Mutual Fund's R^2 as Predictor of Performance

By

Yakov Amihud* and Ruslan Goyenko**

Abstract:

We propose that fund performance is predicted by its R^2, obtained by regressing its return on the Fama-French-Carhart four benchmark portfolios. Lower R^2, or higher idiosyncratic risk relative to total risk, measures selectivity or active management. We show that lagged R^2 has significant negative predictive coefficient in predicting *alpha* or *Information Ratio*. This is consistent with Cremers and Petajisto's (2008) results on the effect of selectivity. Funds ranked into lagged lowest-quintile R^2 and highest-quintile *alpha* produce significant *alpha* of 2.8%. Also, both fund *RMSE* and return volatility predict the following year's performance with significant positive and negative coefficients, respectively. Across funds, R^2 is an increasing function of fund size and a decreasing function of its age, its manager tenure and its past performance, but better performance induces funds to subsequently increase their R^2.

First version: March 17, 2008
This version: February 25, 2009

* Ira Leon Rennert Professor of Finance, Stern School of Business, New York University
** Desautels Faculty of Management, McGill University
We thank Alex Kane, Martijn Cremers, Marcin Kacperczyk, Anthony Lynch and Antti Petajisto for helpful comments and discussions.

1. Introduction

Fama (1972) suggests that a portfolio's overall performance in excess of the beta-adjusted return on a benchmark (or naïve) portfolio is due to selectivity, which "measures how well the chosen portfolio did relative to a naively selected portfolio with the same level of risk" (Fama, 1972, p. 557). Recent studies show that fund performance is positively affected by fund selectivity or active management, measured by the deviation of funds holdings from some diversified benchmark portfolio (see review below). The problem is that this measure of selectivity requires knowledge of the portfolio composition of all mutual funds and of their benchmark indexes, which is hard for many investors to obtain and calculate. It also hard to measure selectivity when the benchmark portfolio is not well-defines, that is, when funds opt to outperform some combination of benchmark indexes.

We propose a simple and intuitive measure of mutual fund selectivity: the fund's R^2 from the standard four-factor regression model of Fama-French (1993) and Carhart's (1997), which includes four factor-mimicking portfolios: *RM-Rf* (the market portfolio excess return), *SMB* (small minus big size stocks), *HML* (high minus low book-to-market ratio stocks) and *UMD* (winner minus loser stocks). R^2, the proportion of the return variance that is explained by broad portfolios or indexes, is a traditional measure of diversification, and thus $1-R^2$ measures the weight of idiosyncratic risk or selectivity. The closer is R^2 to 1, the closer does the fund track the benchmark portfolios and the lower the selectivity. If selectivity enhances mutual fund performance, R^2 should negatively predict the fund's performance.

This is indeed what we find: *R^2 has a negative and significant predictive effect on fund performance*, using two conventional measures: the intercept *alpha* from the Fama-French-Carhart four-factor regression model, and the *Information Ratio*, which is *alpha* scaled by the idiosyncratic (regression residual) risk. We also identify an R^2-based strategy that earns significantly positive average excess return (factor-adjusted): at the beginning of each year, select funds whose lagged R^2 is in the lowest quintile and whose *alpha* is in the highest quintile. These funds generate a significant *alpha* of 2.81%.

Our results are robust to the indexes used. A recent study by Cremers, Petajisto and Zitzewitz (2008) criticizes the use of Fama-French (1993) indexes *SMB* and the *HML* in evaluating mutual fund performance. We re-do our analysis using instead the returns on the six Fama-French portfolios (2x3) classified by size (small and big) and value, neutral or growth, in addition to the market excess return and Carhart's *UMD*. Our results remain unchanged: R^2 has negative and highly significant predictive effect on the following year's *alpha* and *Information Ratio*.

R^2 is also lower due to another aspect of active fund management: rotation between characteristics or factors over time, which may reflect timing. We estimate yearly four-factor regressions with fixed factor coefficients, while active fund managers change their portfolio such that it rotates between factors. Mamaysky, Spiegel and Zhang (2007) estimate funds' factor betas over five-year periods by Kalman Filter and find that the coefficients vary over time. Our estimation period is only one year, during which factor rotation is more limited, but such rotation can still be done to some extent, resulting in lower estimated R^2.

R^2 is decreasing in the regression *RMSE* (root mean squared error) and increasing in the standard deviation of the fund return, denoted by *SDR* (their squared ratio equals $1\text{-}R^2$). The *RMSE* is related to the "tracking error" in studies of active fund management. Wermers (2003) finds that the standard deviation of S&P500-adjusted fund return is positively related to the contemporaneous fund performance, measured by *alpha* from the Carhart (1997) four-factor model. Cremers and Petajisto (2008) find that the tracking error (the standard deviation of the fund's benchmark-adjusted returns) has insignificant predictive effect on performance. However, studies of the effect of tracking error on performance often omit from the estimation model the *SDR*, which correlates positively with the tracking error. Such omitted-variable specification may result in a biased estimation of the effect of tracking error on performance. We include both *RMSE* and *SDR* in the regression and find that *RMSE* has a *positive* and significant predictive effect on fund performance while *SDR* has *negative* and significant predictive effect on fund performance. Together, these effects are consistent with the effect of R^2.

Recent studies of hedge fund performance use R^2 as a measure of fund strategy and find similar results: lower R^2 predicts better fund performance. Titman and Tiu (2008) conclude that hedge fund performance is better when they do less hedging against common benchmarks, using Fung and Hsieh (2001, 2004) benchmark indices, and suggest that choosing smaller exposure to factor risk reflects hedge funds managers' confidence in their ability. Wang and Zheng (2008) define *$1\text{-}R^2$* as the "hedge fund distinctiveness index," where R^2 is obtained from a regression of the hedge fund return on the return on its hedge fund style index, or on the aggregate hedge fund index, or on the Fung and Hsieh (2001) 7-factor model.

Fund selectivity is shown to enhance performance. Daniel, Grinblatt, Titman and Wermers (1997) analyze selectivity at the securities level, finding that securities that are picked by mutual funds outperform a characteristic-based benchmark, although the gain from stock picking approximately equals the funds' average management fee. Other studies examine selectivity at the fund level. Brand, Brown and Gallagher (2005) measure a fund active management by a divergence index, defined as the sum of squared deviations of the fund portfolio's stock weights from the market portfolio (or portfolio's deviations from the benchmark with respect to holdings the industry and sector level), using Australian data. They find that the divergence index positively predicts fund performance. Cremers and Petajisto (2008) show that Active Share, which represents the share of portfolio holdings that differ from the fund's benchmark index holdings, significantly predicts fund performance, after controlling for other fund characteristics. And, sorting funds on prior one-year performance and on Active Share, they identify a group of funds with active share and high prior performance that generates significantly positive four-factor *alpha*, after controlling for benchmark (or style) returns. Notably, these returns are net of expenses. Kacperczyk, Sialm and Zheng (2005) find that funds exhibit better performance if they have greater industry concentration of holdings compared to the weights of these industries in a diversified portfolio, and Kacperczyk and Seru (2007) find that funds whose stocks holdings are related to company-specific information from analysts' expectations exhibit better performance.

Our study examines the effect of fund selectivity on performance, using measures which do not require knowledge of the fund portfolio holdings. We proceed as follows. Section 2 presents the fund performance measures that we use and their estimation procedure, and then it presents the performance predictors that we use, R^2 and its components, the residual mean-squared error and the return standard deviation. Section 3 describes data and sample selection procedure. Section 4 presents the results on the prediction of next-year fund performance, employing two performance measures – *alpha* and *InfRatio* – and various predictive methods. We also explain why the predictive power of our measures is weaker in early period and stronger in more recent periods. In Section 5 we show how using information about past fund performance and R^2 enables to choose a portfolio of funds which produces significant positive performance in the following year. In Section 6 we present estimation of the association between fund characteristics and our performance predictor R^2. Concluding remarks are in Section 7.

2. Performance measures and performance predictors

2.1. Performance measures

We employ two standard measures of fund performance. The first is the intercept $alpha_j$ from the four-factor regression model of Fama and French (1997) and Carhart (1997),

$$R^e_{j,t} = alpha_j + \beta 1_j(RM_t - rf_{,t}) + \beta 2_j SMB_t + \beta 3_j HML_t + \beta 4_j UMD_t + e_{j,t} \,. \qquad (1)$$

$R^e_{j,t} = R_{j,t} - r_{f,t}$ is the excess return on fund j in period t in excess of the risk-free rate, the four factors are defined above and $e_{j,t}$ is the residual. (In Section 4.8 below we present results using an alternative set of indexes.)

The second performance measure is the *Information Ratio* or the *Appraisal Ratio*, which measures the extent of the fund's excess performance relative to its idiosyncratic risk.

$$InfRatio_j = \frac{alpha_j}{RMSE_j} \,. \qquad (2)$$

$RMSE_j$ is the squared root of the mean squared errors or residuals $e_{j,t}$ from (1). Treynor and Black (1973), who introduce the *Appraisal Ratio* in the context of the single-index (CAPM) model, show that considering an asset j as part of an optimal portfolio, the fraction of the investor's capital devoted to the jth asset is proportional to the *InfRatio*. If evaluate a mutual fund as an active investment component in an efficient portfolio rather than a sole repository of the investor's wealth, Bodie, Kane and Markus (2009, pp. 262-263) show that the larger is the *InfRatio* of a fund, the greater is the demand for the fund. Following Treynor and Black (1973) they show that an optimally constructed risky portfolio *P*, composed of a passive index portfolio *M* and an active investment portfolio *A*, has the following Sharpe ratio, SR_p:

$$SR_P^2 = SR_M^2 + [\frac{alpha_A}{RMSE_A}]^2 ,$$

where $alpha_A$ and $RMSE_A$ are measured with respect to the passive index *M*. Thus, the contribution of mutual fund *A* to the Sharpe ratio of the investor's portfolio is increasing in the fund's *Information Ratio*. This means that a higher fund's *InfRatio* makes the fund more attractive to investors. The fund's *Information Ratio* has been used as a performance measure by Brands, Brown and Gallagher (2005) and by Kacperczyk, Sialm and Zheng (2005).

The use of *Information Ratio* also helps mitigate the survivorship bias in studies of persistence in mutual funds performance. Brown, Goetzmann, Ibbotson and Ross (1992) note that choosing a risky strategy may result in high *alpha* but it also increases the probability of failure. Because we observe the survivors, the apparent pattern is that of persistence of high performance and ex post, superior *alphas* are positively related to idiosyncratic risk. Therefore, scaling *alpha* by the fund idiosyncratic risk reduces the survivorship bias.[1] The Information Ratio, which scales the abnormal fund performance by the volatility of the abnormal fund returns, mitigates this bias.

In what follows, we estimate for each fund both *alpha* and *InfRatio* and analyze how these performance measures can be predicted by various fund characteristics.

2.2 Performance predictors

We predict fund performance in one period by its estimated R^2 in the preceding period, where R^2 is estimated from the regression model (1). As detailed below, because we use daily data and because some stocks that constitute the fund returns are slow to adjust to information, we use in practice the regression model (1) where the fund return is

[1] Brown, Goetzmann and Ross (1995) show that the magnitude of the survivorship bias in the calculation of average stock returns is an increasing function of the return volatility.

regressed on the current and one-lag returns of the benchmark indexes (following Dimson (1979)). We also use as predictors the two components of R^2 (in squared-root values): *RMSE*, the root mean squared errors from (1), and *SDR*, the standard deviation of the excess fund return R^e.

3. Data and Sample Selection

We use the CRSP Survivorship Bias Free Mutual Fund Database with the CDA/Spectrum holdings database and merge the two databases using Mutual Fund Links tables available at CRSP. The monthly returns for mutual funds are from the CRSP Mutual Fund Database from 1989 to 2007. These are net returns, i.e. after fees, expenses, and brokerage commissions but before any front-end or back-end loads. The daily returns from 1989 to 1998 are obtained from the International Center for Finance at Yale School of Management.[2] These data include Standard and Poor's database of live mutual funds.[3] The S&P data are not survivorship-bias free. They are supplemented by another daily database which is used by Goetzmann, Ivkovic, and Rouwenhorst (2001) and obtained from the Wall Street Web. This combined database is survivorship-bias free and is also used by Cremers and Petajisto (2008). CRSP data on daily mutual fund returns begins in March, 1998. Therefore, from 1999 to 2007 we use the CRSP daily data. Altogether, our final sample spans the period from January 1989 to December 2007.

The CRSP database also contains data on total net assets, the fund's turnover ratio, expense ratio, investment objective, and other fund characteristics. We use the end-of-year values of these variables. We also use Cremers and Petajisto (2008) Active Share measure, for which data are available only for funds reporting share holdings on CDA/Spectrum. The criteria for fund selection with Active Share estimated are the same as in Cremers and Petajisto (2008).[4]

The CRSP database identifies each shareclass separately, whereas the CDA database lists only the underlying funds. The Mutual Fund Links tables assign each shareclass to

---

[2] We are grateful to William Goetzmann for providing these data.

[3] This is also previously known as Micropal mutual fund data

[4] We are grateful to Martijn Cremers and Antti Petajisto for providing the Active Share data which are available from 1980 to 2006.

the underlying fund. Whenever a fund has multiple shareclasses at the CRSP database, we compute the weighted CRSP net returns, expenses, turnover ratio and other characteristics for each fund. The weight is based on the most recent total net assets of that shareclass.

Our analysis employs actively managed all-equity funds. Included are funds with investment objective codes from Weisenberg and Lipper to be aggressive growth, growth, growth and income, equity income, growth with current income, income, long-term growth, maximum capital gains, small capitalization growth, micro-cap, mid-cap, unclassified or missing. When both the Weisenberg and the Lipper codes are missing, we use Strategic Insight Objective Code to identify the style, and if Weisenberg, Lipper and Strategic Insight Objective Code are missing, we use investment objective codes from Spectrum, if available, to identify the style. If no code is available for a fund-year and a fund has a past year with the style identified, that fund-year is assigned the style of the previously identified style-year. If the fund style cannot be identified, it is not included in the sample.[5] We classify funds into four style categories which roughly follow the categorizations in Brown and Goetzmann (1997): (i) "Growth" which includes: Aggressive growth, Growth, Long-term growth, Maximum capital gains, (ii) "Income", (iii) "Growth and Income", (iv) "Small cap" which includes: small cap, small-cap growth, micro-cap, mid-cap. We eliminate index funds by deleting those whose name includes the word "index" or the abbreviation "ind". Following Elton, Gruber and Blake (1996), we eliminate funds with total net assets of less than $15 million at the end of the year preceding the test year because inclusion of such funds can cause survivorship bias in estimation due to reporting conventions. Addressing Evans's (2004) comment on incubation bias, we eliminate observations before the reported starting year by CRSP. And, following Cremers and Petajisto (2008), we delete funds with missing name in CRSP. We require funds to have at least 125 daily return data in the first year of two consecutive years, which we use to estimate lagged values of R^2, return variances and *alpha*, and only 50 daily return data in the second year, where we estimate the fund

[5] In case that Wiesenberger and Lipper Code are missing, in which case we use another style identifier, we check if the fund name corresponds to the style. If it does not, we consider the style as un-identified. There are about 5% of fund-years with missing styles.

performance measures *alpha* and *InfRatio,*[6] thus reducing the survivorship bias problem. We also require funds to have data in the first year on expenses, turnover, total net assets, age and managerial tenure.

For the funds that satisfy these requirements, we estimate their R^2 from the regression model (1) for the first year of the two-year pair, using the indexes' current and one-lag returns, following Dimson (1979). We rank all resulting R^2 estimates and symmetrically trim the top and bottom 1% of the observations. The funds with R^2 close to 1.0 are effectively "closet indexers," and very low R^2 may reflect outlier-type strategy or estimation error. We thus obtain a final sample of 16,646 fund-year pairs of 2,314 funds with R^2 ranging between 0.240 and 0.989. This is the sample that we analyze. The mean R^2 is 0.86 and the median is 0.90. Finally, we apply to R^2 a logistic transformation,

$$TR^2 = log[\sqrt{R^2}/(1- \sqrt{R^2})].$$

The resulting distribution of TR^2 is more symmetric than that of R^2. As an alternative to R^2 in prediction performance we use its components: *RMSE*, the root mean squared errors, and *SDR*, the fund return's standard deviation.

The control variables in the predictive cross-fund regression are those that commonly appear in studies of fund performance. For example, Cremers and Petajisto (2008) use Total Net Assets, *TNA*, ($mm), *Expenses*, the expense ratio of the most recently completed fiscal year,[7] *Turnover*, defined as the minimum of aggregated sales or aggregated purchases of securities divided by the average 12-month *TNA* of the fund. Other fund characteristics are *Age*, computed as the difference in years between current date and the date the fund was first offered, and *Manager Tenure* in logarithm, the difference in years between the current date and the date when the current manager took control. An important predictor of future performance is lagged *alpha* or *InfRatio* which may reflect managerial skill and strategy and is shown to be a significant predictor of future performance (see Brown and Goetzmann (1995) and Gruber (1996)).

INSERT TABLE I

[6] Cremers and Petajisto (2008) require 125 days in the performance estimation year (the second of the two-year pair). Our results do not materially change under this requirement.

[7] Expense ratio is the ratio of total investment that shareholders pay for the fund's operating expenses, which include 12b-1 fees. Expense ratio may include waivers and reimbursements, causing it to appear to be less then the fund management fee.

Table I presents the statistics of our sample. Panel A presents fund characteristics, while Panel B presents the correlations between them. We observe that R^2 is larger for large funds, which cannot be niche investors and must hold a broad portfolio, which makes their performance closer to that of broad indexes. Funds with more idiosyncratic investment – being more active – have higher expense ratio, as evident from the negative correlation between R^2 and *Expenses*. A detailed analysis of the relationship between R^2 and the other control variables is presented in Table X.

4. Fund Performance prediction in cross-sectional regressions

We study the relationship between fund performance and lagged R^2 by regressing the fund annualized *alpha* from Model (1) and *InfRatio* (Information Ratio) defined in (2) on the fund's previous-year TR^2 (logistic transformation of R^2) and control variables. All fund characteristics that are used to predict performance are known at the end of year *y-1* while performance is measured over the following year *y*.

4.1. Fund alpha as a measure of performance

We expect *alpha* to be a negative function of the fund's R^2. Table II presents the results of a regression of *alpha* on TR^2 or on its components *RMSE* and *SDR*, the root mean squared errors from regression (1) and the standard deviation of the fund excess return R^e_t. We estimate the performance over the 18 years, 1990-2007 (the first year for parameter estimation is 1989) in a pooled regression with year dummy variables and four style dummy variables. Errors are clustered at the fund level.

INSERT TABLE II

The estimation results in Table II, column (1) show that R^2 is a strong predictor of *alpha*. The coefficient of TR^2 is –0.680 with t = 7.69. This means that funds with low R^2, which may be more active in pursuing stock selection strategies, perform better. R^2 is decreasing in *RMSE* and increasing in *SDR*. In column (2) we estimate the effect of these components of R^2 on *alpha*. The coefficient of *RMSE* is 3.955 (t = 6.33) and the coefficient of *SDR* is –6.688 (t = 19.28). This pair of results is consistent with the results on the negative effect of R^2. Our result on the significant positive effect of *RMSE* should be compared to the mixed results on its effect obtained in previous studies that use *RMSE* as a measure of "tracking error," a proxy for fund active management or selectivity. These studies omit the total fund risk *SDR* which is positively correlated with *RMSE* and itself has negative coefficient in the performance equation. The correlation between *RMSE* and *SDR* in our sample is 0.686. The omission of *SDR* produces a downward bias in the coefficient of *RMSE*, hence the mixed results on its effect.

The effect of fund size (*TNA*) on performance is negative, although this negative effect is mitigated for very large funds, as evident from the positive and significant coefficient of $log(TNA)^2$. *Expenses* negatively affect performance, as observed by Gruber (1996). Given that R^2 is negatively correlated with *Expenses* (see Table I, Panel B), we re-estimate the model excluding the variable *Expenses*. The coefficient of TR^2 changes very little, remaining negative and highly significant. The effect of *Manager Tenure* (in logarithm) is negative, meaning that managers who are longer time on the job generate worse performance. However, the coefficient is not statistically significant. We revisit the effect of this variable later in this paper.

Our result on the superior performance of funds with higher R^2 is consistent with the findings of Cremers and Petajisto (2008) on better performance of funds with active management, measured by *AS* (Active Share), the sum of absolute deviations of the fund's stock holdings (weights) from those of its benchmark portfolio. We replicate their result in column (3): *AS* has a positive coefficient, 1.579, with t = 3.01. The sample decreases to 1,890 funds because the calculation of *AS* requires fund portfolio holdings data, and their sample ends in 2006. When including both TR^2 and *AS* in the regression (column 4), TR^2 retains its negative and highly significant effect while the coefficient of *AS* becomes insignificant (with negative sign). Similarly, the effects of *RMSE* and *SDR* remain practically unchanged when *AS* is included in the model (column 5), while the coefficient of *AS* switches to become negative and marginally insignificant. Notably, however, Cremers and Petajisto (2008) measure the performance of their Active Shares measure relative to a specific fund's benchmark portfolio, not relative to *alpha* from the Fama-French-Carhart's multi-factor model.

In the rest of the table, we split the sample into two equal nine year subperiods. The year 1999, which begins the second subperiod, coincides with the beginning year of CRSP data. The first nine-year subperiod (1990-1998) has ¼ of the sample fund years while the second subperiod (1999-2007) that utilizes CRSP data has ¾ of the sample fund years. In the first subperiod, TR^2 is insignificant and also *RMSE* is insignificant, while *SDR* retains its negative and significant effect. We explain the weak performance of TR^2 during the first nine-year period in Section 4.3 below. In the recent nine-year subperiod that includes most of the data, TR^2 has a negative and highly significant effect on *alpha*, and the pair *RMSE* and *SDR* have the expected signs – positive and negative, respectively – with high statistical significance. The results obtained for the whole sample hold stronger for the last nine years of the sample.

4.2. Information Ratio as a measure of performance

The second performance measure is the fund's *Information Ratio*, $InfRatio_j = alpha_j/RMSE_j$. Theoretically, the demand for an additional asset by an investor who holds an efficient portfolio is an increasing function of the asset's *InfRatio*. Also, Brown, Goetzmann, Ibbotson and Ross (1992) discuss the merit of dividing *alpha* by *RMSE* as a way to mitigate the survivorship bias. We estimate whether *InfRatio* is affected by the fund's lagged TR^2 or its *RMSE* and *SDR*, controlling for other fund characteristics.

INSERT TABLE III HERE

The results in Table III show that TR^2 has negative and highly significant effect on the following year fund's *InfRatio*. *RMSE* and *SDR* also predict fund performance with positive and negative coefficients, respectively, which are highly significant. As before, the effect is stronger in the second subperiod than it is in the first. For TR^2, its negative effect here in the first subperiod is more significant than it is in Table II, first period.

Active Share, *AS*, is a positive and highly significant predictor of *InfRatio* for the whole sample (column (3)) and it remains so after including in the model either TR^2 or *RMSE* and *SDR*. However, its coefficient changes signs between the two subperiods, being negative in the first. Overall, TR^2 consistently predicts the fund *Information Ratio* for the whole sample and for the two subperiods, and after controlling for other fund characteristics as well as for *Active Share*. The effects of fund *Expenses* and size (*TNA*) are similar to that in the *alpha* model.

4.3. Why is the predictive power of R^2 stronger in recent years than in early years?

Our results show that during the first nine years of the sample (Period 1), the coefficient of $TR^2_{j,y-1}$ as predictor of $\alpha_{j,y}$ is negative but small and insignificant, while in the second nine-year period (Period 2), the coefficient $TR^2_{j,y-1}$ is more negative and statistically it is highly significant. Notably, there is a big difference in the sample size and data source between the two periods. Period 1 has 3,999 fund years while Period 2 has 12,647 fund years, more than three-fold. The data source for Period 2 is CRSP, which provides broader data. In addition, we propose the following explanation.

We want to measure the relationship between the fund performance ($\alpha_{j,y}$) in year y and the fund's strategy for that year, the *planned* $R^2_{j,y}$, using $R^2_{j,y-1}$ as an estimate of the *planned* $R^2_{j,y}$. This follows, for example, the convention in asset pricing empirical test procedures such as that of Fama and MacBeth (1973) who use lagged portfolio β as an instrument for current β. But if $R^2_{j,y-1}$ is a poor predictor of planned $R^2_{j,y}$, this procedure produces poor results on the relationship between performance and *planned* $R^2_{j,y}$.

Indeed, we observe that in Period 1, $Corr(TR^2_{j,y}, TR^2_{j,y-1})$ is far lower than in Period 2, and therefore in Period 1, $TR^2_{j,y}$ does a poor job predicting $\alpha_{j,y}$. We do an annual regression

$$TR^2_{j,y} = b_{0,y} + b_{1,y}\, TR^2_{j,y-1} + e_{j,y}$$

for y = 1990, 1991, … 2007 and obtain the following results for the *R-sqr* from these regressions:

Period 1, 1990-1998**:** Average *R-sqr* = 0.24. Median *R-sqr* = 0.28.

Period 2, 1999-2007: Average *R-sqr* = 0.62. Median *R-sqr* = 0.70.

This means that in Period 2, $R^2_{j,y-1}$ is a more reliable (less noisy) estimate of the fund's next year's $R^2_{j,y}$. This partly accounts for lagged $R^2_{j,y-1}$ being a stronger predictor of year-y performance in Period 2, as seen in Tables II and III.

We further do the following regression for the entire 18-year period. Define *PERIOD2* = 1 for the years 1999-2007. Then,[8]

[8] The *t*-statistics in the regressions below employ heteroskedasticity-consistent standard errors (White (1980)), clustered by funds.

$$TR^2_{j,y} = 0.524\ TR^2_{j,y-1} + 0.227\ PERIOD2{*}TR^2_{j,y-1} + \text{year fixed effects}$$
$$\qquad (23.73) \qquad\qquad (9.94)$$

The positive and significant coefficient of $PERIOD2{*}TR^2_{j,y-1}$ means that during Period 2, R^2_j is greatly more persistent between the years compared to the persistence in Period 1. We also estimate the model as panel regression with fund fixed effects:

$$TR^2_{j,y} = 0.178\ TR^2_{j,y-1} + 0.136\ PERIOD2{*}TR1^2_{j,y-1} + \text{year fixed effects}$$
$$\qquad (7.67) \qquad\qquad (5.73) \qquad\qquad + \text{fund fixed effects}$$

This shows again a large increase in persistence over time of funds' R^2_j in Period 2 than it is in Period 1. Notably, Period 2 is when funds' $TR^2_{j,y-1}$ strongly predicts year-*y* performance.

4.4. Fund Fixed Effects

Table IV presents estimations with fund *fixed effects*, which effectively remove inter-fund differences that relate to fixed fund characteristics that could account for the negative performance-TR^2 relationship. Here, the hurdle is raised because if a fund has a constant strategy that results in low R^2, its performance will be captured by its fixed effect and will not show as a function of its R^2.

INSERT TABLE IV

The estimation results with fund fixed effect show that TR^2 significantly predicts fund performance, measured either by *alpha* or by *InfRatio*. Higher TR^2 predicts lower performance in the following year, after controlling for fund characteristics, both those that are fixed and those that vary over time. *RMSE* and *SDR* too are significant predictors of fund performance. In this regression, *Expenses* is insignificant because it changes very little for a given fund. The results also show that as the fund becomes larger, its performance deteriorates. The coefficient of *Log*(*TNA*) is negative and significant, but this effect is attenuated as the fund becomes very large, as evident from the positive and significant coefficient on $Log(TNA)^2$. The coefficient of lagged *alpha* is positive and significant while that of lagged *InfRatio* is insignificantly different from zero. The latter result is greatly different from that in Table III, where the coefficient of lagged *InfRatio* is positive and highly significant. The two results can, however, be reconciled. Across funds, we do not control for the unobserved fund strategy, which differs across funds and may be persistent for each fund. Therefore, the noisy estimate of the quality of the fund's strategy, i.e., its lagged performance, positively predicts the fund's future performance. However, for given fund characteristics (controlled by the fund fixed effects), the performance should hover around the mean, and therefore the coefficient on lagged performance should be around zero, as it is for lagged *InfRatio*. The positive coefficient of lagged *alpha* means that a fund with better performance keeps improving it, and an underperforming fund keeps deteriorating. These results, however, are somewhat changed when we change the benchmark portfolios used; see section 4.8 below.

We estimate the effect of Active Share in a fixed-effect regression without TR^2 and obtain that its coefficient in the *alpha* regression is negative and significant. When adding Active Share to the *alpha* regression that includes TR^2, its coefficient is again negative and significant, while TR^2 retains its negative and significant coefficient. When adding Active Share to the *alpha* regression that includes *RMSE* and *SDR*, its effect is negative and significant, while the results for *RMSE* and *SDR* are qualitatively unaltered. In the *InfRatio* equations, Active Share has positive but statistically insignificant coefficient in the fixed-effect regressions.

4.5. Annual cross-sectional regressions (Fama-MacBeth procedure)

We now estimate the predictive power of TR^2 and the pair *RMSE* and *SDR* by the Fama-MacBeth (1973) procedure, performing annual cross-sectional estimates which allow the slope coefficients of the explanatory variables to vary over time. The control variables are the same as in the previous regression, including the style dummy variables.

INSERT TABLE V

The results in Table V are consistent with the previous results although they are not always as statistically significant. TR^2 has a negative predictive effect on *alpha* and its average coefficient is significant at the 6% level. The lower statistical significance may be due to the fact that in this procedure, all years have the same weight regardless of the number of funds in each, while in the pooled panel regression, the estimation results are largely influenced by the number of observations (fund-years) in recent years which is much greater than in earlier years, and it is in recent years that the negative *alpha*-TR^2 is more significant. Still, in a binomial test for the coefficient of TR^2 being negative against the null that it is equally-likely to be positive or negative, the null is rejected at the 0.05 level. Another feature of this procedure that may account for the results is the coefficient of all control variables are allowed to vary between years. In this estimation, only the coefficients of *Expenses* and lagged *alpha* are statistically significant.

Measuring fund performance by *InfRatio*, the coefficient of TR^2 is negative and significant at the 0.01 level. The binomial test too rejects the null hypothesis that the coefficient of TR^2 is equally likely to be positive or negative in favor of the alternative hypothesis that the coefficient of TR^2 is negative.

RMSE and *SDR* have the expected signs – positive and negative, respectively – in both the *alpha* model and in the model of *InfRatio*. However, their coefficients are statistically significant only in the *InfRatio* regression.

4.6. Testing for nonlinearity in the predictive effects of R^2 and of past performance (above and below median).

We now examine non-linearity in the predictive performance of both R^2 and *alpha* or *InfRatio*. In the first year of each two-year pair we divide R^2 into those above and below the median for the year. The dummy variable $HiDUMR^2$ = 1 if R^2 is above the median for the year. Then we split TR^2 into $HiTR^2 = HiDUMR^2 * TR^2$ and its complement $LoTR^2=(1- HiDUMR^2)*TR^2$. We follow the same procedure with *alpha*, splitting it in each year above\below the median into *Hialpha* and *Loalpha*, with the related dummy variable *HiDUMalpha*, and with *InfRatio*, creating the variables *HiDUMInfRatio*, *HiInfRatio* and *LoInfRatio*. We then estimate the models that we have estimated before, replacing TR^2 and *alpha* (or *InfRatio*) by the respective three variables which allow for different intercept and different slope coefficients for values above and below the median.

INSERT TABLE VI

The results in column (1) of Table VI show that in the *alpha* equation, the effects of both TR^2_{y-1} and $alpha_{y-1}$ on $alpha_y$ are non-linear, with their above-median values having weaker predictive effects (in absolute term) than their below-median values. The coefficient of $HiTR^2$, while negative and significant, is less negative than the coefficient of $LoTR^2$, and the coefficient of *Hialpha* is less positive than the coefficient of *Loalpha*. The overall median of R^2_{y-1} is 0.90 and its maximum is 0.989, leaving smaller variance in its above-median values (note, however, that the transformation into TR^2 increases the variance of above-median values). The below-median values of R^2_{y-1} range from 0.24 to 0.90, and it is for this range that there is a more negative predictive effect of TR^2_{y-1} on $alpha_y$. As for lagged *alpha*, $Loalpha_{y\text{-}1}$ has much stronger predictive power on $alpha_y$ than does $Hialpha_{y\text{-}1}$, implying greater persistence of bad performance, a pattern noted by Gruber (1996) who predicts alpha by the rank of lagged alpha.

However, in the *InfRatio* model, column (4), there are no asymmetric effects. The coefficients of $HiTR^2$ and $LoTR^2$ are very similar, both being negative and significant. Nor is there asymmetry in the effects of *HiInfRatio* and *LoInfRatio*, both having positive and significant coefficients which are almost the same. Weighting *alpha* by *RMSE*, which produces *InfRatio*, seems to eliminate the asymmetry in performance prediction.

4.7. Interaction effects of R^2 with alpha and Manager Tenure

We examine the interaction predictive effect of TR^2 with *alpha* and with managerial tenure in column (2). The question is whether the effect of selectivity or idiosyncrasy employed by funds depends on their past performance. The pattern of the mean *alpha* when funds are sorted by their lagged R^2 and *alpha* (Panel A in Tables VIII and IX below) suggests that among the weakly-performing funds, lower R^2 predicts worse *alpha*. We therefore include in the model the interaction term $alpha_{y-1}*TR^2$ or $InfRatio*TR^2$.

Another hypothesis relates to the connection between manager tenure and fund strategy. Chevalier and Ellison (1999) propose that a fund manager’s propensity to take unsystematic risk is positively related to her age, which we translate here to managerial tenure. We therefore add to the model the interaction term $Log(Manager\ Tenure)*TR^2$.

The estimated effects of these two interaction terms are presented in column (2) for the *alpha* model and in column (5) for the *InfRatio* model. The results are:

(a) The coefficients of $alpha*TR^2$ and of $InfRatio*TR^2$ are negative and significant, meaning that that the negative effect of TR^2_{y-1} on performance is stronger for funds that have been performing better in the past year. If managerial skill is positively related to track record of performance (lagged alpha), the results mean that selectivity (low R^2) is more valuable if applied by more skillful managers.

(b) The coefficients of $Log(Manager\ Tenure)*TR^2$ are positive and significant, meaning that the positive effect of selectivity (low TR^2_{y-1}) on performance is stronger in funds with newer managers. The coefficient of *Log(Manager Tenure)* in itself is negative and significant as opposed to being insignificant in Tables II and III, implying a detrimental effect of longevity in the fund on performance. It seems that for a longer-tenure manager, whose performance is worse, it is better to follow the indexes (have higher R^2).

Notably, in both these equations, the negative effect of TR^2 is negative and highly significant.

Finally, columns (3) and (6) combine the models of the non-linear effects of TR^2 and *alpha* or *InfRatio* with the two interaction effects. The results remain qualitatively the same as for each model separately. Focusing on the effect of TR^2, it remains negative and highly significant for both above and below median values, with its effect being attenuated for funds with longer-tenure managers and funds with bad past performance.

4.8. Robustness check: using different benchmark indexes

Our analysis employs the conventional Fama-French benchmark portfolios, which are supposed to mimic unobserved factors. The use of these portfolios in performance evaluation of mutual funds is criticized by Cremers, Petajisto and Zitzewitz (2008) who point out that the small-minus-big portfolio gives equal weight to both its components while the market value of the “small” portfolio is far smaller than the market value of “big.” Similarly, the value-minus-growth portfolio gives equal weights to both while the market value of the former greatly exceeds that of the latter. Also, the benchmark portfolios small-minus-big and high-minus-low book/market involve holding short positions in major portfolios which funds cannot do. Finally, a fund's beta coefficients on the SMB portfolio, for example, constrain the fund's beta on small and big stock portfolios to be the very same in absolute value (but with opposite signs).

We reexamine our results using as benchmarks the Fama-French six long-only portfolios which cover most of the market and are thus feasible benchmarks for mutual funds. The six (2x3) portfolios are based on sorting by size (two groups) into big and small stocks and by book-to-market (three groups) into value, neutral and growth stocks. The average return of each portfolio is value weighted. Then, the betas of a mutual fund on these portfolios reflect the loading of the characteristics of each of these portfolios unto the fund returns, as opposed to being constrained in the traditional benchmarks.

We replicate our analysis by regressing the fund's daily returns on the following eight benchmark return series: The excess return (over the risk-free rate) of the market and of the six Fama-French portfolios, and the Carhart momentum portfolio. We repeat our procedure: We estimate each fund's R^2, *RMSE*, *SDR* and *alpha*, and proceed by regressing the fund's *alpha* or *InfRatio* on the previous year's R^2 (transformed into TR^2) or on the pair *RMSE* and *SDR*, adding control variables (fund characteristics) and lagged performance. The estimation of R^2 being based on eight instead of four indexes, the mean R^2 is slightly higher, 0.87 compared to 0.86 before. The sample selection criteria are similar to those before.

INSERT TABLE VII

Table VII shows that our results are robust to the change in the benchmark portfolios. The table includes cross-section regressions (with year and style dummy variables) and a panel estimation with fund fixed effect. The predictive coefficients of TR^2 in the alpha model (column (1)) and in the *InfRatio* model (Column (3)) are negative and highly significant, as they were in Tables II and III, respectively. Similarly, the coefficients of the pair *RMSE* and *SDR* are positive and negative, respectively, and both are highly significant.

The fund fixed-effect regressions too are qualitatively similar to those reported in Table IV. The coefficient of TR^2 is negative and significant, meaning that a decline in the fund R^2 – or greater idiosyncrasy in investment – improves performance, measured by either *alpha* or *InfRatio*, after controlling for inter-fund differences. Similarly, the coefficient of *RMSE* is positive and that of *SDR* is negative, both being highly significant. The difference in results from the previous analysis pertains to the coefficients of lagged performance. In Table IV, the coefficient of lagged *alpha* is positive and significant and that of lagged *InfRatio* is insignificantly different from zero. Here, the coefficient of lagged *alpha* is insignificant while the coefficient of lagged *InfRatio* is negative and significant, implying reversal in performance over time.

Notably, the coefficients of lagged performance differ from those in the cross section. They are positive in the cross-section regressions and zero or negative in the fixed-effects regressions. The cross-section results mean that better performing funds are more likely to continue to outperform. This may reflect the effect of an unobserved fund characteristic, such as the fund's investment strategy, for which the fund's past performance is a noisy proxy. Once we control for the fund characteristics (including its strategy) by the fund fixed effects, we obtain that better performance in one year does not predict better performance in the following year, or it even shows a reversal in performance (when using *InfRatio*). This means that a fund strategy produces some average level of performance which is reverted to over time.

Finally, follow the recommendation of Cremers, Petajisto and Zitzewitz (2008) and use the following benchmark portfolios: the excess return on the market, midcap index, small cap index, three value factors (large, mid and small) and the momentum factor. We use these benchmark portfolios to estimate *alpha* and R^2 and then we estimate a panel regression model of $alpha_{j,y}$ on $TR^2_{j,y-1}$ and the other control variables that appear in our analysis. The resulting coefficient of $TR^2_{j,y-1}$ is –0.736 with t = 7.44. Again, our predictive measure R^2 is robust to the set of benchmark used to evaluate performance.

5. Fund performance based on sorting on lagged R^2 and lagged performance

We identify a group of funds which generate significant positive performance. In each year *y* we sort funds into five portfolios by their R^2 in *y-1* and within each quintile we sort the funds into five portfolios by their *alpha* (or *InfRatio*) in *y-1*. Then, we estimate the average year-*y* *alpha* (or *InfRatio*) for all funds that are included in each of the 25 portfolios.

INSERT TABLE VIII

Panel A of Table VIII reports the average portfolio *alpha* and Panel B reports the average portfolio *InfRatio*. In Panel A, average $alpha_y$ increases in $alpha_{y\text{-}1}$ and decreases in $R^2_{y\text{-}1}$, as in the regressions. The lowest $R^2_{y\text{-}1}$-highest $alpha_{y\text{-}1}$ portfolio produces annual alpha of 2.81% with t = 5.84, and the lowest $R^2_{y\text{-}1}$-next to highest $alpha_{y\text{-}1}$ portfolio had average *alpha* of 1.05% with t = 2.91. Also, among the highest $alpha_{y\text{-}1}$ portfolios, the average alpha on the lowest R^2 portfolio exceeds the average alpha on the highest R^2 portfolio by 3.62% (t = 6.96). The difference in the mean *alpha* between low and high R^2 quintile portfolios is positive for the four highest quintile portfolios of *alpha*, with the difference being significant for the top three *alpha* quintiles. However, in the bottom-performing funds, measured by low $alpha_{y\text{-}1}$, low $R^2_{y\text{-}1}$ predicts worse rather than better performance. Perhaps in such funds, low R^2 does not indicate selectivity but rather unreasonable idiosyncratic bets.

The results for *InfRatio* as a performance measure are qualitatively similar. Performance is decreasing in $R^2_{y\text{-}1}$ and it increases with $InfRatio_{y\text{-}1}$. The portfolio of funds with the highest $InfRatio_{y\text{-}1}$ and lowest $R^2_{y\text{-}1}$ has a positive $InfRatio_y$, 0.02, with t = 6.16. Here, unlike the case of the *alpha*-sorted funds, the average $InfRatio_y$ is monotonically and significantly decreasing in $R^2_{y\text{-}1}$ for all five $InfRatio_{y\text{-}1}$ quintile portfolios, even for the worst-performing funds by $InfRatio_{y\text{-}1}$.

INSERT TABLE IX HERE

We repeat the above analysis doing *independent* sorting on $R^2_{y\text{-}1}$ and on $alpha_{y\text{-}1}$. The results, presented in Table IX, Panel A, are qualitatively the same. There are two low-$R^2_{y\text{-}1}$ portfolios, with the forth and fifth highest $alpha_{y\text{-}1}$, that have positive and significant $alpha_y$. In particular, the portfolio of the lowest $R^2_{y\text{-}1}$ and highest $alpha_{y\text{-}1}$ has average $alpha_y$ of 2.235% with t = 6.21. For the four higher quintile portfolios of $alpha_{y\text{-}1}$, the average *alpha* of the lowest $R^2_{y\text{-}1}$ portfolio is significantly higher than that of the highest $R^2_{y\text{-}1}$ portfolio. The results for *InfRatio* (Panel B) are qualitatively similar, with the average performance of low-R^2 portfolios being higher than that of the high-R^2 portfolios for all five $InfRatio_{y\text{-}1}$ quintiles. The fund portfolio of the highest-$InfRatio_{y\text{-}1}$ and lowest $R^2_{y\text{-}1}$ has average $InfRatio_y$ of 0.016 with t = 6.03.

6. Factors related to funds' R^2

We suggest that a fund chooses a strategy, such as the extent of selectivity that we measure by R^2, which subsequently affects its performance. We now examine whether there are systematic fund characteristics that are associated with the fund's R^2 by regressing TR^2 on *lagged* fund characteristics.

INSERT TABLE X HERE

The results in Table X show that the funds with high expenses have lower R^2, as evident from the negative and significant coefficient of *Expenses* in model (1). While *Expenses* is lagged, it is quite persistent so the estimated relationship suggests persistent fund policy on expenses and strategy regarding selectivity. More actively-managed funds expend more resources on selectivity and thus incur higher expenses, and at the same time investors are willing to pay more for investing in these funds because of their superior performance. In the fixed-effect regression (model (2)), the coefficient of *Expenses* is practically zero reflecting almost no change over time in the expense ratio that is related to R^2. The positive coefficient of *Log(TNA)* in both models means that larger funds hold broader and more diversified portfolio, which increases their R^2. As the fund size increases, so does its R^2. Another explanation is due to Koijen's (2008) model of fund managers who derive utility from improving their ranking or status by raising their fund size. He proposes that managers of smaller fund that have room to grow and provide better status have an incentive to "deviate from the pack" and employ active investment strategy. Here, it means that smaller fund employ less benchmark-based policy and more idiosyncratic policy, producing a positive relationship between *TNA* and R^2. This relationship is weaker as the fund size grows, following the negative coefficient of $Log(TNA)^2$ (significant only in Model (1)).

Older funds (higher *Age*) have lower R^2 after controlling for other characteristics, including fund size which usually grows with fund age. This result suggests that one reason for fund longevity is its greater selectivity (lower R^2) which produces better performance. There is negative relationship between R^2 and *Managerial Tenure*, which is consistent with Chevalier and Ellison's (1999, p. 391) suggestion that younger managers tend to herd or "avoid unsystematic risk when selecting their portfolio." Here it means larger R^2 – which is greater proportion of the risk due to systematic risk – for managers with lower tenure.

The effect of past performance (*alpha*) on fund's R^2 has different sign in the two models. Model (1) mostly reflects the R^2-*alpha* relationship *across* funds. Stable and persistent strategy and performance by funds produces a negative R^2-*alpha* relationship across funds regardless of which variable lags the other.

The estimated R^2-lagged *alpha* relationship in the *fixed-effect* Model (2) shows that better performing funds tend to reduce their level of selectivity and do more indexing. (Notably, we control for fund size that is itself affected by past performance.) Our result is consistent with Elton, Gruber and Blake's (2003, p. 785) proposition that funds adopt the following strategies: "Greater risk-taking after a period of underperforming the benchmarks," and "Less risk-taking after periods of outperforming the benchmarks." We obtain that outperforming funds subsequently increase the extent of indexing, or choose higher R^2, thus taking less idiosyncratic risk. This helps preserve their past good performance in multi-year performance comparisons, although this strategy is costly: higher R^2 lowers future performance. And, funds that underperform subsequently take more idiosyncratic strategy – choose lower R^2 – in the hope of hitting a successful strategy. For such funds, indexing will preserve their past bad performance and will not give them a chance to improve. The results on the R^2-lagged performance relationship are qualitatively similar when using *InfRatio* instead of *alpha*, and are not reported here.

The estimation model of the determinants of R^2 includes year dummy variables. The coefficients of these variables generally increase over time. In the first nine years, the average R^2 is lower than it is in the last nine years, suggesting a growing propensity to follow the indexes over time. The last nine-years, which use different data source and produce more significant negative relationship between performance and R^2, seem to have different characteristics.

7. Conclusions

We propose an intuitive and convenient measure of mutual fund selectivity or active management: the R^2 from a regression of fund return on the Fama-French (1993) and Carhart (1997) factors. We find that the fund R^2, estimated from one year's daily returns, predicts the following year's fund performance, measured either by the fund's *alpha* or by its *Information Ratio (InfRatio)*, which is the fund *alpha* scaled by the regression's *RMSE*. The predictive coefficient of R^2 is negative and highly significant. That is, lower R^2 or greater fund selectivity predicts better performance. We also obtain that the pair of volatility measures which constitute R^2, *RMSE* and return standard deviation *SDR* (their squared ratio equals 1-R^2) predict funds' subsequent performance, with positive and negative coefficients, respectively, which are highly significant. These effects are obtained after controlling for commonly-used fund characteristics. Our results are the same in both cross-sectional analysis of funds, with year and style fixed effects, and in panel regressions with fund fixed effects in addition to year and style fixed effects.

Our analysis is shown to be robust to the indexes used. We re-estimate our model using eight benchmark portfolios: the excess return on the market, the excess returns on the Fama-French six size-by-value\growth portfolios and Carhart's momentum factors. The results are qualitatively unchanged. In particular, R^2 has a negative and highly significant coefficient in regressions where the dependent variable is the following year's *alpha* or *InfRatio*.

We obtain that the negative predictive effect of R^2 is particularly strong for funds with better past performance and for funds whose manager's tenure is shorter. We also find that in predicting *alpha*, the effect of R^2 is greater for funds whose R^2 is in the lower range of its distribution. In predicting *InfRatio*, the negative predictive effect of R^2 is similar in both the high and low range of its distribution across funds.

Using our strategy, it is possible to identify a portfolio of funds that produces positive and significant performance, measured either by *alpha* or by its *InfRatio*. We sort at the end of each year funds by their R^2 and by their past *alpha* and invest in funds that are in the bottom quintile of R^2 and the highest quintile of *alpha*. The resulting portfolio has an average annual alpha of 2.81% with t = 5.84. Similar results are obtained when replacing *alpha* by the *InfRatio*. The results are similar for independent sorting of funds by their R^2 and past performance.

Fund R^2 is negatively related to another measure of active fund management and selectivity developed by Cremers and Petajisto (2008), called Active Share, the sum of absolute differences between the portfolio holdings of the fund and its benchmark portfolio. When including Active Shares in the model that predicts performance, the predictive effect of R^2 remains negative and highly significant.

R^2 is related to identifiable fund characteristics. It is negatively related to expenses, fund age and manager tenure, and positively related to fund size. Importantly, funds react to performance by changing their R^2. Across funds, funds with better past performance also have lower R^2. However, in a fund fixed-effect analysis we find that the effect of past performance on R^2 is positive, meaning that following better performance, funds tend to index more to preserve their rank, while worse-performing funds increase subsequently their idiosyncratic risk.

Altogether, this study offers a new convenient way to predict mutual fund performance using only their return data.

References

Bodie, Z., Kane, A. and A.J. Marcus, 2009, *Investments*, 8th edition, McGraw-Hill Irwin

Brands, S., Brown, S.J. and D.R. Gallagher, 2006, Portfolio Concentration and Investment Manager Performance, *International Review of Finance*, 149-174

Brown, S.J. and W.N. Goetzmann, 1995, Performance Persistence, *Journal of Finance* 50, 679-698.

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

Brown, S.J., W.N. Goetzmann, and S. Ross, 1995, Survival, *Journal of Finance* 50, 853-873.

Brown, S.J., W.N. Goetzmann, R.G. Ibbotson and S. Ross, 1992, Survivorship Bias in Performance Studies, *Review of Financial Studies* 5, 553-580.

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

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

Cremers, M. and A. Petajisto, 2008, How Active is Your Fund Manager? A New Measure That Predicts Performance, *Review of Financial Studies*, forthcoming

Cremers, M., Petajisto, A. and E. Zitzewitz, 2008, Should Benchmark Indices Have Alpha? Revisiting Performance Evaluation, Working Paper, Yale School of Management

Daniel, K., M. Grinblatt, S. Titman, and R. Wermers, 1997, Measuring Mutual Fund Performance with Characteristic-Based Benchmarks, *Journal of Finance*, 52, 1035-1058

Dimson, E., 1979, Risk Measurement When Shares Are Subject to Infrequent Trading, *Journal of Financial Economics*, 7, 197-226

Elton, E. J., M. J. G., and C. R. Blake, 1996, Survivorship Bias and Mutual Fund Performance, *Review of Financial Studies*, 9, 1097-1120

Elton, E. J., M. J. G., and C. R. Blake, 2003, Incentive Fees and Mutual Funds, *Journal of Finance*, 58, 779-804

Evans, R. B., 2006, Does Alpha Really Matter? Evidence from Mutual Fund Incubation, Termination and Manager Change. Working paper

Fama, E.F., 1972, Components of Investment Performance, *Journal of Finance* 27, 551-567.

Fama, E.F. and J.D. MacBeth, 1973, Risk, Return and Equilibrium: Empirical Tests, *Journal of Political Economy*, 71, 607-636

Fama, E.F. and K.R. French, 1993. Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics* 33, 3–56.

Fung, W. and D. Hsieh, 2001, The Risk in Hedge Fund Strategies: Theory and Evidence from Trend Followers, *Review of Financial Studies* 14, 313-341.

Fung, W. and D. Hsieh, 2004, Hedge Fund Benchmarks: A Risk Based Approach, *Financial Analysts Journal* 60, 65-80

Goetzmann, W.N., Z. Ivkovic, and K.G. Rouwenhorst, 2001, Day Trading International Mutual Funds: Evidence and Policy Solutions, *Journal of Financial and Quantitative Analysis*, 36, 287-309

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

Jensen, M. C., 1968, The Performance of Mutual Funds in the Period 1945-1964, *Journal of Finance,* 23, 389-416.

Kacperczyk, M.T., C. Sialm, and L. Zheng, 2005, On Industry Concentration of Actively Managed Equity Mutual Funds, *Journal of Finance,* 60, 1983-2011.

Kacperczyk, M.T. and A. Seru, 2007, Fund Manager Use of Public Information: New evidence on managerial skills, *Journal of Finance* 62, 485-528.

Koijen, R.S.J., 2008, The Cross-Section of Managerial Ability and Risk Preference, Working Paper, University of Chicago.

Mamaysky, H., M. Spiegel, and H. Zhang, 2007, Estimating the Dynamics of Mutual Fund Alphas and Betas, *Review of Financial Studies,* 21, 233-264

Titman, S. and C. Yiu, 2008, Do the Best Hedge Funds Hedge? Working Paper.

Treynor, J. L., and F. Black, 1973, How to Use Security Analysis to Improve Portfolio Selection, *Journal of Business* 46, 66–86.

Wang, A. and L. Zheng, 2008, Strategy Distinctiveness and Hedge Fund Performance. Working Paper.

Wermers, R., 2003, Are Mutual Fund Shareholders Compensated for Active Management "Bets"? Working Paper.

White, H., 1980, A Heteroscedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity, *Econometrica* 48, 817-838

Table I. Summary statistics

Statistics on actively managed equity mutual funds included in our sample. The performance measure *alpha* is the intercept from an annual regression of daily fund excess returns on the factor returns *RM-Rf*, *SMB*, *HML* and *MOM* (momentum), and their lagged values. R^2 is obtained from the above regression, and $TR^2 = log(\sqrt{R^2}/(1- \sqrt{R^2}))$. The Total Net Assets (*TNA*) in $mm, *Expenses* and *Turnover* are as of the end of the year. *Age* is fund age, the number of years since the fund was first offered. *Tenure* is the tenure of the manager, the number of years since the current manager took control. *AS* is *Active Share* measure from Cremers and Petajisto (2008). The sample period is from 1/1989 to 12/2007.

Panel A: Fund characteristics

	Mean	Median	Minimum	Maximum
Total number of funds: 2,314				
TNA (total net assets, in $millions)	1,501.65	266.69	15.1	193,453.1
Age (years)	13.65	8.92	0.7	84.92
Expenses (%)	1.27	1.23	0.01	4.54
Turnover (%)	89.38	66.00	0.20	3,603
Manager Tenure (years)	3.44	2.82	0.08	45.08
Alpha (%)	-0.92	-1.11	-131.04	149.00
R^2	0.86	0.90	0.240	0.989
TR^2	2.88	2.95	-0.026	5.158

Panel B: Correlations (**: 1% significance, *: 5% significance)

	Log(TNA)	*Age*	*Expenses*	*Turnover*	*Log(Manager Tenure)*	*Alpha*	R^2	TR^2
Log(TNA)	1.00							
Age	0.35**	1.00						
Expenses	-0.32**	-0.23**	1.00					
Turnover	-0.12**	-0.09**	0.19**	1.00				
Log(Manager Tenure)	0.12**	0.09**	-0.06**	-0.07**	1.00			
Alpha	0.05**	-0.03**	-0.04**	-0.03**	-0.04**	1.00		
R^2	0.13**	-0.02*	-0.10**	-0.02*	0.09**	-0.07**	1.00	
TR^2	0.15**	-0.002	-0.13**	-0.04**	0.10**	-0.09**	0.92**	1.00

Table II. Predictive regressions of fund performance: Four-factor *alpha*

Panel regressions of *alpha*, the intercept from an annual regression of daily fund excess returns on the factor returns *RM-Rf*, *SMB*, *HML* and *MOM* (momentum) and their lagged values. All independent variables are as of the end of the previous year. $TR^2 = log(\sqrt{R^2}/(1- \sqrt{R^2}))$, where R^2 is obtained from the above regression. *RMSE* is the root mean squared error from this regression and *SDR* is the standard deviation of the daily fund excess returns. The Total Net Assets (*TNA*) in $mm, *Expenses* and *Turnover* are as of the end of the year. *Age* is fund age, the number of years since the fund was first offered. *Tenure* is the tenure of the manager, the number of years since the current manager took control. *AS* is *Active Share* measure from Cremers and Petajisto (2008). Each regression also includes year and style dummy variables. The *t*-statistics in parentheses are based on standard errors clustered by fund. The sample period is from 1/1990 to 12/2007.

Variables	**1990-2007**					**1990-1998**		**1999-2007**	
lagged one year	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
TR^2	-0.680			-0.736		-0.030		-0.944	
	(7.69)			(5.90)		(0.17)		(9.56)	
RMSE		3.955			4.330		0.390		4.451
		(6.33)			(5.86)		(0.32)		(5.60)
SDR		-6.688			-6.990		-2.205		-7.204
		(19.28)			(16.45)		(1.95)		(20.19)
Expenses	-0.927	-0.523	-0.622	-0.694	-0.343	-0.713	-0.654	-1.043	-0.490
	(5.63)	(3.19)	(3.02)	(3.39)	(1.71)	(1.83)	(1.68)	(6.08)	(2.91)
Log(TNA)	-0.678	-0.479	-0.829	-0.737	-0.576	-0.901	-0.845	-0.578	-0.387
	(3.26)	(2.30)	(3.62)	(3.23)	(2.50)	(1.63)	(1.54)	(2.69)	(1.77)
$Log(TNA)^2$	0.048	0.035	0.057	0.051	0.040	0.071	0.067	0.038	0.025
	(2.98)	(2.12)	(3.22)	(2.91)	(2.25)	(1.62)	(1.54)	(2.25)	(1.46)
Turnover	-0.002	-0.0001	-0.002	-0.002	0.001	0.001	0.002	-0.003	-0.001
	(1.68)	(0.01)	(0.65)	(0.74)	(0.55)	(0.19)	(0.39)	(2.90)	(0.58)
Fund Age	-0.002	0.003	0.004	0.002	0.006	-0.018	-0.015	0.007	0.010
	(0.34)	(0.53)	(0.77)	(0.46)	(1.24)	(1.88)	(1.49)	(1.25)	(1.71)
Log(Manager	-0.090	-0.079	-0.136	-0.147	-0.127	-0.125	-0.126	-0.073	-0.052
Tenure)	(1.78)	(1.57)	(2.32)	(2.52)	(2.19)	(1.05)	(1.07)	(1.40)	(1.01)
Alpha	0.172	0.172	0.184	0.184	0.187	0.183	0.173	0.166	0.173
	(11.84)	(11.82)	(13.63)	(13.58)	(13.86)	(6.27)	(6.07)	(9.78)	(9.87)
AS			1.579	-0.691	-0.999				
			(3.01)	(1.15)	(1.84)				
N of funds	2,314	2,314	1,890	1,890	1,890	871	871	2,177	2,177
Fund-years	16,646	16,646	13,204	13,204	13,204	3,999	3,999	12,647	12,647
R^2	0.17	0.21	0.19	0.19	0.23	0.08	0.08	0.21	0.26

Table III. Predictive regressions of fund performance: *Information Ratio*

Panel regressions of the *Information Ratio*, *InfRatio = alpha/RMSE*, where *alpha* is the intercept from an annual regression of daily fund excess returns on the factor returns *RM-Rf*, *SMB*, *HML* and *MOM* (momentum) and their lagged values, and *RMSE* is the root mean squared error from this regression. $TR^2 = log(\sqrt{R^2}/(1-\sqrt{R^2}))$, where R^2 is obtained from the above regression and *SDR* is the standard deviation of the daily fund excess returns. All independent variables are as of the end of the previous year. The Total Net Assets (*TNA*) in $mm, *Expenses* and *Turnover* are as of the end of the year. *Age* is fund age, the number of years since the fund was first offered. *Tenure* is the tenure of the manager, the number of years since the current manager took control. *AS* is *Active Share* measure from Cremers and Petajisto (2008). Each regression also includes year and style dummy variables. The *t*-statistics in parentheses are based on standard errors clustered by fund. The sample period is from 1/1990 to 12/2007.

Variables	**1990-2007**					**1990-1998**		**1999-2007**	
lagged one year	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
TR^2	-0.010			-0.006		-0.003		-0.007	
	(12.77)			(5.16)		(1.93)		(5.18)	
RMSE		0.051			0.030		0.010		0.032
		(12.05)			(6.01)		(1.01)		(5.16)
SDR		-0.053			-0.045		-0.019		-0.048
		(21.30)			(15.76)		(2.16)		(15.83)
Expenses	-0.012	-0.009	-0.012	-0.012	-0.010	-0.008	-0.007	-0.014	-0.011
	(8.29)	(6.10)	(6.57)	(6.92)	(5.66)	(2.28)	(2.04)	(6.80)	(5.48)
Log (TNA)	-0.005	-0.004	-0.007	-0.006	-0.005	-0.006	-0.006	-0.006	-0.005
	(2.62)	(2.12)	(3.01)	(2.68)	(2.26)	(1.33)	(1.25)	(2.28)	(1.93)
$Log(TNA)^2$	0.00	0.00	0.0001	0.00	0.00	0.00	0.00	0.0004	0.00
	(2.46)	(2.01)	(2.63)	(2.39)	(2.03)	(0.97)	(0.88)	(2.09)	(1.81)
Turnover	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	0.00	-0.00	-0.00
	(3.43)	(1.69)	(1.51)	(1.63)	(0.28)	(0.01)	(0.28)	(3.35)	(1.07)
Fund Age	-0.00	0.00	0.00	0.00	0.00	-0.00	-0.00	0.00	0.00
	(0.20)	(0.41)	(0.35)	(0.11)	(0.60)	(0.85)	(0.60)	(0.54)	(0.84)
Log(Manager	-0.00	-0.00	-0.001	-0.001	-0.001	-0.00	-0.00	-0.001	-0.001
Tenure)	(0.80)	(0.47)	(1.41)	(1.58)	(1.34)	(0.44)	(0.34)	(1.61)	(1.36)
InfRatio	0.162	0.158	0.159	0.156	0.152	0.183	0.180	0.142	0.138
	(20.14)	(20.03)	(17.19)	(16.85)	(16.55)	(10.24)	(10.12)	(12.72)	(12.53)
AS			0.051	0.034	0.034	-0.033	-0.029	0.044	0.048
			(9.81)	(5.55)	(6.29)	(2.60)	(2.33)	(6.07)	(7.78)
N of funds	2,314	2,314	1,890	1,890	1,890	727	727	1,812	1,812
Fund-years	16,646	16,646	13,204	13,204	13,204	3,282	3,282	9,922	9,922
R-sqr	0.17	0.19	0.18	0.18	0.20	0.12	0.12	0.20	0.22

Table IV. Predictive regressions of fund performance: Estimation with fund fixed effects in

Panel regressions with fund fixed-effects. The dependent variables are *alpha*, the intercept from an annual regression of daily fund excess returns on the factor returns *RM-Rf*, *SMB*, *HML* and *MOM* (momentum) and their lagged values, and *InfRatio* = *alpha*/*RMSE*, where *RMSE* is the root mean squared error from this regression. ll independent variables are as of the end of the previous year. $TR^2 = log(\sqrt{R^2}/(1- \sqrt{R^2}))$, where R^2 is obtained from the above regression. *SDR* is the standard deviation of the daily fund returns. The Total Net Assets (*TNA*) in $mm, *Expenses* and *Turnover* are as of the end of the year. *Age* is fund age, the number of years since the fund was first offered. *Tenure* is the tenure of the manager, the number of years since the current manager took control. Each regression also includes year and style dummies, and *t*-statistics in parentheses are based on standard errors clustered by fund. The sample period is from 1/1990 to 12/2007.

Variables lagged one year	Dependent variables			
	alpha		*InfRatio*	
	(1)	(2)	(3)	(4)
TR^2	-0.420		-0.003	
	(3.21)		(2.93)	
RMSE		3.851		0.028
		(4.94)		(5.27)
SDR		-7.578		-0.041
		(12.89)		(12.52)
Expenses	0.213	-0.163	-0.003	-0.005
	(0.44)	(0.33)	(0.73)	(1.17)
Log(TNA)	-3.985	-3.455	-0.028	-0.025
	(9.22)	(8.51)	(7.49)	(6.89)
$Log(TNA)^2$	0.126	0.112	0.00	0.00
	(3.78)	(3.54)	(1.39)	(1.18)
Turnover	0.003	0.004	0.00	0.00
	(1.72)	(2.08)	(1.16)	(1.36)
Fund Age	-0.005	0.004	0.00	0.00
	(0.20)	(0.17)	(0.43)	(0.67)
Log(Manager Tenure)	-0.039	-0.028	-0.00	-0.00
	(0.60)	(0.43)	(0.28)	(0.19)
Alpha	0.051	0.068		
	(3.05)	(3.86)		
InfRatio			0.004	0.005
			(0.43)	(0.64)
N of funds	2,314	2,314	2,314	2,314
Fund-years	16,646	16,646	16,646	16,646
R-sqr	0.21	0.24	0.19	0.20

Table V. Fama-MacBeth regressions of fund performance

The dependent variables are *alpha*, the intercept from an annual regression of daily fund excess returns on the factor returns *RM-Rf*, *SMB*, *HML* and *MOM* (momentum) and their lagged values and *InfRatio* = *alpha*/*RMSE*, where *RMSE* is the root mean squared error from this regression. All independent variables are as of the end of the previous year. $TR^2 = log(\sqrt{R^2}/(1-\sqrt{R^2}))$, where R^2 is obtained from the above regression. *SDR* is the standard deviation of the daily fund returns. The Total Net Assets (*TNA*) in $mm, *Expenses* and *Turnover* are as of the end of the year. *Age* is fund age, the number of years since the fund was first offered. *Tenure* is the tenure of the manager, the number of years since the current manager took control. The numbers presented are the means of the annual coefficients, the *t*-statistics (in parentheses) and (in brackets) the probability of the null hypothesis being rejected. At the bottom three lines, the numbers in brackets are the probability under null of the coefficient being equally likely positive or negative. The sample period is from 1/1990 to 12/2007.

Variables lagged one year	Dependent variable			
	alpha	*alpha*	*InfRatio*	*InfRatio*
TR^2	-0.471		-0.008	
	(2.02)		(3.02)	
	[0.060]		[0.008]	
RMSE		1.070		0.041
		(1.01)		(2.84)
		[0.328]		[0.011]
SDR		-2.674		-0.035
		(1.77)		(2.33)
		[0.095]		[0.033]
Expenses	-0.874	-0.664	-0.012	-0.010
	(2.58)	(2.17)	(4.03)	(3.58)
	[0.019]	[0.044]	[0.001]	[0.002]
Log(TNA)	-0.326	-0.266	-0.003	-0.003
	(1.10)	(0.92)	(1.13)	(1.01)
	[0.288]	[0.369]	[0.275]	[0.326]
$Log(TNA)^2$	0.026	0.020	0.0003	0.0002
	(1.07)	(0.86)	(0.98)	(0.86)
	[0.301]	[0.403]	[0.342]	[0.404]
Turnover	0.001	0.002	-0.00	-0.00
	(0.35)	(0.71)	(0.78)	(0.45)
	[0.727]	[0.490]	[0.447]	[0.660]
Fund Age	-0.009	-0.006	-0.0001	-0.00
	(1.23)	(0.93)	(1.18)	(0.84)
	[0.237]	[0.366]	[0.252]	[0.414]
Log(Manager Tenure)	-0.352	-0.344	-0.006	-0.006
	(1.21)	(1.13)	(1.31)	(1.32)
	[0.242]	[0.273]	[0.208]	[0.204]
Dependent variable,	0.149	0.138	0.164	0.159
Lagged	(4.14)	(4.69)	(6.35)	(6.35)
	[0.001]	[0.000]	[0.000]	[0.000]
R-sqr	0.214	0.246	0.242	0.259
TR^2: pos/neg	5/13		3/15	
(prob. under null)	[0.048]		[0.004]	
RMSE: pos/neg		8/10		12/6
(prob. under null)		[0.407]		[0.119]
SDR: pos/neg		5/13		4/14
(prob. under null)		[0.048]		[0.015]

Table VI: Predictive Regressions of fund performance, with (1) high and low split of lagged *TR^2*, *alpha* and *InfRatio* and with (2) interaction between variables

Panel regressions of *alpha*, the intercept from an annual regression of daily fund excess returns on the factor returns *RM-Rf*, *SMB*, *HML* and *MOM* (momentum) and their lagged values, and *InfRatio* = *alpha/RMSE*. All independent variables are as of the end of the previous year. $TR^2 = log(\sqrt{R^2}/(1- \sqrt{R^2}))$, where R^2 is obtained from the above regression. The Total Net Assets (*TNA*) in \$mm, *Expenses* and *Turnover* are as of the end of the year. *Age* is fund age, the number of years since the fund was first offered. *Tenure* is the tenure of the manager, the number of years since the current manager took control. $HiDUMR^2$ equals 1 if in year *y-1* the fund's R^2 is higher than the cross-sectional median R^2 for that year (zero otherwise). $HiTR^2=TR^2*HiDUMR^2$ and $LoTR^2=TR^2*(1-HiDUMR^2)$. The dummies specifications for *Alpha (Lo/Hi)* and *InfRatio (Lo/Hi)* are defined in a similar way. Each regression also includes year and style dummy variables. The *t*-statistics in parentheses are based on standard errors clustered by fund. The sample period is from 1/1990 to 12/2007.

Variables lagged one year	Dependent variable					
		alpha			*InfRatio*	
	(1)	(2)	(3)	(4)	(5)	(6)
TR²		-0.898			-0.011	
		(9.55)			(13.52)	
HiTR²	-0.374		-0.716	-0.012		-0.014
	(2.64)		(4.78)	(7.57)		(8.60)
LoTR²	-1.150		-1.273	-0.009		-0.009
	(6.56)		(7.34)	(6.06)		(6.32)
HiDUMR²	-2.476		-1.858	0.008		0.014
	(4.44)		(3.40)	(1.68)		(2.70)
Expenses	-0.781	-0.940	-0.784	-0.012	-0.012	-0.012
	(4.81)	(5.66)	(4.81)	(8.40)	(8.31)	(8.44)
Log(TNA)	-0.579	-0.647	-0.547	-0.005	-0.005	-0.005
	(2.82)	(3.09)	(2.65)	(2.66)	(2.62)	(2.66)
Log(TNA)²	0.041	0.046	0.039	0.00	0.00	0.000
	(2.57)	(2.80)	(2.39)	(2.50)	(2.43)	(2.49)
Turnover	-0.001	-0.003	-0.002	-0.00	-0.00	-0.000
	(1.00)	(1.86)	(1.13)	(3.46)	(3.52)	(3.53)
Fund Age	-0.00	-0.002	-0.00	-0.000	-0.00	-0.000
	(0.10)	(0.36)	(0.10)	(0.20)	(0.14)	(0.17)
Log(Manager Tenure)	-0.097	-1.313	-1.347	-0.00	-0.003	-0.004
	(1.91)	(6.41)	(6.52)	(0.80)	(2.47)	(2.68)
*Log(Manager Tenure)*TR²*		0.423	0.433		0.001	0.001
		(7.00)	(7.09)		(2.43)	(2.65)
Alpha		0.324				
		(7.91)				
*Alpha*TR²*		-0.068	-0.072			
		(4.74)	(4.79)			
Hialpha	0.082		0.220			
	(3.15)		(5.58)			
Loalpha	0.330		0.480			
	(11.05)		(9.29)			
HiDUMalpha	-0.721		-0.328			
	(3.23)		(1.57)			
InfRatio					0.255	
					(10.13)	
*InfRatio*TR²*					-0.033	-0.039
					(3.99)	(4.55)
HiInfRatio				0.180		0.275
				(10.93)		(9.98)
LoInfRatio				0.170		0.290
				(9.76)		(9.04)
HiDUMInfRatio				-0.002		-0.002
				(1.33)		(1.28)
N of funds	2,314	2,314	2,314	2,314	2,314	2,314
Fund-years	16,646	16,646	16,646	16,646	16,646	16,646
R^2	0.18	0.18	0.19	0.17	0.17	0.17

Table VII: Predictive regressions of fund performance using different Fama-French-Carhart indexes

Panel regressions of *alpha*, the intercept from an annual regression of daily fund excess returns on 8 factors: the market factor, the Fama-French 6 portfolios classified as 2x3 by size and by book-to-market, and Carhart's momentum factor, and their lagged values. All independent variables are as of the end of the previous year. $TR^2 = log(\sqrt{R^2}/(1-\sqrt{R^2}))$, where R^2 is obtained from the above regression. *RMSE* is the root mean squared error from this regression and *SDR* is the standard deviation of the daily fund returns over the year. *Information Ratio*, *InfRatio* = *alpha*/*RMSE* The Total Net Assets (*TNA*) in $mm, *Expenses* and *Turnover* are as of the end of the year. *Age* is fund age, the number of years since the fund was first offered. *Tenure* is the tenure of the manager, the number of years since the current manager took control. Each regression also includes year and style dummy variables. The *t*-statistics in parentheses are based on standard errors clustered by fund. The sample period is from 1/1990 to 12/2007.

Variables lagged one year	Panel Regression				Panel regression with fund fixed effects			
	Alpha		*InfRatio*		*Alpha*		*InfRatio*	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
TR^2	-0.770		-0.012		-0.380		-0.005	
	(8.61)		(15.33)		(2.88)		(4.60)	
RMSE		5.547		0.062		5.406		0.039
		(7.89)		(13.93)		(6.62)		(7.50)
SDR		-4.142		-0.038		-6.100		-0.037
		(10.96)		(15.04)		(8.90)		(10.87)
Expenses	-0.534	-0.389	-0.010	-0.008	0.478	0.251	0.00	-0.001
	(3.13)	(2.26)	(6.55)	(5.12)	(0.87)	(0.45)	(0.09)	(0.19)
Log(TNA)	-0.397	-0.286	-0.004	-0.004	-3.160	-2.761	-0.025	-0.022
	(1.88)	(1.32)	(2.04)	(1.88)	(7.63)	(6.69)	(6.67)	(6.11)
$Log(TNA)^2$	0.037	0.030	0.00	0.00	0.103	0.094	0.00	0.00
	(2.22)	(1.73)	(2.41)	(2.21)	(3.15)	(2.92)	(1.08)	(0.92)
Turnover	-0.00	0.001	-0.00	-0.00	0.005	0.005	0.00	0.00
	(0.03)	(0.64)	(0.96)	(0.15)	(2.76)	(2.91)	(2.60)	(2.76)
Fund Age	-0.003	-0.003	-0.00	0.00	0.002	0.007	0.00	0.00
	(0.65)	(0.47)	(0.05)	(0.24)	(0.07)	(0.32)	(0.96)	(1.08)
Log(Manager Tenure)	-0.059	-0.054	-0.00	-0.00	0.002	0.012	0.00	0.00
	(1.15)	(1.04)	(0.40)	(0.10)	(0.03)	(0.18)	(0.39)	(0.48)
Alpha	0.112	0.125			-0.022	0.003		
	(6.65)	(7.05)			(1.18)	(0.14)		
InfRatio			0.152	0.162			-0.025	-0.019
			(18.71)	(19.83)			(2.91)	(2.23)
N of funds	2,312	2,312	2,312	2,312	2,312	2,312	2,312	2,312
Fund-years	16,646	16,646	16,646	16,646	16,646	16,646	16,646	16,646
R^2	0.12	0.13	0.14	0.14	0.13	0.15	0.12	0.12

Table VIII. Fund Performance, sorting on R^2 and *alpha\InfRatio*

The table presents the average portfolio *alphas* or *InfRatio* for year *y*, based on sorting all fund-year observations in the sample into quintiles by R^2 and within that by *alpha* or *InfRatio* based on year *y-1* estimation. *alpha* is the intercept from a regression of daily fund excess returns on the factor returns *RM-Rf*, *SMB*, *HML* and *MOM* (momentum) and their lagged values. R^2 is obtained from this regression. *InfRatio* is *alpha/RMSE* from this regression. Panel A shows the average annualized *alphas* with *t*-statistics in parentheses. Panel B presents the results for *InfRatio*. The sample period is from 1/1990 to 12/2007.

Panel A. Four-factor $alpha_y$

	R^2_{y-1}						
$alpha_{y-1}$	Low	2	3	4	High	**All**	**Low-High**
Low	-3.580 (-7.75)	-2.692 (-7.89)	-2.429 (-7.65)	-2.410 (-10.37)	-2.344 (-11.32)	-2.691 (-18.51)	-1.236 (-2.44)
2	-1.440 (-3.81)	-0.925 (-2.49)	-1.449 (-5.63)	-1.850 (-8.21)	-1.592 (-9.37)	-1.451 (-11.10)	0.152 (0.37)
3	0.258 (0.82)	0.278 (0.84)	-1.320 (-4.17)	-0.993 (-3.07)	-1.410 (-7.53)	-0.638 (-4.73)	1.668 (4.54)
4	1.046 (2.91)	-0.002 (-0.01)	-0.583 (-1.80)	-1.314 (-4.14)	-1.452 (-7.72)	-0.461 (-3.34)	2.498 (6.15)
High	2.805 (5.84)	0.582 (1.25)	0.954 (2.37)	-0.326 (-1.03)	-0.817 (-4.10)	0.638 (3.66)	3.622 (6.96)
All	-0.179 (-0.97)	-0.547 (-3.30)	-0.965 (-6.55)	-1.377 (-10.70)	-1.522 (-17.78)	-0.918 (-14.01)	1.343 (6.61)
High-Low	6.385 (9.58)	3.274 (5.68)	3.383 (6.61)	2.084 (5.30)	1.527 (5.31)	3.328 (14.67)	

Panel B. Four-factor *Information Ratio*$_y$

	R^2_{y-1}						
$InfRatio_{y-1}$	Low	2	3	4	High	**All**	**Low-High**
Low	-0.029 (-9.09)	-0.032 (-10.09)	-0.035 (-11.24)	-0.037 (-12.79)	-0.047 (-15.64)	-0.036 (-26.08)	0.018 (4.00)
2	-0.019 (-5.97)	-0.015 (-5.15)	-0.023 (-8.13)	-0.031 (-10.98)	-0.037 (-13.16)	-0.025 (-19.08)	0.019 (4.45)
3	-0.003 (-1.11)	-0.009 (-3.05)	-0.020 (-6.73)	-0.020 (-6.64)	-0.032 (-11.61)	-0.017 (-12.77)	0.028 (7.06)
4	0.004 (1.42)	-0.007 (-2.54)	-0.011 (-3.66)	-0.019 (-6.71)	-0.029 (-10.83)	-0.013 (-9.72)	0.034 (8.49)
High	0.019 (6.16)	0.004 (1.42)	0.003 (0.91)	-0.005 (-1.57)	-0.019 (-6.72)	0.0004 (0.29)	0.039 (9.09)
All	-0.006 (-4.03)	-0.012 (-8.68)	-0.017 (-12.73)	-0.022 (-16.97)	-0.033 (-25.79)	-0.018 (-29.78)	0.027 (14.45)
High-Low	0.048 (10.83)	0.037 (8.21)	0.037 (8.71)	0.032 (7.76)	0.028 (6.62)	0.037 (18.76)	

Table IX. Fund Performance: Independent sorting on R^2 and *alpha\InfRatio*

The table presents the average portfolio *alphas* or *InfRatio* for year *y*, based on independent sorting all fund-year observations in the sample into quintiles by R^2 and by *alpha* or *InfRatio* based on year *y-1* estimation. *alpha* is the intercept from a regression of daily fund excess returns on the factor returns *RM-Rf*, *SMB*, *HML* and *MOM* (momentum) and their lagged values. R^2 is obtained from this regression. *InfRatio* is *alpha/RMSE* from this regression. Panel A shows the average annualized *alphas* with *t*-statistics in parentheses. Panel B presents the results for *InfRatio*. The sample period is from 1/1990 to 12/2007.

Panel A. Four-factor $alpha_y$

	R^2_{y-1}						
$alpha_{y\text{-}1}$	Low	2	3	4	High	**All**	**Low-High**
Low	-3.526 (-8.99)	-2.814 (-9.77)	-2.417 (-7.56)	-2.372 (-9.12)	-2.346 (-8.26)	-2.787 (-18.48)	-1.180 (-2.44)
2	-0.421 (-1.08)	-0.474 (-1.18)	-1.603 (-6.35)	-1.764 (-8.40)	-1.879 (-11.40)	-1.360 (-11.36)	1.459 (3.45)
3	-0.154 (-0.36)	0.060 (0.16)	-1.436 (-4.59)	-0.833 (-2.75)	-1.477 (-9.73)	-0.888 (-6.75)	1.323 (2.94)
4	0.910 (2.33)	0.207 (0.61)	-0.438 (-1.52)	-1.424 (-4.71)	-1.229 (-7.85)	-0.495 (-3.75)	2.139 (5.08)
High	2.235 (6.21)	0.765 (1.84)	0.900 (2.15)	-0.538 (-1.53)	-0.115 (-0.36)	0.935 (5.11)	2.349 (4.88)
All	-0.179 (-0.97)	-0.547 (-3.30)	-0.965 (-6.55)	-1.377 (-10.70)	-1.522 (-17.78)	-0.918 (-14.01)	1.343 (6.61)
High-Low	5.761 (10.82)	3.579 (7.06)	3.317 (6.29)	1.834 (4.19)	2.231 (5.23)	3.722 (15.69)	

Panel B. Four-factor *Information* $Ratio_y$

	R^2_{y-1}						
$InfRatio_{y\text{-}1}$	Low	2	3	4	High	**All**	**Low-High**
Low	-0.034 (-8.57)	-0.035 (-10.54)	-0.034 (-10.60)	-0.036 (-13.02)	-0.044 (-17.20)	-0.037 (-27.09)	0.010 (2.22)
2	-0.019 (-5.98)	-0.015 (-5.11)	-0.023 (-8.15)	-0.029 (-10.51)	-0.035 (-13.57)	-0.025 (-19.36)	0.016 (3.89)
3	-0.007 (-2.41)	-0.010 (-3.40)	-0.025 (-8.31)	-0.018 (-5.95)	-0.031 (-11.18)	-0.018 (-13.72)	0.024 (6.06)
4	0.001 (0.20)	-0.006 (-2.31)	-0.011 (-3.68)	-0.020 (-6.81)	-0.030 (-10.10)	-0.013 (-9.56)	0.031 (7.42)
High	0.016 (6.03)	0.004 (1.32)	0.004 (1.24)	-0.006 (-2.03)	-0.013 (-3.74)	0.003 (2.07)	0.029 (6.66)
All	-0.006 (-4.03)	-0.012 (-8.68)	-0.017 (-12.73)	-0.022 (-16.97)	-0.033 (-25.79)	-0.018 (-29.78)	0.027 (14.45)
High-Low	0.050 (10.47)	0.039 (8.65)	0.038 (8.64)	0.030 (7.26)	0.031 (7.30)	0.040 (20.73)	

Table X. Determinants of TR^2

Panel regressions of $TR^2 = log(\sqrt{R^2}/(1- \sqrt{R^2}))$, where R^2 is obtained from the an annual regression of daily fund excess returns on the factor returns *RM-Rf*, *SMB*, *HML* and *MOM* (momentum) and their lagged values. All independent variables are as of the end of the previous year. The performance measure *alpha* is the intercept from the above regression. The Total Net Assets (*TNA*) in $mm, *Expenses* and *Turnover* are as of the end of the year. *Age* is fund age, the number of years since the fund was first offered. *Tenure* is the tenure of the manager, the number of years since the current manager took control. Each regression also includes style dummy variables and year dummy variables. The *t*-statistics in parentheses are based on standard errors clustered by fund. The sample period is from 1/1990 to 12/2007.

Variables lagged one year	(1) OLS	(2) Fund fixed effect
Expenses	-0.283 (9.82)	0.006 (0.15)
Log(TNA)	0.203 (5.26)	0.129 (2.99)
$Log(TNA)^2$	-0.011 (3.35)	-0.001 (0.37)
Turnover	0.00 (0.10)	-0.001 (4.46)
Fund Age	-0.004 (3.13)	0.002 (0.67)
Log(Manager Tenure)	-0.031 (4.45)	-0.005 (0.95)
Alpha	-0.005 (5.16)	0.002 (3.86)
Growth	--	--
Income	-0.359 (5.79)	-0.153 (1.61)
Growth and Income	0.142 (4.31)	-0.108 (1.65)
small cap	-0.287 (10.41)	-0.074 (1.22)
N of funds	2,314	2,314
Fund-years	16,646	16,646
R-sqr	0.39	0.46