Mutual Fund R^2 as Predictor of Performance

By

Yakov Amihud* and Ruslan Goyenko**

Abstract:

This paper proposes that fund performance can be predicted by its R^2, obtained from a regressing its return on the standard Fama-French-Carhart four-factor benchmark portfolios. In a cross-sectional regression, fund *alpha* and *Information Ratio* in one year has a negative and significant coefficient on the previous-year R^2. Lower R^2, which implies higher idiosyncratic risk relative to the fund total risk, is viewed as a measure of selectivity or active management, which has been shown by Cremers and Petajisto (2008) to predict performance. Selecting funds with previous year's lowest quintile R^2 and highest quintile *alpha* or *Information Ratio* produces significantly positive performance in the following year. Also, both fund RMSE and return volatility predict the following year's performance with positive and negative coefficients, respectively.

This version: December 22, 2008

* Ira Leon Rennert Professor of Finance, Stern School of Business, New York University
** School of Business, McGill University
We thank Alex Kane 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, based on the fund's R^2 from the standard 4-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, which is the proportion of the return variance that is explained by broad portfolios or indexes, is a traditional measure of diversification, and its complement, $1\text{-}R^2$ is thus a measure of proportion of idiosyncratic risk or selectivity. The closer is R^2 to 1, the closer does the fund track the benchmark portfolios and the less is the selectivity. If selectivity enhances mutual fund performance, R^2 should negatively predict the fund's performance.

This is what we find: R^2 *has a negative and significant predictive effect on fund performance in the following year*, using two conventional measures of fund performance: the intercept *alpha* from the four-factor regression model, and the *Information Ratio*, which is *alpha* scaled by the idiosyncratic (residual) risk from that regression. We also identify an R^2-based strategy that earns significantly positive average excess return (factor-adjusted) on mutual funds: at the beginning of each year, select funds whose last year's R^2 was in the lowest quintile and whose *alpha* was in the highest quintile. These funds generate a significant risk-adjusted excess return of 2.87% in the following year.

R^2 also captures another kind of active fund management: rotation between characteristics or factors over time, which may reflect timing. In our estimation of the four-factor model, the factors' coefficients are constant through the year, while active fund managers may change their portfolio such that it rotates between factors. Mamaysky, Spiegel and Zhang (2007) estimate the factor betas over 60-month periods by Kalman Filter and find that they vary over time. Our estimation period is only one year, during which factor rotation is naturally more limited, but such rotation can still be done to some extent.

By definition, R^2 is decreasing in the regression residuals standard deviation, or *RMSE*, and increasing in the standard deviation of the fund return (their squared ratio equals 1- R^2). The *RMSE* (or its square) is "tracking error," a measure 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 the intercept *alpha* from the Carhart (1997) four-factor model. Cremers and Petajisto (2008), who estimate the tracking error as the standard deviation of the fund's benchmark-adjusted returns, find that it has insignificant predictive effect on performance. However, models that estimate the performance-tracking error relationship omit the return standard deviation, which is correlated positively with the tracking error. Such model misspecification may result in a biased estimation of the effect of the tracking error on performance. We find that the fund's *RMSE* has a positive and significant predictive effect on fund performance when it is included in the prediction equation together with the standard deviation of the fund's return, which has negative and significant predictive effect on fund performance. Moreover, the negative effect of the return standard deviation on fund performance is no less important than the positive effect of the *RMSE* on performance. As pointed out, together these effects are summarized by the fund's R^2.

Studies of fund selectivity and performance take a number of forms. Daniel, Grinblatt, Titman and Wermers (1997) test whether securities that are picked by mutual funds outperform a characteristic-based benchmark, and whether fund managers' timing of choice of characteristics is successful. They find that mutual funds pick stocks that outperform simple mechanical strategies, but that the gain approximately equals the funds' average management fee. Daniel et al. (1997) carry the analysis at the fund style level, and note that aggressive growth and growth funds exhibit the best performance. Brand, Brown and Gallagher (2005) analyze selectivity at the fund level. They 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, in a way that is significant both statistically and economically, when the divergence is due to overweighting of stocks relative to the index. 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) test the effects on fund performance of deviations from diversified holdings, as reflected in the industry concentration of their holdings, measured as the difference between the industry weights of a mutual fund and the industry weights of the total market portfolio (sum of squared deviations, using 10 industry groups). They find that mutual funds with greater industry concentration exhibit better performance.

Our paper proceeds 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. Section 5 explains why the predictive power of our measures is weaker in early period and stronger in more recent periods. In Section 6 we show how using information about past fund performance and R^2 enable to choose a portfolio of funds which produces significant positive performance in the following year. Section 7 we present estimation of the association between fund characteristics and our performance predictor R^2. Concluding remarks are in Section 8.

2. Fund 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.

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 standard deviation of the residual series $e_{j,t}$ 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, p. 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 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. That is, the contribution of mutual fund A to the Sharpe ratio of the investor's portfolio is increasing in the fund's Information Ratio. Therefore, a higher fund's *InfRatio* makes the fund more attractive to investors. The Information Ratio is used as a performance measure by Brands et. al. (2005) and by Kacperczyk et al. (2005).

Another virtue of the Information Ratio is that it mitigates the survivorship bias in mutual funds. 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. The Information Ratio, which scales the abnormal fund performance by the volatility of the abnormal fund returns, mitigates this bias.

In summary, we estimate for each fund *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 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 residual standard deviation 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.[1] These data include Standard and Poor's database of live mutual funds.[2]

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

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 identified only if they have reported share holdings on CDA/Spectrum. The criteria for fund selection with Active Share estimated are the same as in Cremers and Petajisto (2008).[3]

The CRSP database identifies each shareclass separately, whereas the CDA database lists only the underlying funds. The Mutual Fund Links tables reliably assign each shareclass to 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. We therefore include 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. Whenever Weisenberg or Lipper codes are missing, we use Strategic Insight Objective Code to identify the style. Whenever Weisenberg, Lipper or 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 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

---

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

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

the sample.[4] We then classify funds into four style categories:[5] (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".

We also eliminate funds with total net asset value (TNA) less than $15 million, following the suggestion of Elton, Gruber and Blake (1996) that including funds with TNA below $15 million can cause survivorship bias in estimating mutual fund performance because of reporting convention. Addressing Evans's (2004) comment on incubation bias, we eliminate observations before the starting year reported by CRSP. As in Cremers and Petajisto (2008), we delete fund with missing name in CRSP and funds with less than 125 daily return data in each of any two consecutive years. In addition, the fund should have data on expenses, turnover, total net assets, age and tenure non-missing to be included in the sample.

For the funds that satisfy these requirements, we estimate R^2 from a regression of model (1) for the first year of the two-year pair, using current and one-day lag of benchmark returns, following Dimson (1979). We rank the resulting R^2 estimates and symmetrically trim the top and bottom 1% of the observations. The funds with R^2 near 1.0 are effectively "closet indexers" while very low R^2 represents an outlier-type strategy or estimation error. We thus obtain a final sample of 16,431 fund years (in fact, pairs of fund-years) of 2,295 funds with R^2 ranging between 0.989 and 0.240. This is the sample that we analyze. The mean R^2 is 0.86 and the median is 0.90. To address the boundaries on R^2, we use its logistic transformation

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

The resulting distribution of TR^2 is fairly symmetric, as opposed to the distribution of R^2, which is concentrated in high values of R^2. As an alternative to R^2, we use the components of R^2: *RMSE,* the root mean squared error of the regression from which we estimate R^2, and the fund's return standard deviation, *SDR*.

---

[4] We identified about 5% of fund-years with missing styles.

[5] These groups roughly follow the categorizations in Brown and Goetzmann (1997).

Control variables included in the predictive cross-fund regression are those that commonly appear in other studies on fund performance, see e.g. the recent study by Cremers and Petajisto (2008). They include Total Net Assets, *TNA*, ($mm), *Expenses*, which is the expense ratio of the most recently completed fiscal year,[6] *Turnover* or turnover ratio defined as the minimum of aggregated sales or aggregated purchases of securities divided by the average 12-month *TNA* of the fund. Among other fund characteristics we use fund age, *Age*, computed as the difference in years between current date and the date the fund was first offered, and manager tenure, *Tenure*, the difference in years between current date and the date the current manager took control.

INSERT TABLE I

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 more detailed analysis of the relationship between R^2 and the other control variables is presented in Table VIII.

4. Fund Performance prediction in cross-sectional regressions

We study the relationship between fund performance and 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* and performance is measured over the following year *y*.

4.1. Mutual Fund alpha

Table II presents the results of pooled panel regressions. As an alternative to TR^2 we use *RMSE* and *SDR*, the regression mean squared error and the excess return's standard deviation, which constitute the components of TR^2. We estimate the performance over the

[6] 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.

years 1990-2007 (the first year for parameter estimation is 1989). The control variables included in the regression are those that commonly appear in other studies on fund performance, see e.g. the recent study by Cremers and Petajisto (2008). We include the four style dummy variables that were discussed above. We also include the lagged *alpha* which may reflect managerial skill and strategy. The estimation is done in a panel pooled cross-section and time series regression with year dummy variables and style dummy variables, and with errors 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 –0.689 with t = 7.70. The negative sign means that funds with low R^2, which may be considered more active in pursuing stock selection strategies, have better performance. R^2 is a decreasing function of *RMSE* from regression (1) and an increasing function of *SDR*, the standard deviation of the fund excess return $R^e{}_t$. In column (2) we estimate the effect of the components of R^2 on *alpha*. Consistent with the results in some previous literature that uses *RMSE* as a measure of "tracking error" and thus a proxy for fund active management or selectivity, it has positive and significant effect on fund performance. The coefficient of *RMSE* is 4.055 (t = 6.37) and the coefficient of *SDR* is –6.805 (t = 19.27). This pair of results is consistent with the results on the effect of R^2.

Larger-size funds (measured by *log*(*TNA*)) perform worse, but this is mitigated for very large funds. *Expenses* have negative effect on performance, consistent with the finding of Gruber (1996). Given that R^2 is negatively correlated with *Expenses* (see Table I, Panel B), one could doubt whether low-R^2 funds would still show superior performance if it were not conditioned on them having higher expenses. We therefore re-estimate model (1) excluding the variable *Expenses*. We obtain that the coefficient of TR^2 is –0.606 with t = 6.63.

Our results on the superior performance of funds with higher R^2 are consistent with Cremers and Petajisto's (2008) results on higher performance of funds with active management, measured by *AS* (Active Share), the sum of absolute deviations of the stock weights in the fund's portfolio from these weights in its benchmark portfolio. We replicate their results in column (3): *AS* has a positive coefficient, 1.589, with t = 2.99.

The number of funds in the sample decreases to 1,875 because *AS* is estimated only for funds with available portfolio holdings data. When TR^2 and *AS* are both included in the regression, TR^2 retains its negative and highly significant effect while *AS* becomes insignificant (with negative sign). Similarly, the effects of *RMSE* and *SDR* remain practically unchanged when *AS* is included in the model.

The rest of the table reports sub-sample analysis. The year 1999 is the beginning of CRSP data. The first nine-year subperiod (1990-1998) has ¼ of the sample fund years while the second nine-year subperiod (1999-2007) that utilizes CRSP data has ¾ of the sample fund years. The results show that in the first subperiod, TR^2 is insignificant while *SDR* retains its negative and significant effect. We explain the weak performance of TR^2 during first nine years of the sample later in the Section 5. In the recent nine-year period 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 level of statistical significance. The results obtained for the whole sample hold even stronger for the last nine years of the sample.

4.2. Information Ratio

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*. Dividing *alpha* by *RMSE* also mitigates the survivorship bias (Brown, Goetzmann and Ross (1995)). We estimate whether TR^2 and the pair *RMSE* and *SDR* predict next year's *InfRatio*, 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*. Notably, the significant prediction power of TR^2 holds for the entire sample period and for each of the two subperiods (columns (6) and (8)). *RMSE* and *SDR* also predict fund performance with positive and negative coefficients, respectively, which are highly significant for the entire period. As before, the effect is stronger in the second subperiod than it is in the first.

Active Share, *AS*, is a positive and highly significant predictor of Information Ratio for the whole sample (column (3)) and it remains so after including in the model either TR^2 or *RMSE* and *SDR*. However, the coefficient of *AS* in period 1 is negative and significant while it is expected to be positive. This shows the effect of AS to be unstable over time after controlling for TR^2 and the pair *RMSE* and *SDR*. Overall, TR^2 consistently predicts the fund *Information Ratio* for the whole sample, for the two subperiods and controlling for various fund characteristics as well as for Active Share. *RMSE* has positive effect throughout, although it is not statistically significant in Period 1, while *SDR* has a negative and significant effect on subsequent *InfRatio* in both subperiods.

4.3. Fund Fixed Effects

Table IV replicates the panel regressions of Tables II and III with fund *fixed effects*, which effectively remove inter-fund differences that relate to fixed fund characteristics and account for the performance-TR^2 negative relationship. The hurdle here is raised because if a fund has a constant strategy of selectivity which results in low R^2, its performance will be shown as a result of the fixed effect and not of its R^2.

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 other fund characteristics, both fixed effects and those that vary over time. Also, *RMSE* and *SDR* 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 declines – 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 $Log(TNA)^2$.

INSERT TABLE IV

Estimating the effect of Active Share in a fixed-effect regression which excludes TR^2, its coefficient in the *alpha* regression is -2.35 with t = 2.32, which is inconsistent with the proposition that Active Share enhances performance. When adding Active Share to the *alpha* regression that includes TR^2, its coefficient is again negative and significant (-3.71, t = 3.38), while TR^2 retains its negative coefficient that is statistically significant. When adding Active Share to the *alpha* regression that includes *RMSE* and *SDR*, its effect is negative and significant. Again, 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.

4.4 Annual cross-sectional regressions

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. (In the panel regression, we allow the intercepts to vary over time by including year dummy variables.) The control variables are the same as in the panel regression, including the style dummy variables.

INSERT TABLE V

The results in Table V are consistent with the panel regression 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. One possibility for this weaker statistical significance is that in the pooled panel regression, the estimation results are largely influenced by the observations (fund-years) in recent years which are three-fold greater than those in the earlier years. We have seen there that in recent years, the negative *alpha*-TR^2 is highly significant. In contrast, in the Fama-MacBeth procedure the years are equally weighted, so the estimation in early years with fewer fund-years weighs as much as the results for recent years that have many more fund-years. Still, when we do a binomial test for the coefficient of TR^2 to be negative against the null that it is equally-likely to be positive or negative, the null is rejected at better than 5%. Interestingly, in this estimation, only the coefficient of *Expenses* is significant, in addition to the coefficient of lagged *alpha*.

When fund performance is measured by *InfRatio*, the coefficient of TR^2 is negative and significant at the 1% level. The binomial test too rejects with high statistical significance 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, the coefficients are statistically significant only in the *InfRatio* regression.

5. 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 3,967 fund years while Period 2 has 12,464 fund years, more than 3 times greater. The data source for Period 2 is CRSP, which provides broader data which may be more reliable. In addition to that, 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 $R^2_{j,y}$. This follows, for example, the convention in asset pricing empirical procedure such as that of Fama and Macbeth (1973) who use past portfolio β as an instrument for the current β. But if funds strategies change over time, $R^2_{j,y-1}$ is a poor estimator of of $R^2_{j,y}$ and then 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, $TR^2_{j,y}$ is a poor predictor of $\alpha_{j,y}$. We estimate the following regression for the entire 18-year period:[7]

$$TR^2_{j,y} = \underset{(92.68)}{0.687}\, TR^2_{j,y-1} + \text{year dummy variables}$$

The results show that the funds' R^2s are generally quite persistent from one year to the next. We then estimate a model that allows for a different slope coefficient in the two periods. Define *PERIOD2* = 1 for the years 1999-2007 (Period 2). Then,

[7] The *t*-statistics in the regressions below employ heteroskedasticity-consistent standard errors (White (1980)).

$$TR^2_{j,y} = \underset{(28.40)}{0.522}\, TR^2_{j,y-1} + \underset{(11.66)}{0.230}\, PERIOD2*TR^2_{j,y-1} + \text{year dummy variables}$$

The positive and significant coefficient of $PERIOD2*TR^2_{j,y-1}$ means that during Period 2, there was a rise of 44% in persistence in R^2_j between the years compared to the persistence in Period 1. We also estimate the model as a panel with fund fixed effects:

$$TR^2_j = \underset{(8.28)}{0.176}\, TR1^2_j + \underset{(6.19)}{0.135}\, PERIOD2*TR1^2_j + \text{year dummy variables}$$

In this estimation, the persistence in funds' R^2_j over time is higher by 77% in Period 2 than it is in Period 1. Notably, it is in Period 2 that we obtain that funds' $R^2_{j,y-1}$ strongly predict year-y performance.

But the magnitude of the slope coefficient of $TR^2_{j,y-1}$ is insufficient to tell whether we can use lagged $R^2_{j,y-1}$ as predictors of the planned $R^2_{j,y}$. The question is whether the prediction is noisy. We then do annual cross-sectional regression of $TR^2_{j,y}$ on $TR^2_{j,y-1}$ (and a constant) for y = 1990, 1991, … 2007. We obtain the following results for the average *R-sqr* from this regression:

Period 1: Average *R-sqr* = 0.24. Median *R-sqr* = 0.27.

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

These estimations means that in the second nine-year period, $R^2_{j,y-1}$ is a far more reliable (less noisy) estimate of the fund's next year's $R^2_{j,\mathrm{y}}$. This accounts at least partially for the greater significance of performance prediction by lagged R^2 in Period 2 that we observe in Tables II and III.

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

We examine mutual fund performance in a double sorting of funds by their past R^2 and past performance, and whether such sorting enables to identify and predict a group of funds with strictly positive performance. In each year y we sort funds into quintile portfolios by their R^2 in y-1 and within each quintile we sort them into five portfolios by their *alpha* (or *InfRatio*) in *y-1*. Then, for each of the resulting 25 portfolios we estimate the average *alpha* (or *InfRatio*) for year *y*.

INSERT TABLE VI

Panel A of Table VI reports the average portfolio *alpha* and Panel B reports the average portfolio *InfRatio*. Consider Panel A. Average $alpha_y$ is increasing in $alpha_{y\text{-}1}$ and decreasing in $R^2_{y\text{-}1}$, as it is in the regressions. Out of the 25 portfolios we identify 2 fund portfolios with high $alpha_{y\text{-}1}$ and low $R^2_{y\text{-}1}$ which have positive and significant $alpha_y$. In particular, the highest $alpha_{y\text{-}1}$-lowest $R^2_{y\text{-}1}$ portfolio produces annual alpha of 2.87% with t = 5.87. Notably, in the bottom-performing funds, as 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 indicates 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 is strictly increases with $InfRatio_{y\text{-}1}$. The portfolio of funds with the highest $InfRatio_{y\text{-}1}$ and lowest $R^2_{y\text{-}1}$ produces a positive $InfRatio_y$, 0.02, with t = 6.37. Here, even for the funds that perform the worst in year *y-1*, $InfRatio_y$ is monotonically decreasing in $R^2_{y\text{-}1}$.

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 VII, 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 average $alpha_y$ for the portfolio of the highest $alpha_{y\text{-}1}$ and the lowest $R^2_{y\text{-}1}$ is 2.243% (t = 6.15). The results for estimations using *InfRatio* (Panel B) are qualitatively similar.

INSERT TABLE VII HERE

7. Factors related to funds R^2

We suggest that a fund chooses a strategy, such as the extent of selectivity, whose outcomes are captured by the fund's 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 VIII HERE

The results in Table VIII show that $Expenses_{y-1}$ is negatively associated with TR^2_y. While the model is predictive, the direction of causality may run from R^2 to *Expenses*: funds that are more actively managed and expend more resources on selectivity incur higher expenses. Because *Expenses* changes very little for a given fund from year to year, the results suggest persistence in the fund policy on strategy, captured by R^2, and expenses. The positive coefficient of *Log(TNA)* is reasonable: Larger funds hold broader and more diversified portfolio, which increases their R^2, although the positive TNA-R^2 relationship is mitigated for larger funds. Older funds and fund managers with longer tenure have lower R^2 after controlling for other characteristics, in particular for fund size which is positively correlated with fund age and manager tenure. This result is interesting because it may suggest a reason for the fund longevity: greater selectivity (lower R^2) which produces better performance.

Funds with higher *alpha* subsequently have lower R^2. By one interpretation, funds have relatively stable strategy and performance. Those with more idiosyncratic investments have both higher *alpha* and lower R^2, and this produces a cross-sectional negative relationship between $alpha_{y\text{-}1}$ and R^2_y across funds. The second interpretation is behavioral: funds with exceptionally good performance in one year tend to take idiosyncratic bets in the following year because they can "afford" it, given the past success and given that funds are usually judged over a number of years. Fund styles are also associated with fund R^2: relative to growth funds, income and small-cap funds have lower R^2 while growth and income funds have higher R^2. We also observe a trend in R^2 over time. In the first nine years, the average R^2 was lower than it was in the last nine years. Apparently, the last nine-years, which use different data source and produce more significant negative relationship between performance and R^2, have different characteristics.

8. Conclusion

We propose a convenient measure of mutual fund activity or selectivity: 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*. Lower R^2 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 fund subsequent performance, with positive and negative coefficients, respectively. These results are obtained after controlling for commonly-used fund characteristics.

We also find that 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 2.87% with t = 5.87. Similar results are obtained when replacing *alpha* by the *InfRatio*.

Fund R^2 is negatively related to another measure of active fund management and idiosyncratic selectivity developed by Cremers and Petajisto (2008), called Active Share, the difference between the portfolio holdings of the fund and its benchmark portfolio. R^2 provides superior predictive power, while the effect of Active Share is insignificant in some equations and in others it has the opposite sign to the one hypothesized.

R^2 is related to identifiable fund characteristics. It is negatively related to expenses and fund age, and positively related to fund size.

Altogether, this study offers a new way to predict mutual fund performance, based on the fund R^2 which may be viewed as a measure of selectivity.

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

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

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

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

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

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.

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.

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

Treynor, J. L., and F. Black, 1973, How to use security analysis to improve portfolio selection, *Journal of Business,* 46, 66–86.

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 Weisenberg and Lipper categories that are included are aggressive growth, growth, growth and income, equity income, growth with current income, income, long-term growth, maximum capital gains, small capitalization growth. Index and sector funds are excluded from the sample. The performance measure *alpha* is the intercept from an annual regression of daily fund excess returns on the factors mkt, smb, hml and 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 January 1989 to December 2007.

Panel A: Fund Characteristics

	Mean	Median	Minimum	Maximum
Total number of funds 2,295				
TNA (total net assets)(in millions)	1359.32	243.30	15.1	161,911.9
Age (years)	12.85	8.12	1.5	83.92
Expenses (%)	1.27	1.23	0.01	4.54
Turnover (%)	89.38	66.00	0.20	3,603
Tenure (years)	3.08	2.08	0.08	44.08
Alpha (%)	-0.69	-1.12	-77.40	90.63
R^2	0.86	0.90	0.240	0.989
TR^2	2.88	2.95	-0.026	5.158

Panel B: Correlation Structure

	Log(TNA)	*Age*	*Expenses*	*Turnover*	*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				
Tenure	0.20**	0.22**	-0.12**	-0.08**	1.00			
Alpha	0.04**	-0.03**	-0.04**	-0.03**	-0.04**	1.00		
R^2	0.13**	-0.02	-0.10**	-0.02*	0.07**	-0.08**	1.00	
TR^2	0.15**	-0.002	-0.13**	-0.04**	0.08**	-0.09**	0.92**	1.00

**1% significance, *5% significance

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 factors mkt, smb, hml and 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 returns over the 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 dummies, and *t*-statistics in parentheses are based on standard errors clustered by fund. The sample period is from January 1990 to December 2007.

Indep. Vars.	**1990-2007**					**1990-1998**		**1999-2007**	
(lagged one year)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
TR^2	-0.689			-0.747		-0.031		-0.959	
	(7.70)			(5.92)		(0.18)		(9.47)	
RMSE		4.055			4.472		0.463		4.544
		(6.37)			(5.96)		(0.37)		(5.58)
SDR		-6.805			-7.121		-2.261		-7.333
		(19.27)			(16.43)		(1.97)		(20.20)
Expenses	-0.965	-0.549	-0.671	-0.743	-0.373	-0.732	-0.674	-0.109	-0.520
	(5.68)	(3.25)	(3.16)	(3.51)	(1.80)	(1.80)	(1.67)	(6.17)	(2.98)
Log(TNA)	-0.688	-0.472	-0.841	-0.747	-0.561	-0.802	-0.749	-0.613	-0.396
	(3.25)	(2.22)	(3.61)	(3.22)	(2.40)	(1.45)	(1.36)	(2.80)	(1.78)
$Log(TNA)^2$	0.049	0.034	0.058	0.052	0.039	0.064	0.060	0.040	0.026
	(2.97)	(2.06)	(3.22)	(2.91)	(2.16)	(1.47)	(1.39)	(2.35)	(1.47)
Turnover	-0.003	-0.00	-0.001	-0.002	0.002	0.001	0.002	-0.003	-0.001
	(1.65)	(0.00)	(0.58)	(0.67)	(0.62)	(0.23)	(0.43)	(2.89)	(0.59)
Fund Age	-0.002	0.002	0.003	0.002	0.006	-0.018	-0.015	0.006	0.009
	(0.45)	(0.39)	(0.70)	(0.40)	(1.11)	(1.91)	(1.52)	(1.17)	(1.58)
Manager	-0.006	-0.002	-0.017	-0.019	-0.011	-0.060	-0.061	-0.004	0.002
Tenure	(0.43)	(0.17)	(1.08)	(1.21)	(0.72)	(0.73)	(0.76)	(0.29)	(0.20)
Alpha	0.172	0.172	0.183	0.183	0.186	0.179	0.169	0.167	0.173
	(11.66)	(11.60)	(13.51)	(13.43)	(13.65)	(6.27)	(6.08)	(9.60)	(9.64)
AS			1.589	-0.717	-1.067				
			(2.99)	(1.18)	(1.95)				
N of funds	2,295	2,295	1,875	1,875	1,875	863	863	2,166	2,166
Fund-years	16,431	16,431	13,043	13,043	13,043	3,967	3,967	12,464	12,464
R^2	0.18	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 factors mkt, smb, hml and momentum, and their lagged values, and *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 over the 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 dummies, and *t*-statistics in parentheses are based on standard errors clustered by fund. The sample period is from January 1990 to December 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		-0.004		-0.007	
	(13.00)			(5.46)		(2.15)		(5.44)	
RMSE		0.052			0.031		0.013		0.034
		(12.38)			(6.41)		(1.30)		(5.50)
SDR		-0.054			-0.046		-0.021		-0.048
		(21.67)			(16.09)		(2.39)		(16.16)
Expenses	-0.012	-0.009	-0.012	-0.013	-0.010	-0.008	-0.007	-0.014	-0.011
	(8.54)	(6.31)	(6.86)	(7.24)	(5.94)	(2.30)	(2.05)	(7.16)	(5.80)
Log (TNA)	-0.006	-0.005	-0.008	-0.007	-0.006	-0.006	-0.006	-0.007	-0.006
	(3.34)	(2.77)	(3.64)	(3.28)	(2.79)	(1.33)	(1.24)	(2.92)	(2.49)
$Log(TNA)^2$	0.000	0.000	0.001	0.000	0.0004	0.000	0.000	0.001	0.000
	(3.05)	(2.55)	(3.15)	(2.89)	(2.47)	(0.94)	(0.84)	(2.65)	(2.28)
Turnover	-0.00	-0.00	-0.00	-0.00	-0.000	0.00	0.00	-0.00	-0.00
	(3.17)	(1.45)	(1.28)	(1.41)	(0.03)	(0.14)	(0.45)	(3.00)	(0.76)
Fund Age	-0.000	0.000	0.000	0.000	0.000	-0.000	-0.000	0.000	0.000
	(0.28)	(0.30)	(0.41)	(0.19)	(0.61)	(0.81)	(0.56)	(0.59)	(0.83)
Manager	-0.000	0.000	-0.000	-0.000	-0.0001	-0.000	-0.000	-0.000	-0.000
Tenure	(0.11)	(0.17)	(0.95)	(1.04)	(0.81)	(0.45)	(0.34)	(1.05)	(0.78)
InfRatio	0.163	0.159	0.161	0.157	0.153	0.182	0.180	0.143	0.140
	(20.38)	(20.27)	(17.52)	(17.15)	(16.84)	(10.29)	(10.15)	(12.94)	(12.76)
AS			0.051	0.033	0.033	-0.033	-0.029	0.043	0.047
			(9.85)	(5.44)	(6.16)	(2.57)	(2.32)	(5.94)	(7.63)
N of funds	2,295	2,295	1,875	1,875	1,875	723	723	1,801	1,801
Fund-years	16,431	16,431	13,043	13,043	13,043	3,260	3,260	9,783	9,783
R-sqr	0.18	0.19	0.18	0.18	0.20	0.12	0.12	0.21	0.23

Table IV. Predictive Regressions of Fund Performance, Fund Fixed Effects

Panel regressions with fund fixed-effects. The dependent variables are *alpha*, the intercept from an annual regression of daily fund excess returns on the factors mkt, smb, hml and 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 January 1990 to December 2007.

Variables lagged one year	Dependent variables			
	alpha		*InfRatio*	
	(1)	(2)	(3)	(4)
TR^2	-0.434		-0.003	
	(3.27)		(3.00)	
RMSE		3.988		0.028
		(5.06)		(5.35)
SDR		-7.709		-0.041
		(12.75)		(12.56)
Expenses	0.247	-0.131	-0.001	-0.003
	(0.49)	(0.26)	(0.21)	(0.64)
Log(TNA)	-4.039	-3.486	-0.029	-0.026
	(9.16)	(8.41)	(7.75)	(7.13)
$Log(TNA)^2$	0.131	0.116	0.000	0.000
	(3.86)	(3.60)	(1.64)	(1.43)
Turnover	0.003	0.004	0.00	0.00
	(1.56)	(1.92)	(0.94)	(1.15)
Fund Age	-0.004	0.005	0.0001	0.000
	(0.17)	(0.21)	(0.71)	(0.96)
Manager Tenure	-0.023	-0.022	-0.0003	-0.000
	(1.04)	(1.00)	(1.21)	(1.22)
Alpha	0.053	0.068		
	(3.10)	(3.84)		
InfRatio			0.003	0.005
			(0.39)	(0.58)
N of funds	2,295	2,295	2,295	2,295
Fund-years	16,431	16,431	16,431	16,431
R-sqr	0.21	0.24	0.19	0.20

Table V. Fama-MacBeth Regressions

The dependent variables are *alpha*, the intercept from an annual regression of daily fund excess returns on the factors mkt, smb, hml and 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 and *t*-statistics (in parentheses) of the coefficients from annual cross-sectional regressions. The sample period is from January 1990 to December 2007.

Variables lagged	Dependent variable			
one year	*alpha*	*alpha*	*InfRatio*	*InfRatio*
TR^2	-0.468		-0.008	
	[0.065]		[0.007]	
RMSE		1.071		0.041
		[0.332]		[0.010]
SDR		-2.647		-0.035
		[0.104]		[0.030]
Expenses	-0.894	-0.685	-0.012	-0.011
	[0.020]	[0.045]	[0.001]	[0.002]
Log(TNA)	-0.291	-0.230	-0.004	-0.004
	[0.317]	[0.416]	[0.188]	[0.231]
$Log(TNA)^2$	0.023	0.018	0.0003	0.0003
	[0.322]	[0.441]	[0.271]	[0.328]
Turnover	0.001	0.002	-0.00	-0.00
	[0.735]	[0.511]	[0.565]	[0.817]
Fund Age	-0.010	-0.007	-0.0001	-0.0001
	[0.180]	[0.272]	[0.177]	[0.282]
Manager Tenure	0.529	0.289	0.030	0.041
	[0.557]	[0.666]	[0.375]	[0.365]
Dependent variable,	0.154	0.143	0.167	0.161
Lagged	[0.001]	[0.000]	[0.000]	[0.000]
R-sqr	0.216	0.248	0.246	0.262
TR^2: pos/neg	5/13		3/15	
	[0.048]		[0.004]	
RMSE: pos/neg		9/9		12/6
		[0.593]		[0.119]
SDR: pos/neg		5/13		4/14
		[0.048]		[0.015]

Table VI. 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 factors mkt, smb, hml and 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 January 1990 to December 2007.

Panel A. Four-factor $alpha_y$

	R^2_{y-1}					
$alpha_{y\text{-}1}$	Low	2	3	4	High	**LOW-HIGH**
Low	-3.714 (-7.83)	-2.645 (-7.04)	-2.506 (-7.74)	-2.445 (-10.24)	-2.406 (-11.38)	-1.308 (-2.52)
2	-1.231 (-3.22)	-0.986 (-2.74)	-1.446 (-5.44)	-1.920 (-8.44)	-1.574 (-9.23)	0.343 (0.82)
3	0.371 (1.15)	-0.017 (-0.05)	-1.360 (-4.24)	-1.033 (-3.14)	-1.473 (-7.75)	1.843 (4.91)
4	0.945 (2.61)	0.118 (0.37)	-0.517 (-1.56)	-1.287 (-3.95)	-1.465 (-7.76)	2.410 (5.91)
High	2.867 (5.87)	0.652 (1.37)	0.936 (2.30)	-0.361 (-1.11)	-0.861 (-4.18)	3.729 (7.04)
High-Low	6.581 (9.67)	3.297 (5.45)	3.443 (6.63)	2.085 (5.16)	1.544 (5.23)	

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

	R^2_{y-1}					
$InfRatio_{y\text{-}1}$	Low	2	3	4	High	**LOW-HIGH**
Low	-0.029 (-9.08)	-0.032 (-10.07)	-0.035 (-11.58)	-0.036 (-12.73)	-0.047 (-15.48)	0.018 (4.06)
2	-0.015 (-5.04)	-0.015 (-5.02)	-0.023 (-7.81)	-0.031 (-11.45)	-0.037 (-13.05)	0.022 (5.24)
3	-0.003 (-0.93)	-0.009 (-3.18)	-0.020 (-6.86)	-0.019 (-6.75)	-0.032 (-11.90)	0.029 (7.27)
4	0.004 (1.39)	-0.006 (-2.30)	-0.011 (-3.67)	-0.019 (-6.57)	-0.028 (-10.21)	0.032 (8.05)
High	0.020 (6.37)	0.006 (1.84)	0.004 (1.17)	-0.005 (-1.76)	-0.020 (-6.77)	0.040 (9.28)
High-Low	0.049 (10.96)	0.038 (8.48)	0.039 (9.08)	0.031 (7.53)	0.027 (6.34)	

Table VII. 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 factors mkt, smb, hml and 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 January 1990 to December 2007.

Panel A. Four factor $alpha_y$

	R^2_{y-1}					
$alpha_{y-1}$	Low	2	3	4	High	**All**
Low	-3.558 (-8.85)	-2.791 (-8.88)	-2.553 (-7.89)	-2.396 (-8.91)	-2.402 (-8.34)	-2.827 (-18.07)
2	-0.306 (-0.77)	-0.612 (-1.59)	-1.582 (-6.07)	-1.814 (-8.50)	-1.881 (-11.32)	-1.378 (-11.53)
3	-0.044 (-0.10)	-0.001 (-0.00)	-1.371 (-4.29)	-0.909 (-2.96)	-1.505 (-9.77)	-0.896 (-6.73)
4	0.897 (2.27)	0.127 (0.38)	-0.457 (-1.53)	-1.344 (-4.33)	-1.248 (-7.81)	-0.503 (-3.75)
High	2.243 (6.15)	0.833 (1.96)	0.914 (2.16)	-0.643 (-1.79)	-0.210 (-0.63)	0.936 (5.02)
All	-0.149 (-0.79)	-0.572 (-3.38)	-0.980 (-6.53)	-1.408 (-10.70)	-1.555 (-17.87)	

Panel B. *Information Ratio*$_y$

	R^2_{y-1}					
$InfRatio_{y-1}$	R1 – 1st	R1-2nd	R1-3rd	R1-4th	R1-5th	**All**
Low	-0.032 (-8.16)	-0.034 (-10.33)	-0.035 (-10.98)	-0.036 (-13.12)	-0.044 (-17.17)	-0.037 (-27.02)
2	-0.017 (-5.54)	-0.015 (-5.15)	-0.023 (-8.07)	-0.030 (-10.82)	-0.035 (-13.38)	-0.024 (-19.20)
3	-0.007 (-2.29)	-0.010 (-3.58)	-0.025 (-8.29)	-0.018 (-6.10)	-0.030 (-10.83)	-0.018 (-13.69)
4	0.001 (0.35)	-0.006 (-2.15)	-0.010 (-3.52)	-0.019 (-6.69)	-0.031 (-10.18)	-0.012 (-9.29)
High	0.017 (6.32)	0.005 (1.75)	0.004 (1.41)	-0.007 (-2.15)	-0.013 (-3.73)	0.003 (2.42)
All	-0.005 (-3.35)	-0.011 (-8.38)	-0.017 (-12.69)	-0.022 (-17.18)	-0.033 (-25.52)	

Table VIII. 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 factors mkt, smb, hml and 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 January 1990 to December 2007.

Variables lagged one year	Model 1		Model 1
Expenses	-0.307	1994	-0.137
	(9.90)		(1.82)
Log(TNA)	0.189	1995	-0.502
	(4.74)		(6.64)
Log(TNA)^2	-0.010	1996	-0.101
	(2.90))		(1.25)
Turnover	0.0001	1997	0.460
	(0.67)		(5.75)
Fund Age	-0.004	1998	-0.288
	(3.36)		(3.55)
Manager	-0.006	1999	-0.007
Tenure	(1.76)		(0.09)
Alpha	-0.003	2000	0.443
	(3.20)		(5.98)
Growth	--	2001	0.859
			(11.53)
Income	-0.354	2002	1.342
	(5.29)		(18.12)
Growth and	0.171	2003	1.256
Income	(5.02)		(17.07)
small cap	-0.240	2004	0.989
	(8.48)		(13.51)
1990	--	2005	1.008
			(13.67)
1991	-0.180	2006	1.046
	(2.41)		(14.07)
1992	-0.270	2007	1.423
	(3.48)		(19.05)
1993	-0.439		
	(5.56)		
N of funds	2,295		
Fund-years	16,431		
R-sqr	0.41		