$ is positive (value of 31.93) but insignificant (*t*-statistic 1.63).

Panels B reports results for Young funds. Regression [1] is similar to that in Panel A. However, there is a positive contemporaneous relationship between $r_{i,t-1,2nd}$ and $\widehat{\Delta Risk}_{i,t-1}$. Decomposing $\widehat{\Delta Risk}_{i,t-1}$into $\widehat{\Delta Risk}_{i,t-1}^{RSI}$ and $\widehat{\Delta Risk}_{i,t-1}^{Other}$ reveals that risk changes made in response to *RSI* have no impact on returns (*t*-statistic 1.31). Similar to Panel A, all control variables are significant; the exception is *ER*. Panel C shows results for Medium funds and these are similar to those in Panel A. In addition, the R^2 are very similar to those obtained for "All Funds".

Panel D reports results for Old funds. Similar to previous panels, Regression [1] indicates an insignificant contemporaneous relationship between returns and (predicted) changes in portfolio risk during the second half of the year. However, Regression [2] indicates a negative and significant relationship between returns and predicted value of risk changes (*t*-statistic of -2.33). Regression [3], which decomposes the predicted risk changes into components arising from the *RSI* terms and other terms in Equation 5.1, indicates that risk changes due to the *RSI* do not affect returns–the coefficient on $\widehat{\Delta Risk}^{RSI}_{i,t-1}$ is positive (value of 20.13) but insignificant (*t*-statistic 0.564). Each of the control variables is significant and the signs are identical to those in previous panels.

Thus, the results of this sub-section indicate that (regardless of the fund category) changes in portfolio risk that are attributable to the risk-shifting incentive have no impact on returns.

6. Tests Using Returns

The previous section examined managerial response and its impact on fund performance using an ex-ante measure of incentives. This section examines the same issues using actual incentives–realized excess returns during the first half of the year, and is therefore structured similar to Section 5. We first examine managerial response (Section 6.1 and Table 7) and then investigate the impact of such changes on fund performance (Section 6.2 and Table 8).

6.1 Managerial Response to Returns

6.1.1 Regression Specification

To investigate whether managers change portfolio risk in response to returns, we estimate the following model:

$$\begin{aligned}\Delta Risk_{i,t-1} = {} & \gamma_1 r^{++}_{i,t-1,1st} + \gamma_2 r^{+}_{i,t-1,1st} + \gamma_3 r^{-}_{i,t-1,1st} + \gamma_4 r^{--}_{i,t-1,1st} + \gamma_5 (r^{++}_{i,t-1,1st} \cdot lnTNA_{i,t-1,1st}) \\ & + \gamma_6 (r^{+}_{i,t-1,1st} \cdot lnTNA_{i,t-1,1st}) + \gamma_7 (r^{-}_{i,t-1,1st} \cdot lnTNA_{i,t-1,1st}) \\ & + \gamma_8 (r^{--}_{i,t-1,1st} \cdot lnTNA_{i,t-1,1st}) + \boldsymbol{\gamma' Controls}_{i,t-1,1st} + v_{i,t-1} \qquad \ldots(6.1)\end{aligned}$$

This model is similar to Equation 5.1, except that the *RSI* terms have been replaced by a piece-wise linear specification for excess returns ($r_{i,t-1,1st}$) during the first half of year *t-1*. The return components are obtained from the following expressions:

$$r_{i,t-1,1st}^{--} = \min[(r_{i,t-1,1st} + b_{0,t-1,1st}),0],$$

$$r_{i,t-1,1st}^{-} = \max\left[-b_{0,t-1,1st}, \min\left(0, r_{i,t-1,1st}\right)\right],$$

$$r_{i,t-1,1st}^{+} = \min\left[b_{1,t-1,1st}, max\left(0, r_{i,t-1,1st}\right)\right], \text{ and}$$

$$r_{i,t-1,1st}^{++} = \max[(r_{i,t-1,1st} - b_{1,t-1,1st}),0]$$

In these expressions, $b_{0,t-1,1st}$ and $b_{1,t-1,1st}$ are positive constants that equal the median negative and positive excess return of funds in the first half of year *t-1* for which Equation 6.1 is being estimated. The variables in $\boldsymbol{Controls}$ vector are the same (See Section 5.1.1 for further details).

6.1.2 Discussion of Results

Panel A reports results for all fund types. Similar to Panel A in Table 5, $\text{Risk}_{t-1,1st}$ is negatively related to $\Delta Risk_{t-1}$ and $\text{Turn}_{t-1,1st}$ is positively related to $\Delta Risk_{t-1}$. All the return components are significant. For example, the coefficient on $r_{t-1,1st}^{++}$ is positive (value of 0.012) and significant (*t*-statistic 8.066). Regression [2] interacts each return component with $lnTNA_{t-1,1st}$. This reduces the *t*-statistic for $r_{t-1,1st}^{++}$ but it is still significant. All other return components become insignificant. Similarly, each of the interacted return components is insignificant. In addition, the R^2 does not change much.

Panel B reports results for Young funds and the results are similar to those in Panel A. The only difference is that interacting the return components with $lnTNA$ makes *all* return components insignificant. Panel C and D show results for Medium and Old funds respectively. In Regression [1], the coefficient on $r_{t-1,1st}^{++}$ is positive and significant and other return components are insignificant. Introducing the interaction terms with $lnTNA$ makes all the terms involving the return components insignificant (Regression [2]). The exception is Old funds that have outperformed the benchmark by a small amount. For such funds, larger are the fund's assets lower are the changes in risk.

These results indicate that funds change risk in response to performance in the first half of the year. However, this effect is more predominant in Young funds than in Medium and Old funds. Young funds that have underperformed or overperformed the benchmark by a large (small) amount subsequently increase (decrease) risk. Amongst the Medium and Old funds, funds that have overperformed the benchmark by a large amount subsequently increase risk.

6.2 Impact of Managerial Response on Fund Performance

To investigate the impact of the risk changes documented above on fund performance, we estimate the following model:

$$r_{i,t-1,2nd} = \theta_1 \cdot \widehat{\Delta Risk}_{i,t-1}^{r_{1st}} + \theta_2 \cdot \widehat{\Delta Risk}_{i,t-1}^{Other} + \boldsymbol{\theta'} \cdot \boldsymbol{Controls}_{i,t-1} + u_{i,t-1} \quad \ldots (6.2)$$

This model is identical to that in *Section 5.2* except that Regression [1] in Table 7 is used to obtain the predicted value of risk changes attributable to the returns ($\widehat{\Delta Risk}_{t-1}^{r_{1st}}$) and those attributable to other components ($\widehat{\Delta Risk}_{t-1}^{Other}$). The choice of variables in $\boldsymbol{Controls}_{t-1}$ and estimation method is identical to that used earlier (see Section 5.2 and Section 5.1.2 for further details).

Table 8 reports results from estimating the above equation. Regression [1] in Panel A indicates a positive but insignificant contemporaneous relationship between $r_{t-1,2nd}$ and $\Delta Risk_{t-1}$. Each of the control variables is significant and the signs of the coefficients are identical to that in Table 6. In the next regression, $\Delta Risk_{t-1}$ is replaced by its predicted value

($\Delta\widehat{Risk}_{t-1}$). It is significant (*t*-statistic 2.50) and the coefficient is positive. The last regression splits the predicted risk changes into the two parts mentioned above. The risk-changes made in response to returns during the first half of the year ($\Delta\widehat{Risk}_{t-1}^{r_{1st}}$) are positively related (value 28.68 and *t*-statistic 8.696) to returns during the second half of the year. Panel B, C and D show results for Young, Medium, and Old funds respectively and these are similar to those in Panel A.

7. Conclusion

The empirical mutual fund literature has documented two stylized facts: (i) the flow-performance is asymmetric in that investor flows respond more to positive past performance than to negative past performance, and (ii) fund managers act in their self-interest by exploiting this asymmetry via altering portfolio risk. This paper shows that the flow-performance relation has become linear for the time period 2000-2009 for Young, Medium and Old funds, and that, on average, managers do not respond to the incentives arising from this relationship. Managers of Young funds respond to the incentives but the risk changes have no impact on fund performance; managers of Medium and Old funds do not respond to such incentives. While managers do not respond to these incentives, they change risk based on past performance. These risk changes have a positive impact on fund performance and thus benefit fund investors.

REFERENCES

Barber, Brad M., Odean, T., Zheng, L., 2005. Out of Sight, Out of Mind: The Effects of Expenses on Mutual Fund Flows. The Journal of Business 78, 2095-2120.

Berk, Jonathan B., Green, Richard C., 2004. Mutual Fund Flows and Performance in Rational Markets. Journal of Political Economy 112, 1269-1295.

Brown, K.C., Harlow, W.V., Starks, L.T., 1996. Of Tournaments and Temptations:An Analysis of Managerial Incentives in the Mutual Fund Industry. The Journal of Finance 51, 85-110.

Busse, J.A., 2001. Another Look at Mutual Fund Tournaments. Journal of Financial and Quantitative Analysis 36, 53-73.

Carhart, M.M., 1997. On Persistence in Mutual Fund Performance. The Journal of Finance 52, 57-82.

Chen, H.-l., Pennacchi, G.G., 2009. Does Prior Performance Affect a Mutual Fund's Choice of Risk? Theory and Further Empirical Evidence. Journal of Financial and Quantitative Analysis 44, 745-775.

Chen, J., Hong, H., Huang, M., Kubik, J.D., 2004. Does Fund Size Erode Mutual Fund Performance? The Role of Liquidity and Organization. American Economic Review 94, 1276-1302.

Chevalier, J., Ellison, G., 1997. Risk Taking by Mutual Funds as a Response to Incentives. Journal of Political Economy 105, 1167-1200.

Chevalier, J., Ellison, G., 1999. Career Concerns of Mutual Fund Managers. Quarterly Journal of Economics 114, 389-432.

Ellison, G., Ellison, S.F., 2000. A simple framework for nonparametric specification testing. Journal of Econometrics 96, 1-23.

Fama, E.F., French, K.R., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33, 3-56.

Glode, V., 2010. Why Mutual Funds 'Underperform'. Journal of Financial Economics (JFE), Forthcoming.

Goriaev, A., Nijman, T.E., Werker, B.J.M., 2005. Yet Another Look at Mutual Fund Tournaments. Journal of Empirical Finance 12, 127-137.

Guerico, D.D., Tkac, P.A., 2002. The Determinants of the Flow of Funds of Managed Portfolios: Mutual Funds Vs Pension Funds. Journal of Financial and Quantitative Analysis 37, 523-557.

Härdle, W., Mammen, E., 1993. Testing parametric versus nonparametric regression. Annals of Statistics 21, 1296-1947.

Hu, P., Kale, J., Pagani, M., Subramanian, A., 2009. Fund Flows, Performance, Managerial Concerns, and Risk-Taking. San Jose State University and Georgia State University.

Huang, J., Sialm, C., Zhang, H., 2009. Risk Shifting and Mutual Fund Performance. McCombs School of Business, University of Texas at Austin, Texas, Austin.

Jacques, O., Tay, A., 2009. Time-Varying Incentives in the Mutual Fund Industry. HEC Paris and Singapore Management University.

Jensen, M.C., Meckling, W.H., 1976. Theory of the firm: Managerial behavior, agency costs and ownership structure. Journal of Financial Economics 3, 305-360.

Kempf, A., Ruenzi, S., Thiele, T., 2009. Employment risk, compensation incentives, and managerial risk taking: Evidence from the mutual fund industry. Journal of Financial Economics 92, 92-108.

Kim, M.S., 2010. Changes in Mutual Fund Flows and Managerial Incentives. University of Southern California.

Kosowski, R., 2006. Do Mutual Funds Perform When it Matters Most to Investors? US Mutual Fund Performance and Risk in Recessions and Expansions. SSRN eLibrary.

Miles, D., Mora, J., 2003. On the performance of nonparametric specification tests in regression models. Computational Statistics & Data Analysis 42, 477-490.

Nanda, V.K., Wang, Z.J., Zheng, L., 2009. The ABCs of mutual funds: On the introduction of multiple share classes. Journal of Financial Intermediation 18, 329-361.

Petersen, M.A., 2009. Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches. Review of Financial Studies 22, 435-480.

Robinson, P.M., 1988. Root-N-Consistent Semiparametric Regression. Econometrica 56, 931-954.

Sirri, E.R., Tufano, P., 1998. Costly Search and Mutual Fund Flows. The Journal of Finance 53, 1589-1622.

Wang, X., 2009. On Time Varying Mutual Fund Performance. Rotman School of Management, University of Toronto.

Zheng, L., 1999. Is Money Smart? A Study of Mutual Fund Investors' Fund Selection Ability. The Journal of Finance 54, 901-933.

Table 1: Definitions of Variables

Unless specified otherwise, all variables are available at the share-class level, their definitions are obtained from the CRSP Mutual Fund Database Guide, and are converted to the fund-level by calculating a TNA-weighted sum across all share classes of a fund. In the "Symbol" column, the subscript *i* denotes a fund, and *t* denotes a year. Depending on the regression, a third subscript (1^{st} or 2^{nd}) is used to denote the first 6 months or last 6 months of year *t.* In the "Variable" column, the first letter in the square brackets indicates the type (Flow/Level) of variable, the second letter indicates the frequency (monthly or quarterly) of the variable available from the data source for the sample period 2000-2009, and the third symbol denotes the units (Percentage, Millions of US Dollars, Millions, or Years) of the variable. A "%" denotes Percentage, "$Mn" denotes Millions of US Dollars, "Mn" denotes "Millions", "I" denotes "Integer", and "." denotes "Not Applicable"

SYMBOL	VARIABLE	DEFINITION
Fund Characteristics		
$ER_{i,t}$	Expense Ratio [F, Q, %]	Dollar amount of Operating Expenses including 12b-1 fees divided by the dollar amount of total assets for the most recently completed fiscal year.
$FEL_{i,t}$	Front-end Load [L, Q, %]	Maximum sales charge incurred by the investor when money is invested in the fund
$BEL_{i,t}$	Back-end Load [L, Q, %]	Maximum fees incurred by the investor when money is withdrawn from the fund.
$MFee_{i,t}$	Management Fee [F, Q, %]	Dollar amount of management fees divided by average net assets
$Fee12b1_{i,t}$	Actual 12b-1 Fee [L, Q, %]	Dollar amount of marketing and distribution costs divided by total assets for the most recently completed fiscal year.
$TNA_{i,t}$	Total Net Assets [L, M, $Mn]	Total Net Assets at the end of the time period. The fund-level variable is calculated by adding the share-class level variable across all share classes of a fund.
$NShares_{i,t}$	Number of Shares Outstanding [L, . , Mn]	A derived variable that is obtained by dividing Total Net Assets by the Net Asset Value (per share). The fund-level variable is calculated by adding the share-class level variable across all share classes of a fund.
$Age_{i,t}$	Age [L, . , Years]	A derived variable that is calculated at the end of year *t* as the number of years since the First-Offer-Date. The fund-level variable is calculated by taking the maximum of the share-class level variable across all share classes of a fund.
$Nshc_{i,t}$	Number of Share Classes [L, . , I]	A derived variable that is obtained by counting the number of share classes that have the same fund name
$Turn_{i,t}$	Turnover [F, Q, %]	A fund-level variable that is defined as the minimum of the total dollar amount of purchases and sales divided by the average total net assets of the fund over the past 12-months.
$Stock_{i,t}$	Stock [L, Q, %]	A fund-level variable that is defined as the percentage of the portfolio invested in domestic common and preferred stocks.
$Cash_{i,t}$	Cash [L, Q, %]	A fund-level variable that is defined as the percentage of the portfolio invested in cash.

Table 1: Definitions of Variables (contd.)

SYMBOL	VARIABLE	DEFINITION
Investor Flows		
$NewM_{i,t}$	NewMoney [F, . , $Mn]	A derived fund-level variable that is defined as the dollar amount of money received by the fund during year *t*. It is obtained by adding the money received by the fund in each month of year *t*. Following Zheng (1999), the money received by the fund each month is calculated as the difference between total net assets of a fund at time *m* and the appreciated value of its assets at time *m-1*: $NewM_{i,m} = [TNA_{i,m} - TNA_{i,m-1} \cdot (1+r_{m,t})]$, where $r_{m,t}$ is obtained as described in "*Returns*" below. The variable is defined in a similar manner for time period shorter than one year–6 months.
$Flow_{i,t}$	Flow [F, . , %]	A derived fund-level variable that expresses the dollar amount of money received by the fund during year *t* as a fraction of total net assets. It is defined as the total dollar amount of money received by the fund during year *t* divided by the average total net assets during year *t*. The variable is defined in a similar manner for time period shorter than one year–6 months
Returns		
$r_{i,t}$	Excess Return [F, M , %]	Net return to fund *i* during year *t* minus return on the CRSP NYSE/AMEX/NASDAQ Index during the same time interval. Monthly net returns are compounded to obtain returns for year *t*. Monthly net returns are obtained taking a TNA-weighted sum across all share classes of a fund. The variable is defined in a similar manner for time period shorter than one year–6 months
Risk Measures		
$\sigma_{i,t}$	Volatility [F, . ,%]	Sample standard deviation of monthly fund returns. Monthly fund returns are obtained by taking a TNA-weighted sum across all share classes of a fund. The variable is defined in a similar manner for time period shorter than one year–6 months.
$\Delta Risk_{i,t}$	Change in Risk [F, . ,%]	Difference between the total risk of a fund in the second half of year *t* and that in the first half of year *t*. The total risk of a fund in the first/second half of a year is the sample standard deviation of excess (of CRSP NYSE/AMEX/NASDAQ Index) weekly fund returns in that half. Daily fund returns for a fund are obtained by taking a TNA-weighted sum across all share classes of a fund. The daily returns are compounded to obtain weekly returns
$Risk_{t,1st}$	Risk in First Half [F, . , %]	Sample standard deviation of weekly excess fund returns during the first six months of year *t*. Weekly fund returns are obtained as described in "$\Delta Risk_{i,t}$"

Table 2: Descriptive Statistics

The table reports fund-level descriptive statistics for all funds (All Funds), funds with ages between 2-5 years (Young Funds), funds with ages between 6-10 years (Medium Funds) and funds with ages more than 11 years (Old Funds). The data are obtained from the CRSP Mutual Fund Database and the sample period is from 2000-2009. See **Table 1** for definitions of the variables. **Panel A** reports various fund characteristics. Since the variables in this panel are available at a higher frequency (monthly or quarterly), they are averaged over each year before computing the statistics below. The exception is *Age*, which is calculated at the end of each year. **Panel B** reports annual investor flows, **Panel C** reports annual returns in excess of the CRSP NYSE/AMEX/NASDAQ Index and annual risk-adjusted returns calculated using 1-,3-, and 4-factor models as described in *Section 3.3.2* of the text, and **Panel D** reports risk measures. The columns labeled "AV", "MD" and "SD" report the sample mean, median and standard deviation for the variable appearing in the column "Variable". In Panels B, C, and D, for the columns labeled "All Funds", "Young Funds", "Medium Funds", and "Old Funds", the second row reports the *p*-values to test the null hypothesis that the statistic is equal to zero. Under the null hypothesis of zero mean, the test statistic is $\frac{\bar{X}}{SD/\sqrt{N}}$, where $\bar{X}$ is the sample mean, SD is the sample standard deviation and N is the number of (non-missing) observations, and follows a *T*-distribution with N-1 degrees of freedom. The test for the median is based on the Wilcoxon Signed Rank Test. See *Base SAS 9.2 Procedures Guide: Statistical Procedures* for further details. The column labeled "Tests of Differences" tests (for the variable in the column "Variable") whether the difference between the two types of funds is statistically significant. Under this column, the column labeled "AV" tests the null hypothesis that the difference between the means of the two types of funds is zero, against a 2-sided alternative. Under the null hypothesis of equal means, the test statistic is $\frac{\overline{X_1}-\overline{X_2}}{se}$, where $\overline{X_1}$ and $\overline{X_2}$ are the sample means, and *se* is calculated based on a test of equality of variances for the two types of funds. If we fail to reject the null hypothesis of equal variances, then *se* = $\sqrt{s_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$, where $s_p^2 = \left(\frac{n_1-1}{n_1+n_2-2}\right)s_1^2 + \left(\frac{n_2-1}{n_1+n_2-2}\right)s_2^2$, n_1 and n_2 are the number of observations for the two types of funds, and the test statistic has a *T*-distribution with (n_1+n_2-2) degrees of freedom. In the other case, *se* = $\sqrt{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)}$, the test statistic follows a *T*-distribution, and the degrees of freedom are obtained using the Satterthwaite method. The column labeled "MD" reports results from the Mann-Whitney-Wilcoxon Test to test the null hypothesis that the distributions for the two types of funds are identical against a 2-sided alternative. See Chapter 62 in the *SAS/STAT 9.2 Users Guide* for further details. For these two columns, the first row reports the difference between the mean or median of the two types of funds and the second row reports the *p*-values. For brevity, only *p*-values significant at the 5% level are shown and a "." indicates insignificant *p*-values. To reduce the impact of outliers, all variables are winsorized at the 1st and 99th percentile.

Variable	All Funds			Young Funds			Medium Funds			Old Funds			Tests of Differences					
													(Young – Med.)		(Young – Old)		(Med. – Old)	
	AV	**MD**	**SD**	**AV**	**MD**	**SD**	**AV**	**MD**	**SD**	**AV**	**MD**	**SD**	**AV**	**MD**	**AV**	**MD**	**AV**	**MD**
Panel A: Characteristics																		
ER	1.25	1.23	0.40	1.36	1.34	0.47	1.28	1.25	0.41	1.19	1.18	0.36	0.08	0.09	0.17	0.16	0.09	0.07
													<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
FEL	1.35	0.11	1.77	1.09	0.02	1.54	1.13	0.04	1.57	1.60	0.27	1.94	-0.03	-0.01	-0.51	-0.25	-0.47	-0.24
													.	.	<.0001	<.0001	<.0001	<.0001
BEL	0.65	0.17	0.84	0.72	0.08	0.97	0.66	0.17	0.84	0.62	0.19	0.79	0.06	-0.10	0.10	-0.11	0.04	-0.02
													0.0231	.	<.0001	.	0.0041	.
MFee	0.74	0.75	0.26	0.75	0.77	0.32	0.77	0.78	0.28	0.72	0.72	0.23	-0.02	0.00	0.03	0.06	0.05	0.06
													0.0416	.	<.0001	<.0001	<.0001	<.0001
Ac12b1	0.20	0.10	0.23	0.23	0.09	0.28	0.19	0.08	0.23	0.19	0.11	0.21	0.04	0.01	0.04	-0.02	0.00	-0.03
													<.0001	0.0101	<.0001	0.0416	.	.
TNA	1455.07	325.55	3860.13	396.70	160.73	654.35	682.72	237.41	1239.29	2383.22	584.24	5302.96	-286.02	-76.68	-1986.52	-423.52	-1700.50	-346.83
													<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
Nshares	69.44	20.73	149.80	26.22	11.40	40.53	41.01	16.37	65.03	104.91	32.46	201.49	-14.79	-4.97	-78.69	-21.06	-63.90	-16.09
													<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
Age	14.89	10.00	13.71	4.28	5.00	0.96	7.82	8.00	1.39	23.68	17.00	15.50	-3.54	-3.00	-19.40	-12.00	-15.87	-9.00
													<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
Nshc	2.93	3.00	1.99	2.55	2.00	1.67	2.85	2.50	1.90	3.12	3.00	2.12	-0.30	-0.50	-0.57	-1.00	-0.27	-0.50
													<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
Turnover	87.33	69.75	74.30	107.76	79.25	108.23	89.28	72.00	70.95	79.58	65.25	61.86	18.48	7.25	28.18	14.00	9.70	6.75
													<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
Stock	95.34	97.00	5.40	94.93	96.63	5.62	95.42	97.08	5.40	95.39	97.01	5.34	-0.49	-0.45	-0.46	-0.38	0.03	0.07
													0.0023	<.0001	0.0028	0.0006	.	.
Cash	3.69	2.53	4.11	4.19	2.99	4.67	3.54	2.48	3.88	3.68	2.50	4.11	0.65	0.51	0.51	0.49	-0.14	-0.02
													<.0001	<.0001	<.0001	<.0001	.	.

Table 2: Descriptive Statistics (contd.)

Variable	All Funds			Young Funds			Medium Funds			Old Funds			Tests of Differences					
													(Young – Med.)		(Young – Old)		(Med. – Old)	
	AV	MD	SD	AV	MD	SD	AV	MD	SD	AV	MD	SD	AV	MD	AV	MD	AV	MD
Panel B: Investor Flows																		
NewM	-7.83	-8.88	480.93	62.48	2.19	225.60	24.51	-5.85	274.83	-54.63	-20.99	635.56	37.97	8.03	117.11	23.18	79.14	15.15
	0.05	<.0001		<.0001	<.0001		<.0001	<.0001		<.0001	<.0001		<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
Flow	-3.11	-5.87	30.01	7.56	2.50	36.48	-3.75	-5.06	32.44	-5.86	-7.97	24.66	11.31	7.56	13.42	10.47	2.11	2.91
	<.0001	<.0001		<.0001	<.0001		<.0001	<.0001		<.0001	<.0001		<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
Panel C: Returns																		
Excess Returns	0.16	-0.96	10.00	1.35	-0.21	12.90	-0.10	-1.14	9.60	0.00	-1.00	9.26	1.45	0.94	1.35	0.79	-0.10	-0.15
	.	<.0001		<.0001	.		.	<.0001		.	<.0001		<.0001	<.0001	<.0001	0.0004	.	.
1-Factor Alpha	-0.21	-1.04	8.48	0.66	-0.98	10.41	-0.63	-1.30	8.27	-0.13	-0.80	8.00	1.29	0.32	0.79	-0.18	-0.50	-0.50
	0.00	<.0001		0.01	.		<.0001	<.0001		.	<.0001		<.0001	0.0042	0.0025	.	0.0009	<.0001
3-Factor Alpha	-1.33	-1.38	6.59	-1.95	-1.88	7.60	-1.78	-1.66	6.48	-0.80	-1.06	6.31	-0.18	-0.22	-1.15	-0.82	-0.98	-0.60
	<.0001	<.0001		<.0001	<.0001		<.0001	<.0001		<.0001	<.0001		.	.	<.0001	<.0001	<.0001	<.0001
4-Factor Alpha	-1.25	-1.36	6.13	-1.91	-1.72	6.91	-1.70	-1.59	5.90	-0.71	-1.04	6.01	-0.21	-0.12	-1.20	-0.67	-0.99	-0.55
	<.0001	<.0001		<.0001	<.0001		<.0001	<.0001		<.0001	<.0001		.	.	<.0001	<.0001	<.0001	<.0001
Panel D: Risk Measures																		
σ	4.61	4.13	2.02	5.09	4.85	2.25	4.57	4.06	1.96	4.52	4.01	1.98	0.52	0.79	0.57	0.84	0.05	0.05
	<.0001	<.0001		<.0001	<.0001		<.0001	<.0001		<.0001	<.0001		<.0001	<.0001	<.0001	<.0001	.	.
ΔRisk	-0.03	-0.01	0.37	-0.14	-0.09	0.41	-0.01	0.00	0.37	-0.02	-0.01	0.36	-0.14	-0.10	-0.12	-0.09	0.02	0.01
	<.0001	<.0001		<.0001	<.0001		.	.		0.00	0.00		<.0001	<.0001	<.0001	<.0001	.	.

Table 3: Semiparametric Model for Flow-Performance Relationship

The table reports estimates for the parametric component of the following model of Chevalier and Ellison (1997):

$$Flow_{i,t} = g(r_{i,t-1}) + \beta_1 r_{i,t} + \beta_2 r_{i,t-2} + \beta_3 r_{i,t-3} + \beta_4 IndustryGrowth_{i,t} + \beta_5 ln\left(\frac{TNA_{i,t-1}}{\bar{\mathrm{A}}_{\mathrm{t-1}}}\right) + e_{i,t},$$

where $r_{i,t-1}$ is the excess (of CRSP NYSE/AMEX/NASDAQ Index) return for fund *i* during year *t-1*, $Flow_{i,t}$ is investors' flow to fund *i* during year *t*, $IndustryGrowth_{i,t}$ is growth in the assets of the equity mutual fund industry during year *t*, and $ln\left(\frac{TNA_{i,t-1}}{\bar{\mathrm{A}}_{\mathrm{t-1}}}\right)$ is the log of total net assets of fund *i* at the end of year *t-1* scaled by the average total net assets of funds for which the equation is being estimated. The data are obtained from the CRSP Mutual Fund Database, the sample period is from 2000-2009, and model is estimated separately for each fund category using the methodology of Robinson (1988). **Panel A**, **Panel B**, and **Panel C** reports estimates for funds with ages between 2-5 years (Young Funds), between 6-10 years (Medium Funds) and more than 11 years (Old Funds) at the end of year *t-1*. Estimates of the non-parametric component, $g(r_{i,t-1})$, for each fund category are plotted in **Figure 1-3**. For each regression specification, the first row reports the OLS coefficient estimate, and the second row reports the *t*-statistic in parentheses. Boldface (italicized) *t*-statistics indicate significance at the 5% (10%) level. Based on Petersen (2009), the standard errors are clustered by fund. The row labeled "*Tests for Linearity*" reports results from the specification test of Härdle and Mammen (1993). The test tests the null hypothesis $E(Flow_{i,t} - \mathbf{Z}_i'\hat{\boldsymbol{\beta}} | r_{i,t-1}) = \gamma_0 + \gamma_1 r_{i,t-1}$, against a nonparametric alternative. The test statistic (*T*) is given by $\sqrt{h} \cdot \sum_{i,t} \left(\{\hat{g}_{NP}(r_{i,t-1}) - \hat{g}_P(r_{i,t-1})\}^2 \cdot w(r_{i,t-1})\right)$, where $\hat{g}_{NP}(\cdot)$ and $\hat{g}_P(\cdot)$ are the nonparametric and parametric estimators of $g(\cdot)$ respectively, $w(\cdot)$ is a weighting function and *h* is the bandwidth. To account for the different convergence rates of the parametric and nonparametric estimators, Härdle and Mammen (1993) replace $\hat{g}_P(\cdot)$ with a kernel-smoothed parametric estimator. The critical values for *T* are obtained using the wild bootstrap procedure with 1000 replications. The row labeled "R^2" reports the R^2 (in decimals) from the kernel regression of $(Flow_{i,t} - \mathbf{Z}_i'\hat{\boldsymbol{\beta}})$ on $r_{i,t-1}$.The last row reports the number of observations in each fund category. To reduce the impact of outliers, all variables are winsorized at the 1st and 99th percentile.

Table 3: Semiparametric Model for Flow-Performance Relationship (contd.)

	Panel A: Young Funds	**Panel B: Medium Funds**	**Panel C: Old Funds**
Parametric Part			
$r_{i,t}$	0.556	0.637	0.509
	(**5.208**)	(**12.798**)	(**13.407**)
$r_{i,t-2}$	0.176	0.303	0.332
	(**3.477**)	(**9.381**)	(**13.407**)
$r_{i,t-3}$	0.166	0.179	0.283
	(**2.568**)	(**5.430**)	(**10.339**)
$IndustryGrowth_{i,t}$	0.008	0.034	0.041
	(0.249)	(**2.261**)	(**4.161**)
$ln(TNA_{i,t-1}/\bar{\mathrm{A}}_{\mathrm{t-1}})$	-1.240	-0.081	-0.181
	(-1.896)	(-0.242)	(-0.973)
Tests of Linearity			
Test Statistic	41316.359	21846.762	14707.074
Critical Value	46558.28	26835.66	18735.21
R^2	0.202	0.073	0.140
N	2030	5218	6657

Table 4: Descriptive Statistics for Risk-Shifting Incentive (RSI)

The table reports descriptive statistics for the Risk-Shifting Incentive of Chevalier and Ellison (1997) defined as

$$RSI_{i,t-1,1st} = \mathrm{E}_{t-1,1st}\left(\mathrm{Flow}_{i,t}^{RS}\right) - \mathrm{E}_{t-1,1st}\left(\mathrm{Flow}_{i,t}^{noRS}\right)$$

where $\mathrm{E}_{t-1,1st}\left(\mathrm{Flow}_{i,t}^{RS}\right)$ is the expected flows to fund *i* in year *t* if the manager increases the total portfolio risk in the second half of year *t-1* by 50% (relative to the first half) and $\mathrm{E}_{t-1,1st}\left(\mathrm{Flow}_{i,t}^{noRS}\right)$ is the expected flows to fund *i* in year *t* if the manager keeps the total portfolio risk in the second half of year *t-1* identical to that in the first half of year *t-1*. See Section 4 in the text for additional details. The column labeled "All Funds" reports statistics for the pooled sample of estimates of *RSI*. For the columns labeled "All Funds", "Young Funds", "Medium Funds", and "Old Funds", the number in the first row reports the statistic appearing in the "Statistic" column. Under the "Statistic" column, the second row, if present, reports the *p*-values (in parentheses) to test the null hypothesis that the statistic is equal to zero. Under the null hypothesis of zero mean, the test statistic is $\frac{\bar{X}}{SD/\sqrt{N}}$, where $\bar{X}$ is the sample mean, SD is the sample standard deviation and *N* is the number of (non-missing) observations, and follows a *T*-distribution with *N*-1 degrees of freedom. The test for the median is based on the Wilcoxon Signed Rank Test. See *Base SAS 9.2 Procedures Guide: Statistical Procedures* for further details. The column labeled "Tests of Differences" tests whether the difference between the Mean/Median *RSI* for the two types of funds is statistically significant. Under the null hypothesis of equal means, the test statistic is $\frac{\overline{X_1}-\overline{X_2}}{se}$, where $\overline{X_1}$ and $\overline{X_2}$ are the sample means, and *se* is calculated based on a test of equality of variances for the two types of funds. The test for medians is based on the Mann-Whitney-Wilcoxon Test, which tests the null hypothesis that the distributions for the two types of funds are identical against a 2-sided alternative. See Chapter 62 in the *SAS/STAT 9.2 Users Guide* for further details. For these three columns, the first row reports the difference between the mean or median of the two types of funds and the second row reports the *p*-values in parentheses. Boldface (italicized) *p*-values indicate significance at the 5% (10%) level. To reduce the impact of outliers, the variable is winsorized at the 1st and 99th percentile.

Statistic	All Funds	Young Funds	Medium Funds	Old Funds	Tests of Differences		
					(Young – Medium)	(Young – Old)	(Medium – Old)
Mean	-0.01609	-0.07031	-0.01527	-0.00019	-0.05504	-0.07012	-0.01508
	(**0.002**)	(**0.016**)	(**0.002**)	(0.971)	(**0.063**)	(**0.018**)	(**0.037**)
Median	0.000375	-0.001544	-0.001326	0.003199	-0.000219	-0.004743	-0.004524
	(0.192)	(0.635)	(0.048)	(0.915)	(0.542)	(0.849)	(0.176)
SD	0.6199	1.3056	0.3512	0.4305			
Skewness	-1.2493	-0.8137	-0.1934	0.1479			
Min	-5.3894	-5.3894	-1.1266	-1.3225			
Max	3.8149	3.8149	1.0127	1.4652			

Table 5: Managerial Response to Risk-Shifting Incentives

The table reports estimates from the following panel regression

$$\Delta Risk_{i,t-1} = \gamma_1 RSI^{+}_{i,t-1,1st} + \gamma_2 RSI^{-}_{i,t-1,1st} + \gamma_3 RSI^{+}_{i,t-1,1st} \cdot D^{-}_{i,t-1,1st} + \gamma_4 RSI^{-}_{i,t-1,1st} \cdot D^{+}_{i,t-1,1st} + \gamma_5 RSI^{+}_{i,t-1,1st} \cdot lnTNA_{i,t-1,1st} + \gamma_6 RSI^{-}_{i,t-1,1st} \cdot lnTNA_{i,t-1,1st} + \boldsymbol{\gamma}' \boldsymbol{Controls}_{i,t-1,1st} + v_{i,t-1},$$

where $RSI^{+}_{i,t-1,1st} = \max(0, RSI_{i,t-1,1st})$, $RSI^{-}_{i,t-1,1st} = \min(0, RSI_{i,t-1})$, $RSI_{i,t-1,1st}$ is the risk-shifting incentive for fund *i* obtained from Equation 4.5 in the text, $D^{-}_{i,t-1,1st}$ is a dummy variable that takes value 1 if the excess return for fund *i* in the first half of year *t-1* is below the median negative excess returns of funds for which the equation is being estimated, $D^{+}_{i,t-1,1st}$ is a dummy variable that takes value 1 if the excess return for fund *i* in the first half of year *t-1* is above the median positive excess returns of funds for which the equation is being estimated, $lnTNA_{i,t-1,1st}$ is the natural logarithm of the average monthly TNA of fund *i* during the first half of the year, and $\boldsymbol{Controls}_{i,t-1,1st}$ is a vector of control variables. All other definitions are in **Table 1**.The data are obtained from the CRSP Mutual Fund Database and the sample period is from 2000-2009. **Panel A** reports results for all funds ("All Funds"), **Panel B** reports results for funds with ages between 2-5 years ("Young Funds"), **Panel C** reports results for funds with ages between 6-10 years ("Medium Funds"), and **Panel D** reports results for funds with ages more than 11 years ("Old Funds"). For each regression specification, the first row reports the OLS coefficient estimate, and the second row reports the *t*-statistic in parentheses. Boldface (italicized) *t*-statistics indicate significance at the 5% (10%) level. All equations are estimated with time fixed-effects and, based on Petersen (2009), the standard errors are clustered by fund. The row labeled "R^2" reports the R^2 in decimals. To reduce the impact of outliers, all variables are winsorized at the 1st and 99th percentile.

	Dependent Variable: $\Delta Risk_{t-1}$											
Independent Variables	**Panel A:** All Funds			**Panel B:** Young Funds			**Panel C:** Medium Funds			**Panel D:** Old Funds		
	[1]	[2]	[3]	[1]	[2]	[3]	[1]	[2]	[3]	[1]	[2]	[3]
$Risk_{t-1,1st}$	-0.201	-0.198	-0.198	-0.244	-0.241	-0.241	-0.164	-0.165	-0.165	-0.201	-0.202	-0.203
	(-28.26)	**(-27.65)**	**(-27.53)**	**(-17.26)**	**(-16.83)**	**(-16.65)**	**(-15.25)**	**(-13.80)**	**(-13.82)**	**(-20.16)**	**(-19.06)**	**(-19.05)**
$Turn_{t-1,1st}$	0.0003	0.0003	0.0003	0.0004	0.0004	0.0004	0.0002	0.0003	0.0003	0.0003	0.0003	0.0003
	(7.70)	**(7.66)**	**(7.69)**	**(4.68)**	**(4.66)**	**(4.67)**	**(4.16)**	**(4.22)**	**(4.23)**	**(4.73)**	**(4.55)**	**(4.67)**
RSI_{t-1}	-0.023			-0.053			0.001			0.021		
	(-1.25)			**(-1.97)**			(0.01)			(0.70)		
RSI^{+}_{t-1}	0.003			0.009			-0.001			-0.004		
	(0.96)			(1.66)			(-0.08)			(-0.80)		
RSI^{-}_{t-1}		-0.035	-0.031		-0.070	-0.070		-0.050	-0.059		0.060	0.060
		(-1.29)	(-1.14)		(-1.61)	(-1.59)		(-0.64)	(-0.72)		(1.35)	(1.34)
$RSI^{+}_{t-1} * D^{-}_{t-1,1st}$		-0.017	-0.009		-0.045	-0.045		0.034	0.036		-0.019	0.006
		(-0.67)	(-0.39)		(-1.30)	(-1.42)		(0.55)	(0.58)		(-0.39)	(0.12)
$RSI^{-}_{t-1} * D^{+}_{t-1,1st}$			-0.014			-0.007			0.021			-0.006
			(-0.91)			(-0.36)			(0.54)			(-0.22)
$RSI_{t-1} * lnTNA_{t-1,1st}$			-0.014			0.000			-0.005			-0.048
			(-1.03)			(0.03)			(-0.16)			(-1.91)
$RSI^{+}_{t-1} * lnTNA_{t-1,1st}$		0.004	0.004		0.011	0.011		0.010	0.010		-0.009	-0.009
		(0.82)	(0.82)		(1.23)	(1.24)		(0.73)	(0.75)		(-1.40)	(-1.40)
$RSI^{-}_{t-1} * lnTNA_{t-1,1st}$		0.003	0.002		0.008	0.008		-0.008	-0.008		0.002	0.001
		(0.65)	(0.55)		(1.19)	(1.24)		(-0.73)	(-0.73)		(0.32)	(0.16)
R^2	0.573	0.573	0.573	0.533	0.533	0.533	0.556	0.557	0.557	0.597	0.597	0.597

Table 6: Impact of Managerial Response due to Risk-Shifting Incentives on Returns

The table reports estimates from the following panel regression

$$r_{i,t-1,2nd} = \theta_1 \cdot \Delta\widehat{Risk}^{RSI}_{i,t-1} + \theta_2 \cdot \Delta\widehat{Risk}^{Other}_{i,t-1} + \boldsymbol{\theta}' \cdot \boldsymbol{Controls}_{i,t-1} + u_{i,t-1},$$

where $\Delta\widehat{Risk}_{i,t-1}$ is the predicted value for the change in risk ($\Delta Risk_{i,t-1}$) for fund i, $\Delta\widehat{Risk}^{RSI}_{i,t-1}$ is the predicted value for the change in risk ($\Delta Risk_{i,t-1}$) for fund i attributable to the risk-shifting incentive terms, $\Delta\widehat{Risk}^{Other}_{i,t-1}$ is the predicted value for the change in risk ($\Delta Risk_{i,t-1}$) for fund i attributable to all other terms, including time-fixed effects, and $\boldsymbol{Controls}_{i,t-1}$ is a vector of control variables that are measured over the first or second half of year *t-1*.The predicted values are obtained from Regression [1] of Table 5 from Panel A,B,C, or D, based on the funds for which the equation is being estimated. All other definitions are in **Table 1**.The data are obtained from the CRSP Mutual Fund Database and the sample period is from 2000-2009. **Panel A** reports results for all funds ("All Funds"), **Panel B** reports results for funds with ages between 2-5 years ("Young Funds"), **Panel C** reports results for funds with ages between 6-10 years ("Medium Funds"), and **Panel D** reports results for funds with ages more than 11 years ("Old Funds"). For each regression specification, the first row reports the OLS coefficient estimate, and the second row reports the *t*-statistic in parentheses. Boldface (italicized) *t*-statistics indicate significance at the 5% (10%) level. All equations are estimated with time fixed-effects and, based on Petersen (2009), the standard errors are clustered by fund. The row labeled "R^2" reports the R^2 in decimals. To reduce the impact of outliers, all variables are winsorized at the 1st and 99th percentile.

Independent Variables	*Dependent Variable:* $r_{t-1,2nd}$											
	Panel A: All Funds			**Panel B:** Young Funds			**Panel C:** Medium Funds			**Panel D:** Old Funds		
	[1]	[2]	[3]	[1]	[2]	[3]	[1]	[2]	[3]	[1]	[2]	[3]
$\Delta Risk_{t-1}$	-0.154			0.557			-0.203			-0.574		
	(-0.57)			(0.80)			(-0.49)			(-1.72)		
$\Delta\widehat{Risk}_{t-1}$		0.288			5.405			-1.804			-2.906	
		(0.28)			**(2.65)**			(-1.03)			**(-2.33)**	
$\Delta\widehat{Risk}^{RSI}_{t-1}$			31.932			18.445			64.985			20.132
			(1.64)			(1.31)			(0.62)			(0.56)
$\Delta\widehat{Risk}^{Other}_{t-1}$			0.202			5.213			-1.792			-2.955
			(0.19)			**(2.53)**			(-1.02)			**(-2.38)**
$Flow_{t-1,1st}$	-0.072	-0.073	-0.073	-0.095	-0.095	-0.095	-0.050	-0.051	-0.051	-0.084	-0.087	-0.086
	(-15.75	**(-15.84)**	**(-15.86)**	**(-8.28)**	**(-8.06)**	**(-8.10)**	**(-8.13)**	**(-8.11)**	**(-8.11)**	**(-11.71)**	**(-12.01)**	**(-12.01)**
$Flow_{t-1,2nd}$	0.088	0.088	0.087	0.090	0.088	0.088	0.082	0.082	0.082	0.096	0.097	0.097
	(17.55)	**(17.42)**	**(17.36)**	**(7.16)**	**(6.99)**	**(6.97)**	**(12.30)**	**(12.26)**	**(12.24)**	**(12.26)**	**(12.20)**	**(12.21)**
$Turn_{t-1,2nd}$	-0.007	-0.007	-0.007	-0.009	-0.009	-0.009	-0.004	-0.004	-0.004	-0.009	-0.008	-0.008
	(-6.07)	**(-6.05)**	**(-6.07)**	**(-2.83)**	**(-2.97)**	**(-2.98)**	**(-3.05)**	**(-3.02)**	**(-3.02)**	**(-6.77)**	**(-6.46)**	**(-6.47)**
$lnTNA_{t-1,2nd}$	-0.140	-0.143	-0.139	-0.461	-0.460	-0.442	-0.109	-0.110	-0.112	-0.103	-0.105	-0.106
	(-3.96)	**(-3.99)**	**(-3.88)**	**(-2.69)**	**(-2.65)**	**(-2.54)**	(-1.63)	(-1.63)	(-1.66)	**(-2.28)**	**(-2.34)**	**(-2.34)**
$ER_{t-1,2nd}$	-0.711	-0.698	-0.705	-0.576	-0.123	-0.143	-0.591	-0.659	-0.659	-0.843	-1.013	-1.016
	(-4.41)	**(-4.26)**	**(-4.31)**	(-1.20)	(-0.25)	(-0.29)	**(-2.70)**	**(-2.91)**	**(-2.91)**	**(-3.67)**	**(-4.29)**	**(-4.30)**
R^2	0.151	0.151	0.151	0.178	0.186	0.187	0.152	0.150	0.150	0.148	0.150	0.150

Table 7: Managerial Response to Actual Incentives

The table reports estimates from the following panel regression

$$\Delta Risk_{i,t-1} = \gamma_1 r^{++}_{i,t-1,1st} + \gamma_2 r^{+}_{i,t-1,1st} + \gamma_3 r^{-}_{i,t-1,1st} + \gamma_4 r^{--}_{i,t-1,1st} + \gamma_5 (r^{++}_{i,t-1,1st} \cdot lnTNA_{i,t-1,1st}) + \gamma_6 (r^{+}_{i,t-1,1st} \cdot lnTNA_{i,t-1,1st}) + \gamma_7 (r^{-}_{i,t-1,1st} \cdot lnTNA_{i,t-1,1st}) + \gamma_8 (r^{--}_{i,t-1,1st} \cdot lnTNA_{i,t-1,1st}) + \boldsymbol{\gamma}' \boldsymbol{Controls}_{i,t-1,1st} + v_{i,t-1},$$

where $r^{--}_{i,t-1,1st} = \min[(r_{i,t-1,1st} + b_{0,t-1,1st}),0]$, $r^{-}_{i,t-1,1st} = \max[-b_{0,t-1,1st}, \min(0, r_{i,t-1,1st})]$, $r^{+}_{i,t-1,1st} = \min[b_{1,t-1,1st}, \max(0, r_{i,t-1,1st})]$, $r^{++}_{i,t-1,1st} = \max[(r_{i,t-1,1st} - b_{1,t-1,1st}),0]$, $r_{i,t-1,1st}$ is the excess return for fund *i* during the first half of year *t-1*, $b_{0,t-1,1st}$ and $b_{1,t-1,1st}$ are positive constants that equal the median negative and positive excess return during the first half of year *t-1* for funds for which the equation is being estimated, $lnTNA_{i,t-1,1st}$ is the natural logarithm of the average monthly TNA of fund *i* during the first half of the year, and $\boldsymbol{Controls}_{i,t-1,1st}$ is a vector of control variables. All other definitions are in **Table 1**.The data are obtained from the CRSP Mutual Fund Database and the sample period is from 2000-2009. **Panel A** reports results for all funds ("All Funds"), **Panel B** reports results for funds with ages between 2-5 years ("Young Funds"), **Panel C** reports results for funds with ages between 6-10 years ("Medium Funds"), and **Panel D** reports results for funds with ages more than 11 years ("Old Funds"). For each regression specification, the first row reports the OLS coefficient estimate, and the second row reports the *t*-statistic in parentheses. Boldface (italicized) *t*-statistics indicate significance at the 5% (10%) level. All equations are estimated with time fixed-effects and, based on Petersen (2009), the standard errors are clustered by fund. The row labeled "R^2" reports the R^2 in decimals. To reduce the impact of outliers, all variables are winsorized at the 1st and 99th percentile.

Independent Variables	*Dependent Variable:* $\Delta Risk_{t-1}$							
	Panel A: All Funds		**Panel B:** Young Funds		**Panel C:** Medium Funds		**Panel D:** Old Funds	
	[1]	[2]	[1]	[2]	[1]	[2]	[1]	[2]
$Risk_{t-1,1st}$	-0.219	-0.220	-0.261	-0.261	-0.181	-0.181	-0.215	-0.217
	(**-30.19**)	(**-30.29**)	(**-17.60**)	(**-17.55**)	(**-15.86**)	(**-15.93**)	(**-20.88**)	(**-21.26**)
$Turn_{t-1,1st}$	0.0003	0.0003	0.0004	0.0004	0.0003	0.0002	0.0003	0.0003
	(**7.84**)	(**7.44**)	(**4.65**)	(**4.56**)	(**4.31**)	(**4.16**)	(**4.93**)	(**4.52**)
$r^{++}_{t-1,1st}$	0.012	0.015	0.016	0.025	0.008	0.012	0.010	0.003
	(**8.07**)	(**2.50**)	(**4.51**)	(1.56)	(**3.50**)	(1.32)	(**4.78**)	(0.32)
$r^{+}_{t-1,1st}$	-0.007	0.000	-0.012	-0.013	-0.001	0.007	-0.006	0.015
	(**-3.310**)	(0.07)	(**-2.42**)	(-0.83)	(-0.23)	(0.54)	(-1.75)	(1.57)
$r^{-}_{t-1,1st}$	0.007	0.002	0.016	0.012	0.001	0.006	0.003	-0.011
	(**2.89**)	(0.27)	(**2.51**)	(0.68)	(0.15)	(0.56)	(0.77)	(-1.31)
$r^{--}_{t-1,1st}$	-0.007	-0.008	-0.011	-0.025	-0.005	0.004	-0.004	-0.009
	(**-3.70**)	(-1.44)	(**-2.51**)	(-1.60)	(-1.80)	(0.43)	(-1.48)	(-0.86)
$r^{++}_{t-1,1st} * lnTNA_{t-1,1st}$		-0.001		-0.002		-0.001		0.001
		(-0.54)		(-0.59)		(-0.46)		(0.76)
$r^{+}_{t-1,1st} * lnTNA_{t-1,1st}$		-0.001		0.000		-0.001		-0.003
		(-1.34)		(0.08)		(-0.67)		(**-2.31**)
$r^{-}_{t-1,1st} * lnTNA_{t-1,1st}$		0.001		0.001		-0.001		0.002
		(1.05)		(0.22)		(-0.49)		(1.77)
$r^{--}_{t-1,1st} * lnTNA_{t-1,1st}$		0.000		0.003		-0.002		0.001
		(0.32)		(0.90)		(-1.01)		(0.55)
R^2	0.577	0.578	0.542	0.543	0.559	0.560	0.599	0.600

Table 8: Impact of Managerial Response due to Actual Incentives on Returns

The table reports estimates from the following panel regression

$$r_{i,t-1,2nd} = \theta_1 \cdot \widehat{\Delta Risk}^{r_{1st}}_{i,t-1} + \ \theta_2 \cdot \widehat{\Delta Risk}^{Other}_{i,t-1} + \boldsymbol{\theta}' \cdot \boldsymbol{Controls}_{i,t-1} + u_{i,t-1},$$

where $\widehat{\Delta Risk}_{i,t-1}$ is the predicted value for the change in risk ($\Delta Risk_{i,t-1}$) for fund *i*, $\widehat{\Delta Risk}^{r_{1st}}_{i,t-1}$ is the predicted value for the change in risk ($\Delta Risk_{i,t-1}$) for fund *i* attributable to the return, $\widehat{\Delta Risk}^{Other}_{i,t-1}$ is the predicted value for the change in risk ($\Delta Risk_{i,t-1}$) for fund *i* attributable to all other terms, including time-fixed effects, and $\boldsymbol{Controls}_{i,t-1}$ is a vector of control variables that are measured over the first or second half of year *t-1*.The predicted values are obtained from Regression [1] of Table 7 from Panel A,B,C, or D, based on the funds for which the equation is being estimated. All other definitions are in **Table 1**.The data are obtained from the CRSP Mutual Fund Database and the sample period is from 2000-2009. **Panel A** reports results for all funds ("All Funds"), **Panel B** reports results for funds with ages between 2-5 years ("Young Funds"), **Panel C** reports results for funds with ages between 6-10 years ("Medium Funds"), and **Panel D** reports results for funds with ages more than 11 years ("Old Funds"). For each regression specification, the first row reports the OLS coefficient estimate, and the second row reports the *t*-statistic in parentheses. Boldface (italicized) *t*-statistics indicate significance at the 5% (10%) level. All equations are estimated with time fixed-effects and, based on Petersen (2009), the standard errors are clustered by fund. The row labeled "R^2" reports the R^2 in decimals. To reduce the impact of outliers, all variables are winsorized at the 1st and 99th percentile.

Independent Variables	*Dependent Variable:* $r_{t-1,2nd}$											
	Panel A: All Funds			**Panel B:** Young Funds			**Panel C:** Medium Funds			**Panel D:** Old Funds		
	[1]	[2]	[3]	[1]	[2]	[3]	[1]	[2]	[3]	[1]	[2]	[3]
$\Delta Risk_{t-1}$	-0.154			0.557			-0.203			-0.574		
	(-0.57)			(0.80)			(-0.49)			(-1.72)		
$\widehat{\Delta Risk}_{t-1}$		2.540			7.903			1.144			-1.791	
		(**2.50**)			(**3.93**)			(0.69)			(-1.47)	
$\widehat{\Delta Risk}^{r_{1st}}_{t-1}$			28.679			35.440			35.627			16.985
			(**8.70**)			(**5.96**)			(**5.80**)			(**3.17**)
$\widehat{\Delta Risk}^{Other}_{t-1}$			2.461			7.125			1.669			-1.645
			(**2.43**)			(**3.56**)			(1.00)			(-1.35)
$Flow_{t-1,1st}$	-0.072	-0.073	-0.076	-0.095	-0.095	-0.102	-0.050	-0.051	-0.052	-0.084	-0.086	-0.089
	(**-15.75**)	(**-15.84**)	(**-16.80**)	(**-8.28**)	(**-8.20**)	(**-8.82**)	(**-8.13**)	(**-8.12**)	(**-8.49**)	(**-11.71**)	(**-11.95**)	(**-12.31**)
$Flow_{t-1,2nd}$	0.088	0.087	0.084	0.090	0.087	0.084	0.082	0.082	0.080	0.096	0.097	0.095
	(**17.55**)	(**17.31**)	(**17.01**)	(**7.16**)	(**6.89**)	(**6.87**)	(**12.30**)	(**12.19**)	(**12.07**)	(**12.26**)	(**12.21**)	(**12.11**)
$Turn_{t-1,2nd}$	-0.007	-0.007	-0.008	-0.009	-0.010	-0.011	-0.004	-0.004	-0.004	-0.009	-0.008	-0.009
	(**-6.07**)	(**-6.24**)	(**-7.02**)	(**-2.83**)	(**-3.11**)	(**-3.74**)	(**-3.05**)	(**-3.12**)	(**-3.58**)	(**-6.77**)	(**-6.53**)	(**-6.65**)
$lnTNA_{t-1,2nd}$	-0.140	-0.147	-0.141	-0.461	-0.478	-0.468	-0.109	-0.111	-0.115	-0.103	-0.108	-0.101
	(**-3.96**)	(**-4.09**)	(**-3.95**)	(**-2.69**)	(**-2.75**)	(**-2.73**)	(-1.63)	(-1.65)	(-1.73)	(**-2.28**)	(**-2.39**)	(**-2.25**)
$ER_{t-1,2nd}$	-0.711	-0.589	-0.787	-0.576	0.030	-0.334	-0.591	-0.541	-0.718	-0.843	-0.961	-1.038
	(**-4.41**)	(**-3.57**)	(**-4.91**)	(-1.20)	(0.06)	(-0.70)	(**-2.70**)	(**-2.38**)	(**-3.23**)	(**-3.67**)	(**-4.06**)	(**-4.39**)
R^2	0.151	0.152	0.162	0.178	0.195	0.215	0.152	0.150	0.161	0.148	0.149	0.153

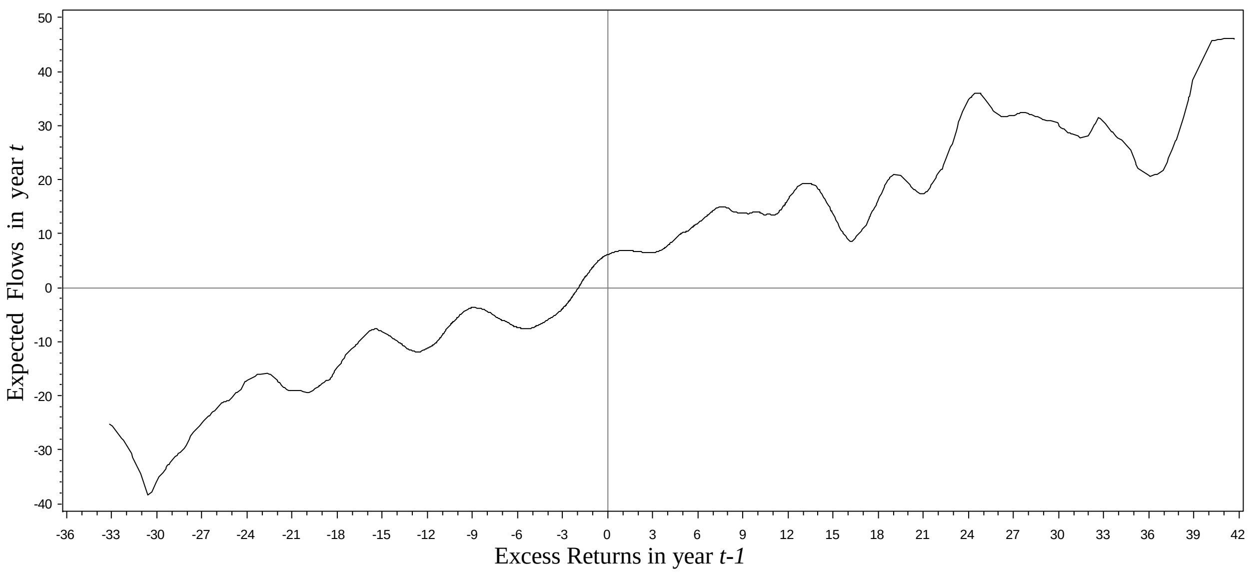

Figure 1: Flow-performance Relationship for Young Funds (Age 2-5 Years)

The figure plots estimates of the function *g*(.) in Equation 4.1 of the text for Young funds using data from 2000-2009. The kernel regression uses the Nadaraya-Watson estimator with the Epanechnikov kernel and the bandwidth is chosen using the Least Squares Cross Validation method.

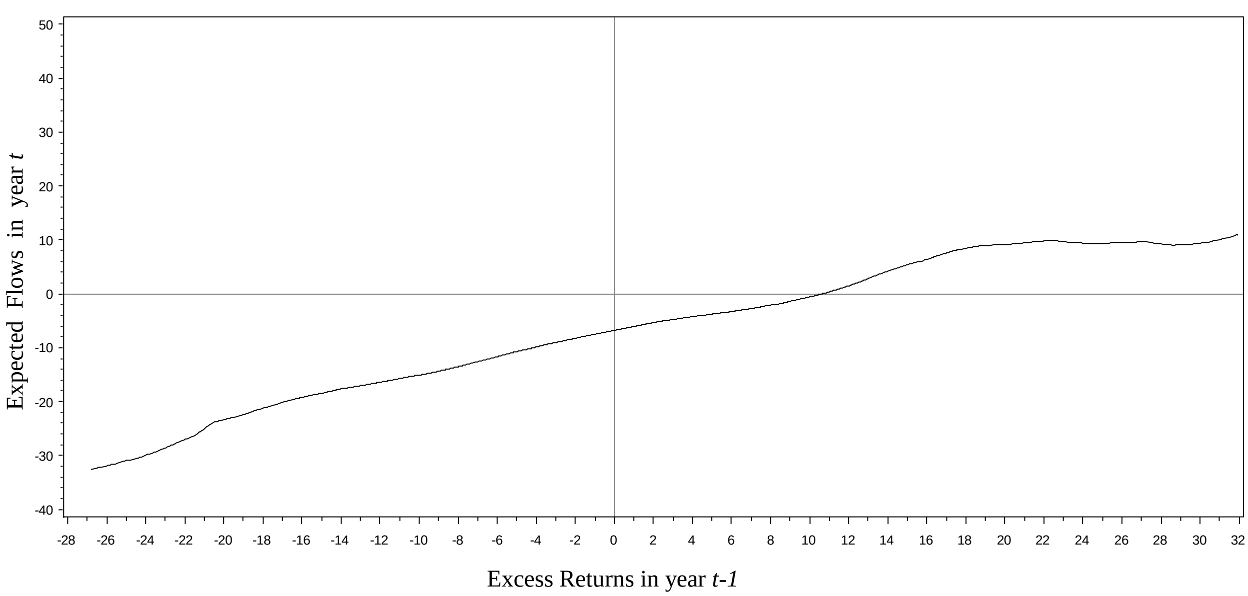

Figure 2: Flow-performance Relationship for Medium Funds (Age 6-10 Years)

The figure plots estimates of the function *g*(.) in Equation 4.1 of the text for Medium funds using data from 2000-2009. The kernel regression uses the Nadaraya-Watson estimator with the Epanechnikov kernel and the bandwidth is chosen using the Least Squares Cross Validation method.

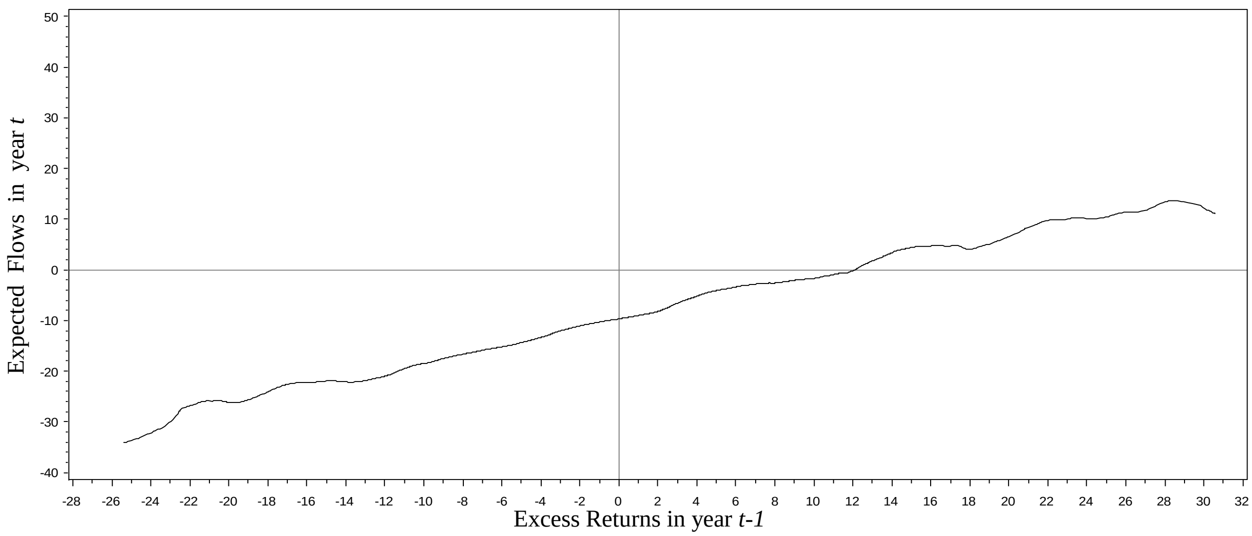

Figure 3: Flow-performance Relationship for Old Funds (Age 11+ Years)

The figure plots estimates of the function *g*(.) in Equation 4.1 of the text for Old funds using data from 2000-2009. The kernel regression uses the Nadaraya-Watson estimator with the Epanechnikov kernel and the bandwidth is chosen using the Least Squares Cross Validation method.

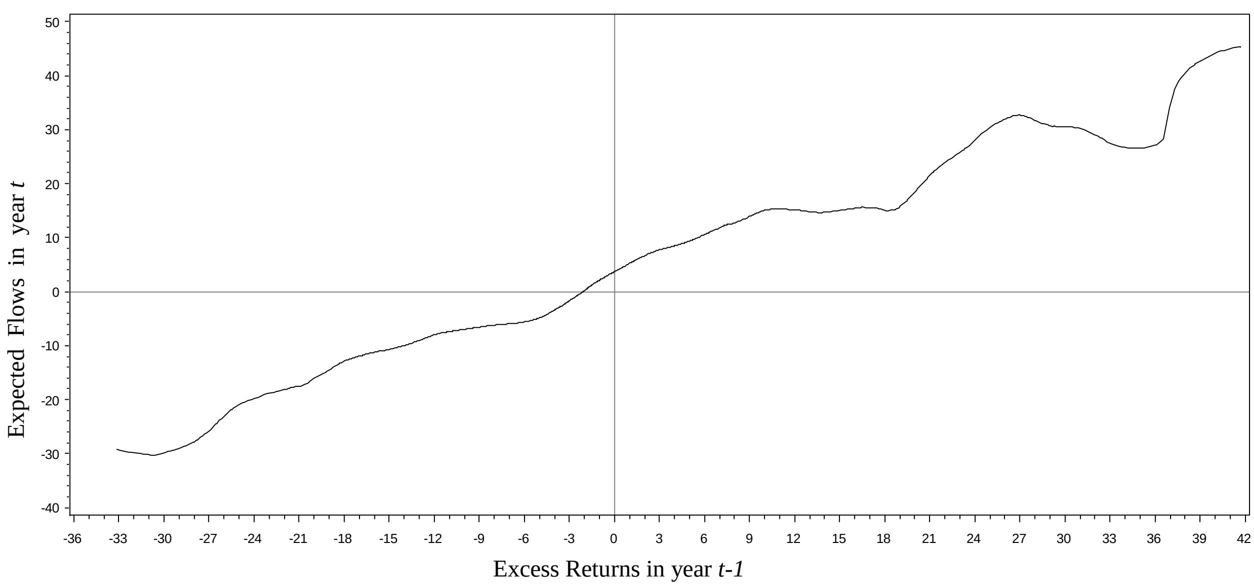

Figure 4: Flow-performance Relationship for Young Funds (Age 2-5 Years)

The figure plots estimates of the function *g*(.) in Equation 4.1 of the text for Young funds using data from 2000-2009 and twice the bandwidth obtained from Least Squares Cross Validation method. The kernel regression uses the Nadaraya-Watson estimator with the Epanechnikov kernel.