Mutual Fund's R^2 as Predictor of Performance

By

Yakov Amihud[*] and Ruslan Goyenko[**]

Abstract:

We propose that fund performance can be predicted by its R^2, obtained by regressing its return on the multi-factor benchmark model. Lower R^2 measures selectivity or active management. We find that lagged R^2 has significant negative predictive coefficient in predicting *alpha* or *Information Ratio*. Funds ranked into lowest-quintile lagged R^2 and highest-quintile *alpha* produce significant *alpha* of 2.5%. Across funds, R^2 is positively related to the fund's size and negatively related to its manager's tenure and its past performance, as well as being negatively related to its expenses. We also find that for funds that hold corporate bonds, R^2 from a benchmark model that includes bond factors predicts fund performance in the same way that it does for stock funds.

This version: July 30, 2010

* 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 as well as to the seminar participants at Erasmus University, McGill University, Tilburg University and EFA 2009 annual meeting. We thank Martijn Cremers and Antti Petajisto for providing us with the data of funds' Active Share and of the benchmark indexes which are used in this study.

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-defined, 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, the proportion of the return variance that is explained by benchmark portfolios, estimated from a multi-factor regression model of its return. R^2 measures diversification and 1-R^2 measures the weight (relative to the fund's variance) of idiosyncratic risk, thus measuring the fund's selectivity. If R^2 is smaller than 1, the fund tracks less closely the benchmark portfolios and thus shows greater selectivity. If selectivity enhances mutual fund performance, R^2 should negatively predict the fund's performance.

Indeed we find that R^2 *has a negative and significant predictive effect on fund performance*. We use two conventional measures of performance: the intercept *alpha* from a multi-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 half-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 2.457% or 2.693% per year depending on benchmark model specification.

Our results are robust to the factor model that is used as benchmark. We use the standard Fama-French (1993) and Carhart (1997) four-factor model, augmented by the returns on the Russell 2000 index, and the Cremers-Petajisto-Zitzewitz (2010) four-factor model that uses market indexes. All our estimates of R^2 and *alpha* use both these benchmark models and our results are similar for both. This flexibility and versatility in the benchmark model is a valuable property of R^2 as a measure of selectivity.[1]

The versatility of our methodology is demonstrated for mutual funds that hold *corporate bonds* as well as stocks. Here, we use benchmark factor models that include both stock factors and bond factors. We show that the R^2 from these models predicts fund performance in the same way that it does for stock funds.

Studies on fund selectivity use fund holdings data. Brand, Brown and Gallagher (2005) and Cremers and Petajisto (2009) show that a fund active management – the divergence of its portfolio composition (the portfolio weights of the stocks that it holds) from the composition of the fund's benchmark index – enhances fund performance. Earlier, Daniel, Grinblatt, Titman and Wermers (1997) show that stocks picked by mutual funds outperform a characteristic-based benchmark. However, the gain from stock picking approximately equals the funds' average management fee. 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 different from analysts' expectations exhibit better performance.

Our analysis does not require fund holdings data or benchmark index composition. We use only fund and benchmark index returns, which are easily accessible, and our measure of fund's strategy – its R^2 – can be calculated easily.

Recent studies of hedge fund performance use R^2 as a measure of fund strategy and find, as we do, that lower R^2 predicts better fund performance; see Wang and Zheng (2008) and Titman and Tiu (2008). The latter paper suggests that choosing smaller exposure to factor risk reflects hedge funds managers' confidence in their ability. However, studies that use hedge fund returns suffer from reporting biases resulting from self reporting and incomplete data on holdings (see Agarwal, Fos and Jiang (2010)).

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

[1] We also do our estimates for the six Fama-French portfolios (2x3) classified by size (small and big) and value, neutral or growth, plus Carhart's momentum factor and the seven-factor model proposed by Cremers et al. (2010). Our results remain qualitatively unchanged.

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. Section 7 presents results for funds that hold corporate bonds. Concluding remarks are in Section 7.

2. Performance measures and performance predictors

2.1. Performance measures

Our study employs two models of benchmark portfolios. The first model consists of the Fama-French (1993) and Carhart (1997) factor-mimicking portfolios which produce the following return vectors: *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). To this we add, following Cremers et all. (2009), the daily excess return on the Russell 2000 index which has a significant *alpha* (-2.059% per year with t = 3.41) when regressing it on the Fama-French-Carhart four factors for our sample period, 1989 to 2007 (using current and lagged returns).[2] Given the high correlation of the Russell 2000 index returns with the Fama-French-Carhart factors, we use the residuals plus intercept from a regression of the Russell 2000 excess return on the Fama-French-Carhart factors returns (current and one-day lag).[3] We thus have a five-factor model, denoted as FFCR (Fama-French-Carhart-Russell 2000).

[2] Cremers et al. (2010) also suggest adding the S&P500 index return. However, for our sample period, the *alpha* of this index daily returns on the four Fama-French-Carhart factors (including one-day lagged return) is insignificant (t-statistic is 1.25).

[3] We add lagged one-day return, following Dimson (1979). See comment in footnote 9.

The second model of benchmark portfolios is the Cremers-Petajisto-Zitzewitz's (2010) three-index model: the excess return on the S&P500 index, the return on the Russell 2000 index minus the return on the S&P500 index and the return on the Russell 3000 value index minus Russell 3000 growth index. To this they add Carhart's momentum factor. We denote this four-factor model by CPZC. This model is also used for mutual fund performance evaluation by Da et al. (2009).

We employ two standard measures of fund performance. The first is the intercept $alpha_j$ from a regression of the excess daily fund-*j* return on the daily factor returns, using either the FFCR model or the CPZC benchmark model.

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

$$InfRatio_j = \frac{alpha_j}{RMSE_j} \ . \qquad (1)$$

$RMSE_j$ is the squared root of the mean squared errors (residuals) obtained from the regression model that we use to estimate $alpha_j$. 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 *j*th asset is proportional to *InfRatio/RMSE* (see Trynor and Black (1973), p. 71). If we 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. Treynor and Black (1973) further 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, and this is the objective that the active fund manager should try to maximize. 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.[4] The Information Ratio, which scales the abnormal fund performance by the volatility of the abnormal fund returns, mitigates this bias.

From the regression of the fund return on the benchmark indexes we also obtain R^2_j, which we propose as a predictor of the fund's performance. By definition

$$R^2 = 1 - \frac{RMSE^2}{VARIANCE} = \frac{SystematicRisk^2}{SystematicRisk^2 + RMSE^2},$$

where $SystematicRisk^2$ is the variance of return which is due to the benchmark indexes. We propose that $1\text{-}R^2$ is a measure of selectivity: a fund with greater *RMSE relative* to its total variance, or with a smaller index-based (systematic) risk, has greater selectivity. Studies on the effect of *RMSE*-related measures on fund performance show inconclusive results. Wermers (2003) find that performance is better in funds with higher volatility of the S&P500-adjusted fund returns, which he uses as a measure of active management or selectivity, whereas Cremers and Petajisto (2009) find that the fund's standard deviation of return relative to its specific benchmark index ("tracking error") does not predict performance. Our measure of fund selectivity, $1\text{-}R^2$, is a multiple factor-based *RMSE* relative to the adherence of the fund's return to these factors.

In what follows, we estimate for each fund the two performance measures, *alpha* and *InfRatio*, and the fund's R^2. We then test whether R^2 predicts fund performance, controlling for other fund characteristics. The main results are presented for both the FFCR factors and the CPZC factor models.

---

[4] 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. Using the *InfRatio*, performance (*alpha*) is muted for funds with higher volatility, thus mitigating the survivorship bias.

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 daily returns for mutual funds are from the CRSP Mutual Fund Database from January 1999 to 2007. These are net returns after fees, expenses, and brokerage commissions but before any front-end or back-end loads. The daily fund returns from 1989 to 1998 are obtained from the International Center for Finance at Yale School of Management.[5] These data include Standard and Poor's database of live mutual funds.[6] 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 (it is also used by Cremers and Petajisto (2009)). The final sample spans the period from January 1989 to December 2007. We also use data on funds' Active Share, following Cremers and Petajisto (2009); see the analysis below.

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-estimation period values of these variables.

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 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 only 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

---

[5] We thank William Goetzmann for providing these data.

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

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 half-year and a fund has a past half-year with the style identified, that fund half-year is assigned the style of the previously identified style-half-year. If the fund style cannot be identified, it is not included in the sample.[7] We use nine style categories: (i) Aggressive Growth, (ii) Equity Income, (iii) Growth, (iv) Long term growth, (v) Growth and Income, (vi) Mid-Cap, (vii) Micro-Cap funds, (viii) Small cap, and (ix) Maximum Capital Gains. We eliminate index funds by deleting those whose name includes the word "index" or the abbreviation ind, S&P, DOW, Wilshire and Russell. We eliminate balanced funds, international funds,[8] sector funds and funds that hold less than 80% in common stocks. Following Elton, Gruber and Blake (1996), we eliminate funds with total net assets of less than $15 million at the end of the period preceding the test period 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 (2009), we delete funds with missing name in CRSP. We also require funds to have data on expenses, turnover, total net assets, age and managerial tenure in the year or half-year before the test period.

We divide the data into overlapping sequences of pairs of periods. These periods are either a year or half year. In the first period of each pair, the estimation period, we estimate for each fund its R^2, *alpha* and *InfRatio*. In the following period, the test period, we estimate the fund's *alpha* and *InfRatio*, and then test whether they can be predicted by lagged R^2, *alpha* and *InfRatio* obtained from the estimation period, and by fund characteristics that are observed before the test period. Notably, the test periods are non-overlapping. We require funds to have at least 120 daily return data in the estimation

[7] If Wiesenberger Code and Lipper Code are missing, we use a style identifier and check if the fund name corresponds to the style. If it does not, we consider the style as unidentified. About 5% of fund-years have their style missing.
[8] We also eliminated international funds by name if they are classified as domestic active equity funds.

period and only 50 daily return data in the subsequent test period.[9] We do the analysis for either 18 annual test periods, 1990-2007, or for 36 half-year test periods.

We estimate R^2, *alpha* and *InfRatio* from a regression of the fund's daily excess returns on the daily returns on the FFCR factors and the CPZC factors. Throughout our analysis, in factor model regressions with daily returns we use the contemporaneous and one-day lagged returns, following Dimson (1979),[10] and in weekly-return models we use the contemporaneous returns only. For funds that satisfy these requirements, we censor 1% at each tail of the distribution of the estimated R^2 because 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.[11] We thus obtain a final sample of 11,230 observations for annual and 22,818 for semi-annual fund-period pairs for 2,026 funds. For the annual estimations and the FFCR five-factor model, R^2 ranges between 0.290 and 0.988 with mean of 0.86 and the median of 0.90 (see Table 1, Panel A). The values using the CPZC four-factor benchmark model are nearly the same.

INSERT TABLE 1

The distribution of R^2 is negatively skewed, with its mass being in the high values of R^2, close to 1.0 which is its upper limit. We therefore apply to R^2 a logistic transformation

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

The resulting distribution of TR^2 is more symmetric than that of R^2. (We also show results for the untransformed R^2.)

The control variables in the predictive cross-fund regressions are those that commonly appear in studies of fund performance. For example, Cremers and Petajisto (2009) use Total Net Assets, *TNA*, ($mm); *Expense*, the expense ratio of the most recently completed fiscal year;[12] *Turnover*, defined as the minimum of aggregated sales or aggregated purchases of securities divided by the average 12-month *TNA* of the fund;

[9] Our requirement for a relatively short period in the test period reduces the extent of survivorship problem. Cremers and Petajisto (2008) require 125 days in the test period (the second year of a two-year pair).

[10] However, Cremers et al. (2010, p. 36) find that the estimation of daily fund performance is not harmed by the use of daily fund returns: "Stale prices would undoubtedly be more important for individual stocks, but mutual funds hold broad portfolios of stocks, so the average staleness in fund return is likely to be close to the average staleness in benchmark index return."

[11] Our results are qualitatively similar if we winsorize the data instead.

[12] Expense ratio is the fraction 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.

Fund Age in logarithm, 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.[13] An important predictor of future performance is lagged *alpha* or *InfRatio* which may reflect managerial skill and strategy and is known to be a significant predictor of performance (see Brown and Goetzmann (1995) and Gruber (1996). Statistics of these variables are presented in Table 1.

The correlation table, Panel B of Table 1, shows that R^2 is larger for large funds (with high *TNA*), which cannot be niche investors and must hold a broad portfolio. This makes their performance closer to that of broad indexes. Funds with lower R^2 have more idiosyncratic investment and higher expense ratio, as evident from the negative correlation between R^2 and *Expenses*. A detailed analysis of the relation between R^2 and the control variables is presented in Table 9.

4. Fund Performance prediction in cross-sectional regressions

We now test whether R^2 predicts fund performance by regressing the fund's $Performance_{j,t}$ – either $alpha_{j,t}$ or $InfRatio_{j,t}$ – estimated in the test period t, on its $TR^2_{j,t-1}$ (logistic transformation of R^2) from the preceding estimation period. The control variables include fund characteristics that are known at the beginning of the test period, the fund's lagged performance and nine style dummy variables. The cross-section estimation employs the Fama-Macbeth method, following Carhart (1997) and Chen, Hong, Huang and Kubik (2004). The estimated model is

$$\begin{aligned} Performance_{j,t} = {} & \gamma_t TR^2_{j,t-1} + \delta_{1t} Expenses_{j,t-1} + \delta_{2t} log(TNA)_{j,t-1} + \delta_{3t}[log(TNA)]^2_{j,t-1} \\ & + \delta_{4t} Turnover_{j,t-1} + \delta_{5t} log(Fund\ Age)_{j,t-1} + \delta_{6t} log(Manager\ tenure)_{j,t-1} \\ & + \delta_{7t} Performance_{j,t-1} + \sum_{n=1}^{9} \lambda_n StyleDummy_{j,n,t-1} \end{aligned} \tag{3}$$

Performance is either *alpha* or *InfRatio*. Our hypothesis is that $\gamma < 0$. That is, fund performance is higher if the fund's R^2 is lower, which means that the fund shows greater

[13] The manager can be an institution with a long tenure.

selectivity in its investment. We estimate the coefficients $\gamma_{,t}$, δ_{mt} and λ_{nt} (m=1, 2, … 7, n = 1, 2, … 9) using both annual frequency (18 years) and semi-annual frequency (36 half years) over the period 1990-2007.

INSERT TABLE 2

The results In Table 2 Show that R^2 is a significant predictor of *alpha* and *InfRatio*. Consistent with out hypothesis, we obtain $\gamma < 0$ for both annual and semi-annual frequency and for both sets of benchmark factors, FFCR and CPZC. Consider first the *alpha* model for the benchmark model FFCR. The mean γ is -0.586 (t = 2.38) in the annual frequency (column (1)) and -0.596 (t = 3.55) in the semi-annual frequency (column (5)). The proportion of negative coefficients is 14/18 for the annual regressions and 27/36 for the semi-annual regressions, significantly rejecting the null hypothesis that this result is obtained by chance (a proportion of 1/2).

For the CPZC benchmark model, TR^2 has stronger predictive power for *alpha* both economically and statistically. The mean γ is -1.003 (t = 3.49) for the annual frequency and -1.042 (t = 5.56) for the semi-annual frequency, and the proportion of negative coefficients is 17/18 for the annual model and 30/36 for the semi-annual model, significantly different from ½ which is the chance result. (In all cases, our tests show that the estimated coefficients γ_t is not serially correlated.)

In the *InfRatio* equations, the coefficient γ of TR^2 is more statistically significant than it is in the *alpha* equations. It has higher t-statistic and has greater proportion of negative coefficients in both annual and semi-annual estimation frequencies and for both the FFCR and the CPZC models. In conclusion, the evidence shows that a fund's R^2 is a significant predictor of the fund performance.

Our analysis employs TR^2, a logistic transformation of R^2 whose distribution is bounded between 0 and 1 and is negatively skewed. Using instead the untransformed R^2, its effect on fund performance remains negative and significant. For the semi-annual FFCR model, the mean coefficient of R^2 is -4.98 with t = 3.23, highly significant. The proportion of negative coefficients, 26/36, is significantly different from ½ at better than the 0.01 level. For the CPZC benchmark model, the coefficient of R^2 is -8.75 with t = 4.65 and a proportion of 27/36 of negative coefficients, significantly different from ½ at the 0.01 level.

The economic meaning of our estimations is illustrated as follows. The estimated coefficient of R^2 is -8.75 (estimated in the semi-annual CPZC-benchmark model). This means, for example, that lowering R^2 from 0.9 (which is the mean) to 0.8 raises the fund's annualized *alpha* by 0.875%. For comparison, using the results from the same model that employs TR^2 where the slope coefficient is -1.042, a decline in R^2 from 0.9 to 0.8 would raise the annualized alpha by 0.814%, which is quite close.

Among the control variables, two are statistically significant: *Expenses*, with a negative and significant coefficient, and lagged *alpha* or lagged *InfRatio*, with a positive and significant coefficient. Both these effects are observed by Gruber (1996). Fund size, which is sometimes observed to have negative effect on performance (see Chen et al., 2004) is insignificant.

4.1. Robustness test I: weekly returns

So far we employed daily returns to calculate R^2 and *alpha*, regressing the funds' daily returns on the current and one-day lag of the benchmark factors, following Dimson (1979). If this procedure does not fully account for lagged adjustment of daily stock prices, we replicate our analysis using weekly returns. The estimation is annual and we require that a fund has at least 40 weeks in the estimation period and 20 weeks in the test period. The estimated alpha is annualized and in percent (multiplied by 5200).

INSERT TABLE 3 HERE

The results in Table 3 are consistent with those obtained from daily data, both in terms of the magnitude of the coefficients and their statistical significance. In the regression of the annual *alpha* on lagged TR^2, the mean γ_1 for the FFCR model is -0.586 (t = 2.82), the same as the coefficient obtained from daily data (Table 2). For the CPZC model, the mean γ_1 is -0.981 (t = 3.18), very close to the coefficients -1.003 from the daily return model. Also, the proportion of negative coefficients is significantly higher than the chance result at the 0.05 level.

4.2. Robustness test II: Alternative benchmark factors and alternative estimation method

We replicate our analysis using an alternative benchmark index model, proposed by Cremers et al.'s (2010), which consists of seven indexes: S&P500, Russell Midcap minus S&P 500, Russell 2000 minus Russell Midcap, S&P500 Value minus S&P 500 Growth, Russell Midcap Value minus Russell Midcap Growth, Russell 2000 Value minus Russell 2000 Growth, and Carhart's UMD. Return data for these indexes are obtained from Antti Petajisto's web site, and they are available starting in 1993, i.e., we lose 6 semi-annual test periods. We estimate R^2, *alpha* and *InfRatio* from this model, using daily returns (including a one-day lagged return), and follow the very same procedure as we have done so far.

We obtain that the coefficient of TR^2 in the *alpha* and *InfRatio* equations are, respectively, -0.553 (t = 3.29) and -0.011 (t = 5.49). The proportion of negative coefficients is 23/30 for *alpha* and 25/30 for *InfRatio*, both are significantly different from ½ (chance result).

We also do the analysis for a seven-factor model that includes the six Fama-French portfolios (2x3) classified by size (small and big) and by value, neutral or growth, plus Carhart's momentum factor. The results are similar: TR^2 is a significant predictor, with a negative sign, of *alpha* and *InfRatio*.

The estimation so far has been done by the Fama-Macbeth method. We also estimate our model by the panel regression method, adding to the model time period dummy variables. The standard errors are clustered by both funds and time periods. We obtain that for the *alpha* equation, the coefficients of TR^2 are –0.487 (t = 2.38) for the FFCR model and –1.040 (t = 3.93) for the CPZC model. For the *InfRatio* equation, the coefficients of TR^2 are –0.010 (t = 3.67) for the FFCR model and –0.014 (t = 6.19) for the CPZC model. (Detailed results are available upon request.) These results are similar to those reported in Table 2.

4.3. Robustness test III: R^2 and Active Share (Cremers and Petajisto, 2009).

Our result on better performance of funds with lower R^2 is consistent with the findings of Cremers and Petajisto (2009) that fund performance is improved by active management, measured by Active Share (*AS*), the sum of absolute deviations of the fund's stock holdings (weights) from those of its benchmark portfolio. The data on *AS* are provided by Cremers and petajisto for 1,890 funds reporting share holdings on CDA/Spectrum and are available up to 2006.

Naturally, our measure of selectivity – R^2 – should be negatively related to *AS*. We use R^2 calculated from the CPZC benchmark model of Cremers et al. (2009) employing daily data in semi-annual frequency, and *AS* from the same frequency. We then calculate Corr($R^2_{j,t}$, $AS_{j,t}$) across funds for each half year. As expected, these correlations are all negative with a mean (median) of -0.47 (-0.50), ranging from -0.04 to -0.75. This means that on average, only a quarter of the cross-fund variability of R^2 is explained by *AS*. Thus, while the two measures of selectivity are related, they are not reflecting the very same information. Notably, *AS* measures deviations from a single benchmark index while R^2 measures deviations from multiple benchmark indexes.

To gauge the contributions of R^2 and *AS* to the prediction of mutual fund performance, we add the Active Share variable of Cremers and Petajisto (2009) to the explanatory variables of Model (3) and re-estimate the model. Because *AS* too is bounded between 0 and 1 and its distribution is negatively skewed, we apply the same transformation as we do for R^2. We define $TAS = log(AS/(1\text{-}AS))$, and as we do with R^2, we censor the upper and lower 1% tails of *AS* to remove outliers.

INSERT TABLE 4

The regression results, presented in Table 4, show that *both* R^2 and *AS* predict fund performance with the correct signs. The coefficient of TR^2 is negative and that of *TAS* is positive, both being statistically significant except in the case of the *alpha* model with weekly observations, were the coefficient of *TAS* is significant at the 0.13 level. (The median coefficient of R^2 is even more negative than the mean, while the median of *AS* is less positive than the mean.) We conclude that R^2, our measure of selectivity, provides significant contribution, in addition to that provided by *AS*, to the prediction of mutual fund performance.

4.4. Robustness Test IV: Is the R^2 effect due to selectivity or to pricing of volatility?

We directly test the *alpha*-R^2 relation for stock portfolios to see if our findings for mutual funds are due to selectivity or due to pricing of the variance components of R^2. Ang et al. (2006, Section II) find that idiosyncratic risk is negatively related to stocks' *alpha* estimated from the Fama and French (1993) model. If this also applies to stock portfolios, it would result in a *positive* R^2-*alpha* relation because R^2 is a declining function of idiosyncratic risk (relative to systematic risk). However, this is the *opposite* of what we find for mutual funds, where the R^2-*alpha* relation is negative.

We replicate our analysis of mutual fund performance on the Fama-French 100 (10x10) portfolios sorted on size and on book-to-market, which can be viewed as passive mutual funds with a constant investment strategy. The dependent variable is *alpha* or *InfRatio* and the explanatory variables are TR^2 and lagged *alpha* or *InfRatio*, all from the benchmark portfolios FFCR, using daily returns over 36 semi-annual estimation periods, 1990-2007. We mimic "style" by three dummy variables: *D-Small cap* = 1 for the smallest three size-based decile portfolios, *D-growth* = 1 for the lowest three book-to-market portfolios, and *D-value* = 1 for the highest three book-to-market portfolios (zero otherwise).

INSERT TABLE 5

The results in Table 5 show that the coefficient of TR^2 is *positive*, quite the opposite of its sign in the analysis of actively-managed mutual funds. The coefficients of TR^2 in both the *alpha* and the *InfRatio* regressions are positive, 0.006 (t = 1.77) and 0.005 (t = 1.76), respectively, both significant at the 0.10 level.[14] This is consistent with Ang et al.'s (2006) result on the negative effect of residual risk on stock expected return. Against this *positive* R^2-fund performance relation in passive portfolios, our results on a *negative* R^2-performance relation are even more remarkable. This test supports our suggestion that the negative effect of R^2 on mutual fund performance reflects the beneficial effect of its manager's strategy.

[14] We also replicate this analysis for the Fama-French 48 industry portfolios which can be viewed as "sector" passive funds. The corresponding coefficients of TR^2 for the 48 Fama-French industry-based portfolios are 0.009 (t = 0.90) for the *alpha* model and 0.002 (t = 0.62) for the *InfRatio* model.

4.5. What affects the predictive power of R^2

The predictive effect of fund R^2 on fund performance is expected to be stronger if the fund's strategy with respect to selectivity is stable and persistent. Otherwise, it would be hard to infer from a fund's R^2 in one period about its strategy and performance in the following period. To test this, we calculate $\rho(R^2)_t = Corr(R^2_{j,t}, R^2_{j,t-1})$, the correlation in each period t between the funds' R^2 and their R^2 of the previous period. The average value of $\rho(R^2)_t$ is 0.564 and the median is 0.518. We then do a regression

$$t(\gamma_t) = b_0 + b_1 * \rho(R^2)_t \quad , \tag{4}$$

where $t(\gamma_t)$ is the t-statistic of γ_t, the coefficients of $TR^2_{j,t-1}$ from the Fama-Macbeth cross-section regression of Model (3) for period t. Thus, $t(\gamma_t)$ is the coefficient γ_t standardized by its estimated standard error. We expect that in periods when R^2 is more persistent over time, lagged TR^2 better predicts performance in the subsequent period. Because the predictive effect of TR^2 is reflected in a negative coefficient, we expect that higher $\rho(R^2)_t$ is associated with a more negative $t(\gamma_t)$, implying b1 < 0. The regression has 36 half-year periods over 1990-2007.

INSERT TABLE 6

The results in Table 6 are consistent with our hypothesis. We obtain that $b_1 < 0$ for both *alpha* and *InfRatio* as measures of performance, and for both FFCR and CPZC as benchmark models. That is, R^2 predicts performance better when mutual funds' strategy was more persistent in terms of selectivity.

Next, we explore why the persistence of funds' R^2 changes over time. We do a time series regression of $\rho(R^2)_t$ on SDM_t, the standard deviation of the CRSP equally-weighted daily return during the half-year period t (measuring market volatility), and RD_t, the residual return dispersion during the half-year period, calculated as the cross-sectional average of the absolute residual returns obtained from regressing each stock daily return on the FFCR benchmark index for the half-year period (including one-day lagged returns). RD_t thus measures the idiosyncratic risk in the market. The regression results are as follows:[15]

$$\rho(R^2)_t = 0.84 + 48.97 SDM_t - 23.07 RD_t$$

(t-stat) (6.37) (3.71) (5.34) R-sqr = 0.45

[15] Standard errors are calculated using White's (1980) robust estimation.

These results show that R^2 is more persistent from one period to the next when the market volatility is higher, but it is less persistent when the average level of idiosyncratic risk in the market rises. Obviously, R^2 declines when the fund's idiosyncratic risk rises relative to the market risk. We regress MR^2_t, the semi-annual cross-funds mean of the funds' R^2, on SDM_t and RD_t and obtain the following:

$MR^2_t = 0.96 + 15.89SDM_t - 7.82RD_t$

(t-stat) (26.62) (4.18) (8.71) R-sqr = 0.63

We observe that an increase in market volatility raises R^2 whereas a market-wide increase in stocks' idiosyncratic risk lowers the average fund's R^2. This could result directly from the definition of R^2, which rises in systematic risk and declines in idiosyncratic risk, except that funds are portfolios of single stocks. We further observe that the same factors that affect the level of R^2 also affect its persistence across funds from one period to the next. These observations suggest that when the market is more volatile, funds stick to their strategy and R^2 is persistent. But when cross-stock return dispersion rises, funds can identify idiosyncratic investment opportunities and they switch into greater selectivity in their strategy.[16] This is reflected in both a decline in the funds' R^2 and in persistence in strategy. These results suggest that funds vary the extent of selectivity in their strategy according to market conditions.

5. Fund portfolios *alpha* based on sorting by lagged R^2 and *alpha*

Our analysis has shown that fund performance can be predicted by the fund lagged R^2 and lagged *alpha*. We now examine a strategy that exploits this result as follows. At the beginning of each half-year period *t*, we sort funds into five portfolios by their R^2 in the *preceding* half-year period *t-1* and then sort the funds in each R^2-quintile into five portfolios by their *alpha* in *t-1*. This sorting generates 25 (5x5) portfolios with equal number of funds in each. Both R^2 and *alpha* are calculated from daily returns using the same models that are used in Section 4 above. The included funds have at least 120 days in period *t-1*, they have $TNA_{t-1} > 15$, they invest at least 80% of their assets in common stocks, and they have a defined style and a name on CRSP. As before, we eliminate index funds by deleting those whose name includes the word "index" or the

[16] The time-series relation between the average fund *alpha* and both *SDM* and *RD* is quite insignificant.

abbreviation "ind" or a name of a recognized index. Because we do not use data on other funds characteristics as we did in the cross-section regressions, we do not require the availability of these data, nor do we censor the distribution of R^2. And, we do not require a minimum number of observations (returns) for any fund in the test period *t*, thus there is no survivorship bias problem. As a result, more funds are included and we use their data over a longer period of time than before. Our sample now consists on 33,146 fund-periods (a period is half-year) for 2,465 different funds, which is about 15% larger than the sample that we use for the cross-section regressions in Section 4.

For the test period (following the portfolio formation), we calculate the average weekly returns for the funds in each portfolio (equally-weighted). Finally, we do a single regression for each of the 25 portfolios over the entire 18-year period. The regression is of the weekly portfolio return on the weekly returns of the benchmark portfolios, using the FFCR or CPZC benchmark models.

INSERT TABLE 7

The *alpha* coefficients of the 25 portfolios and their *t*-statistics are presented in Table 7. In both Panel A (FFCR model) and Panel B (CPZC model), investing in funds in the lowest-quintile R^2_{t-1} and highest-quintile $alpha_{t-1}$ – those with the greatest selectivity and best past performance – generates positive and significant *alpha* in the subsequent period. For the FFCR model, the annualized $alpha_t$ is 2.46% (t = 2.39) and for the CPZC model, $alpha_t$ = 2.69% (t = 2.37). For the CPZC model, there are more portfolios that produce positive and significant *alpha*: the lowest-quintile R^2_{t-1} portfolio for the next-to-highest $alpha_{t-1}$ quintile has $alpha_t$ = 1.957% (t = 2.37) and two other low-R^2_{t-1} portfolios for the highest-*alpha* quintile produce positive and significant *alpha*.[17]

The rightmost column of each table presents $alpha_t$ of the low-minus-high-R^2_{t-1} portfolio. While this strategy is infeasible (because open-end funds cannot be shorted), it indicates the effect of R^2 on fund performance. We obtain that $alpha_t$ of the low-minus-high-R^2_{t-1} portfolio is positive for all $alpha_{t-1}$ quintiles, with statistical significance for the higher $alpha_{t-1}$ quintiles. This significance is higher for the CPZC model: $alpha_t$ of the

[17] The results are similar when using daily or monthly returns. For example, for monthly returns and the FFCR model, annualized *alpha* is 3.03% with t = 2.71.

low-minus-high-R^2_{t-1} portfolio is significant at the 0.05 level for the highest three $alpha_{t-1}$ quintiles and it is significant at the 0.10 for the second lowest $alpha_{t-1}$ quintile.

We estimate the same strategy using the funds *gross* returns, which reflect the skill of the fund managers in beating the benchmark portfolios before subtracting expenses. The results in Table 8 exhibit the same pattern as that for net returns in Table 7, and $alpha_t$ obtained from applying low R^2_{t-1}-high $alpha_{t-1}$ strategy is greater and more statistically significant than it is for net returns. This is consistent with the fact that low R^2 funds, which produce higher $alpha_t$, also have higher expenses (see Table 1). Thus, once we ignore expenses, the performance of low- R^2 is even better. Interestingly, it could be argued that the lower performance of high-R^2 funds is because these are effectively index funds disguised as active funds and thus charge high expenses. If this were the case, then when using *gross* returns, $alpha_t$ of the highest R^2 funds should be zero. This is indeed the case except for the lowest-$alpha_{t-1}$ quintile of funds. There, for both benchmark models, $alpha_t$ is significantly negative. This suggests that these funds are badly managed even though they are effectively indexers, and they lose money relative to the relevant indexes even before expenses. It is also possible that for these funds, actual expenses exceed the stated expenses.

INSERT TABLE 8

6. The determinants of funds' R^2

Funds choose a strategy, such as the extent of selectivity that we measure by R^2, which subsequently affects its performance. We know that this strategy is reasonably persistent, as reflected in the correlation $\rho(R^2)_t = Corr(R^2_{j,t}, R^2_{j,t+1})$, which averages 0.563 for the 36 half-year periods that we study. We now examine the fund characteristics that are systematically associated with the fund's R^2. We regress TR^2 on lagged fund characteristics – those used in Model (3) – employing the Fama-Macbeth method, for both the FFCR and the CPZC benchmarks.

INSERT TABLE 9

The results in Table 9 show that a fund's R^2 is not a random number but it is rather systematically associated with some fund characteristics. Larger funds have higher R^2, as might be expected, because such funds hold portfolios with greater breadth. We obtain an increasing and concave *TNA*-R^2 relation, as evident from the positive coefficient of *log(TNA)* and the negative coefficient of *[log(TNA)]*^{2},[18] both highly significant. Another explanation is due to the Koijen's (2008) model of fund managers who derive utility from improving their ranking or status by raising their fund size. Then, managers of smaller funds who wish to grow and attain better status and ranking have an incentive to "deviate from the pack" and employ active investment strategy (see also Krasny (2010)). Here, it means that smaller funds employ less benchmark-based and more idiosyncratic policy, producing a positive *TNA*-R^2 relation, which naturally weakens as the fund size grows.

Expenses have negative and highly significant coefficient, suggesting that funds with greater selectivity (lower R^2) and thus may incur higher expenses also charge a higher expense ratio. The negative *Expenses*-R^2 relation can also imply a greater willingness of investors to pay for more active management because it is harder for them to replicate the strategy of such funds, and because such funds have superior performance.

Fund *Turnover* has insignificant coefficient, suggesting that selectivity and active management is not reflected in higher turnover but rather in the selection of stocks that differ from the benchmarks, consistent with Cremers and Petajisto (2009).

Funds with higher *Fund Age* have lower R^2 (though this not always significant), which suggests that funds longevity is enhanced by employing a strategy of greater selectivity (lower R^2), which has been shown to produce better performance.

Managerial Tenure has a negative and highly significant coefficient, 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, herding means larger R^2 – which means that greater proportion of the risk is systematic risk – for managers with lower tenure. Also, the association between longer-tenure manager and lower R^2 may imply, as with fund age, that active management (lower R^2) helps managers survive in managing the fund for more years.

The coefficient of past performance (*alpha*) is negative and significant, especially for the model with CPZC benchmark indexes. This suggests that successful funds choose greater selectivity (lower R^2).

---

[18] The function is increasing and concave for the entire range on *TNA* in our sample.

Having suggested that R^2 is a measure of selectivity, we now test whether R^2 also reflects market timing, by which a fund raises (lowers) its investment in high-β stocks when it expects positive (negative) market return. To test that, we add to the basic FFCR daily-return model[19] either RM_t^2 (following Treynor and Mazuy, 1966) or the variable I_t*RM_t where $I_t = 1$ if $RM_t > 0$ and $I_t = 0$ otherwise (following Henriksson and Merton, 1981). *RM* is the value-weighted CRSP market return, and both regressions include the one-day lagged variables. A positive coefficient on the added variables implies that the fund engages in market timing, which should lead to lower fund's R^2 that is estimated from the original model with constant betas and without the added market-timing variables.

We add to the model of Table 9 the estimated coefficients of the RM_t^2 and RM_{t-1}^2 or of $I*RM_t$ and $I*RM_{t-1}$, and estimate their effect across funds by the Fama-Macbeth method, as we do for the other variables. The effects of these coefficients on explaining R^2, while being negative as expected, are quite insignificant. The *t*-statistic of the coefficients of RM_t^2 and RM_{t-1}^2 are, 0.89 and 0.42, respectively, and those of the coefficients of $I*RM_t$ and $I*RM_{t-1}$ are, 1.15 and 0.32. Regressing R^2 on the market timing variables alone, without all the other variables of Table 9, also produces coefficients that are statistically insignificant.

In another test of the connection between market timing and the effect of R^2 on funds' *alpha*, we regress the semi-annual γ_t from model (3) (obtained by the Fama-Mecbeth method) on RM_t over the 36 half-year periods in our sample. If the effect of R^2 on *alpha* is due to market timing, γ_t should be more negative when RM_t is higher, i.e., the coefficient of RM_t in this regression should be negative. We obtain that the coefficient of RM_t is 1.37 with $t = 0.68$, insignificant. Therefore, there is no evidence that the effect of R^2 on fund performance can be attributed to market timing.

7. Funds that include corporate bonds: predicting performance with R^2

We replicate our analysis for open-end mutual funds that hold domestic corporate bonds, excluding funds whose styles indicate that they are Treasury, government or municipal

[19] Bolen and Busse (2001) show that a daily return model is more powerful in testing for market timing ability of mutual funds.

bond funds. We include funds that invest in bonds at least 35% of their net asset value, which accommodates "Balanced Funds"[20] that are excluded from the analysis of stock funds. Our benchmark model consists of six factors based on Bessembinder et al.'s (2008) model, which includes the three Fama-French factors and two bond-spread factors that employ the Barclays (formerly Lehman) indexes: *DEF*, the difference between the return on the BBB bond index and the AAA bond index and *TERM*, the difference between the return on Treasury 30-year bond index and on the three-month Treasury bills index. We augment this five-factor by adding the return on the U.S. aggregate market value bond index, whose regression on the five-factor model results in a positive and significant *alpha*. The sixth factor is the residual return from this regression.[21] The daily data on the BBB and AAA bond indexes begin in the second half of the year 2000, hence our analysis begins in 2001.

We follow the very same procedure which we employ for stock mutual funds, estimating model (3) by the Fama-Macbeth method: we do cross-section regressions over the 14 semi-annual periods over 2001-2007 and then calculate the statistics for the resulting 14 coefficients of each variable. As before, we require at least 120 days of data in the lagged half-year period from which we estimate the fund's R^2 and 50 days in the subsequent half-year period in which we estimate the fund's performance, and we require that *TNA* > $15 million and that there are data for the variables estimated in model (3). And, we censor the extreme 1% in both tails of the distribution R^2. We thus obtain 872 pairs of half-year fund periods, an average of 62.2 observations in each cross-section regression, a smaller number than for stock funds.

The mean (median) R^2 of the funds in the final sample is 0.524 (0.487), which is much lower than it is for stock funds. There, the mean (median) R^2 based on the FFCR benchmark model is 0.882 (0.916). The range of R^2 for the bond funds is 0.150-0.983 while for the stock funds (based on FFCR model) the range is 0.334-0.990. This suggests

[20] Balanced Funds are defined by the CRSP manual as follows: "Funds whose primary objective is to conserve principal by maintaining at all times a balanced portfolio of both stocks and bonds. Typically, the stock/bond ratio ranges around 60%/40%."

[21] Specifically, we use in the order described the following Barclays indexes: US Agg Corp AAA DUP LHIGAAA (LHUC3AD), US Agg Corp BBB DUP LHIGBAA (LHUC3BD), US Treasury Bellwethers 30Y (LHTBL30), US Treasury Bellwethers 3M (LHTBW3M), and US Agg – Market Value- (LHAGGBD(MV)).

that many bond funds pursue investment policies (perhaps odd ones) whose return generating process is not well captured by the six-factor model.

We estimate the cross-section model (3) where *alpha* and *InfRatio* as dependent variables. The explanatory variable of interest is TR^2_{t-1}, the transformed value of lagged R^2_{t-1}. The model also includes 10 style dummy variables.[22] Table 10 presents the estimation results.

INSERT TABLE 10

The mean value of γ_t, the coefficient of TR^2_{t-1}, is –0.834 with t = 3.74, highly significant. The median γ_t is –1.014, and 12 out of the 14 estimated coefficients are negative, significantly different from a chance result of 7/14. In addition, the coefficients of $Expenses_{t-1}$ and $alpha_{t-1}$ are, respectively, negative and positive with statistical significance, while the coefficients of the other variables are statistically insignificant. This is similar to the results for stock funds. We re-estimate the model, replacing TR^2_{t-1} by the untransformed $R^2_{t-1.}$ The mean (median) coefficient of R^2_{t-1} is –4.268 (–5.007) with t = 3.98.

With *InfRatio* as the fund performance measure, we obtain that the mean value of γ_t is –0.014 with t = 2.08, which is marginally significant. The median γ_t is –0.012, and 11 out of the 14 estimated coefficients are negative, significant at the 0.05 level. Again replacing TR^2_{t-1} by the untransformed R^2_{t-1}, we obtain that the mean (median) coefficient is –0.064 (–0.067) with t = 2.10.

It is noteworthy that the estimated coefficient γ_t of TR^2_{t-1} for these bond funds is close in magnitude to that of stock funds. The mean coefficient γ_t in the bond alpha model is –0.834, which is the middle of the range of respective estimate for stock funds, –0.596 and –1.042 for the FFCR and CPZ4 models, respectively. In the *InfRatio* model, the mean coefficient γ_t for bond funds, –0.014, is also within the range –0.010 and –0.013 for the two stock funds models.

As a robustness check, we use the four-factor benchmark model of Elton et al. (1995). It includes the access returns on the stock market index (CRSP value-weighted

---

[22] The styles are: Corporate Debt A Rated, Corporate Debt BBB-Rated, Intermediate Investment Grade Debt, Short-Term Investment Grade Debt, Short-Intermediate Investment Grade Debt, High Current Yield, Balanced, General Bond, Income (including flexible and multi-sector) , Flexible Portfolio, and missing styles.

index) and on the aggregate bond market and two spread factors, *DEF* and *OPTION*. The latter is defined as the return spread between the Barclays GNMA index and the Barclays Government Intermediate index. Their four-factor return model has a lower average R^2 than our six-factor model. We re-do all our estimations using this four factor model. The results are similar in terms of the sign and significance of the effect of lagged R^2 on fund *alpha*. In particular, in estimating Model (3), the mean coefficient of TR^2_{t-1} is -0.769 with t = 2.95. For performance measured by *InfRatio*, the mean coefficient of TR^2_{t-1} is -0.015 with t = 2.14. The results suggest that lagged R^2 is a significant predictor of fund performance also when using as a benchmark model the factor model of Elton et al. (1995).

8. 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 a multi-factor model, predicts the following period'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 fund performance. This is obtained after controlling for commonly-used fund characteristics and past fund performance. The results hold for both annual and semi-annual frequency and using both daily and weekly returns.

Our analysis is shown to be robust to the indexes used. We estimate our model using a number of benchmark multi-factor models and find throughout that R^2 has a negative and highly significant coefficient in regressions predicting next period *alpha* or *InfRatio*.

We are able to identify a portfolio of funds that produces positive and significant *alpha*. At the end of each period, we sort the funds by their estimated R^2 and by *alpha* and divide them into 25 (5x5) portfolios. We next regress the time series weekly return of each portfolio on the two major benchmark portfolios that we use and find that the lowest-R^2 and highest-*alpha* portfolio of funds generates a positive and significant annualized *alpha* of about 2.5%.

Fund R^2 is negatively related to another measure of active fund management and selectivity developed by Cremers and Petajisto (2009), 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 and manager tenure. Also, funds with higher past *alpha* have subsequently lower R^2.

Our method of predicting fund performance by its lagged R^2 holds also for mutual funds that hold corporate bonds. Replicating for such funds our analysis that has been done for stock mutual funds shows that the predictive effect of R^2 is similar

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

References

Agarwal, V. V. Fos and W. Jiang, 2010, Inferring Reporting Biases in Hedge Fund Databses from Hedge Fund Equity Holdings, Working paper, Columbia Business School.

Ang, A., R.J. Hodrick, Y. Xing and X. Zhang, 2006, The Cross-Section of Volatility and Expected Returns. *Journal of Finance* 61, 259-299.

Bessembinder, H., K.M. Kahle, W. F. Maxwell and D. Xu, 2008, Measuring Abnormal Bond Performance. *Review of Financial Studies*, forthcoming.

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

Bollen P.B. and J.A. Busse, 2001, On the Timing Ability of Mutual Fund Managers, *Journal of Finance* 56, 1075-1094.

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.

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

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, 2009, How Active is Your Fund Manager? A New Measure That Predicts Performance, *Review of Financial Studies*, forthcoming

Cremers, M., Petajisto, A. and E. Zitzewitz, 2010, Should Benchmark Indices Have Alpha? Revisiting Performance Evaluation, Working Paper, Yale School of Management

Da, Z., P. Gao, and R. Jagganathan, 2009, Impatient Trading, Liquidity Provision, and Stock Selection by Mutual Funds. Working paper.

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. Gruber and C. R. Blake, 1995, Fundamental Economic Variables, Expected Returns, and Bond Fund Performance. *Journal of Finance* 50, 1229-1256.

Elton, E. J., M. J. Gruber, 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.

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.

Henriksson, R. D., and R.C. Merton, 1981, On Market Timing and Investment Performance. II. Statistical Procedures for Evaluating Forecasting Skills, *Journal of Business* 54, 513-533.

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.

Treynor, J.L., and K. Mazuy, 1966, Can Mutual Funds Outguess the Market? *Harvard Business Review* 44, 131-136.

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 1. Summary Statistics

The table presents summary statistics on actively managed equity mutual funds included in our sample. R^2 is obtained from the regression of daily funds returns on FFCR or CPZC factor daily returns and their one-day lagged values over a year, and $TR^2 = log(\sqrt{R^2}/(1- \sqrt{R^2}))$ computed for each benchmark specification. 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. Panel B presents the correlation matrix between fund characteristic variables. The cross-sectional correlations are computed by year and then average across 18 years. The significance of correlations is evaluated using Swinscow (1997, Ch.11) test statistics as $t= \rho\sqrt{\frac{N-2}{1-\rho^2}}$, where ρ is an estimated correlation, and N is the sample size. The pre-estimation sample period is from 1/1989 to 12/2006.

Panel A: Fund characteristics

		Mean	Median	Minimum	Maximum
Total number of funds: 2,026					
TNA (total net assets, in $millions)		1,314.08	263.94	15.1	105,938.5
Fund Age (years)		12.95	8.33	0.5	78.28
Expenses (%)		1.28	1.24	0.01	8.36
Turnover (%)		88.79	66.00	0.2	3,727
Manager Tenure (years)		6.06	4.83	0.5	62.92
R^2	FFCR	0.86	0.90	0.29	0.988
	CPZC	0.86	0.90	0.28	0.991
TR^2	FFCR	2.89	2.96	0.16	5.137
	CPZC	2.92	2.97	0.12	5.386

Panel B: Average Cross-Sectional Correlation

	Log(TNA)	*Age*	*Expenses*	*Turnover*	*Log(Manager Tenure)*	R^2 *FFCR/ CPZC*	TR^2 *FFCR/ CPZC*
Log(TNA)	1.00						
Log(Fund Age)	0.42*	1.00					
Expenses	-0.36*	-0.23*	1.00				
Turnover	-0.10*	-0.12*	0.17*	1.00			
Log(Manager Tenure)	0.20*	0.32*	-0.11*	-0.20*	1.00		
R^2 FFCR	0.13*	0.02	-0.14*	-0.02	-0.10*	1.00	
CPZ	0.13*	0.02	-0.14*	-0.02	-0.10*	1.00	
TR^2 FFCR	0.15*	0.03	-0.18*	-0.04*	-0.10*	0.94*	1.00
CPZC	0.16*	0.04*	-0.18*	-0.05*	-0.10*	0.93*	1.00

* denotes 5% significance, using Swinscow (1997) test statistics

Table 2. The effect of R^2 on fund performance: Daily data, annual and semi-annual frequencies

The dependent variables are two fund performance variables: *alpha*, the intercept from an annual and semi-annual regression of daily fund return on the returns of the benchmark factors FFCR or CPZC and their lagged values, and *InfRatio* = *alpha*/*RMSE*, where *RMSE* is the root mean squared error from these regressions. The benchmark model FFCR consists of Fama-French-Carhart factors and Russell 2000 (orthogonalized to FFC factors) or CPZC suggested by Cremers et al. (2009) and augmented by Carhart's momentum factor. All independent variables are as of the end of the previous year/half-year. $TR^2 = log(\sqrt{R^2}/(1- \sqrt{R^2}))$, where R^2 is obtained from the above regressions. The Total Net Assets (*TNA*) in \$mm, *Expenses* and *Turnover* are as of the end of the year/half-year. *Fund Age* is the number of years since the fund was first offered. *Manager Tenure* is the number of years since the current manager took control. The estimation is done by the Fama-Macbeth method. The numbers presented are the means of the coefficients. Their *t*-statistics are in parentheses with the corresponding *p*-values in square brackets below. "Med" is the median coefficient. "Neg" is the proportion of negative coefficients. The superscript + indicates that the proportion of negative coefficients is significantly different than the chance proportion of 0.5. The sample period is from 1/1990 to 12/2007.

Variables lagged one year	*ANNUAL*				*SEMI-ANNUAL*			
	FFCR		CPZC		FFCR		CPZC	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	alpha	*InfRatio*	*alpha*	*InfRatio*	*alpha*	*InfRatio*	*alpha*	*InfRatio*
TR^2	-0.586	-0.010	-1.003	-0.011	-0.596	-0.010	-1.042	-0.013
	(2.38)	(3.78)	(3.49)	(5.08)	(3.55)	(4.81)	(5.56)	(7.62)
	[0.029]	[0.002]	[0.003]	[0.000]	[0.001]	[0.000]	[0.000]	[0.000]
	Med: -0.502	Med: -0.008	Med: -0.549	Med: -0.010	Med: -0.617	Med: -0.010	Med:-1.076	Med: -0.014
	Neg: 14/18$^+$	Neg: 15/18$^+$	Neg: 17/18$^+$	Neg: 17/18$^+$	Neg: 27/36$^+$	Neg: 30/36$^+$	Neg: 30/36$^+$	Neg: 31/36$^+$
Expenses	-0.581	-0.011	-0.837	-0.013	-0.746	-0.014	-1.129	-0.017
	(2.13)	(4.38)	(2.66)	(5.52)	(2.59)	(4.44)	(3.74)	(6.15)
	[0.048]	[0.000]	[0.017]	[0.000]	[0.014]	[0.000]	[0.001]	[0.000]
Log(TNA)	0.175	0.001	0.158	-0.001	1.326	-0.0004	1.495	-0.001
	(0.42)	(0.16)	(0.43)	(0.22)	(1.02)	(0.11)	(0.94)	(0.26)
	[0.679]	[0.871]	[0.673]	(0.828)	[0.316]	[0.915]	[0.353]	[0.794]
$Log(TNA)^2$	-0.010	-0.0001	-0.010	0.0001	-0.096	$4*10^{-5}$	-0.112	0.0001
	(0.30)	(0.20)	(0.32)	(0.20)	(0.94)	(0.09)	(0.91)	(0.32)
	[0.771]	[0.871]	[0.752]	[0.846]	[0.352]	[0.927]	[0.368]	[0.748]
Turnover	0.002	$-3*10^{-6}$	0.001	$-1*10^{-5}$	0.002	$-3*10^{-6}$	0.002	$-1*10^{-5}$
	(0.43)	(0.14)	(0.19)	(0.52)	(0.64)	(0.13)	(0.54)	(0.42)
	[0.671]	[0.887]	[0.852]	[0.609]	[0.526]	[0.898]	[0.594]	[0.679]
Log(Fund Age)	-0.188	-0.002	-0.182	-0.002	-0.217	-0.003	-0.233	-0.004
	(1.20)	(1.41)	(1.27)	(1.25)	(1.58)	(1.90)	(1.45)	(2.00)
	[0.245]	[0.178]	[0.223]	[0.228]	[0.123]	[0.065]	[0.155]	[0.054]
Log(Manager	-0.141	-0.001	-0.142	-0.002	-0.054	0.002	-0.103	-0.0003
Tenure)	(1.04)	(1.21)	(1.06)	(1.31)	(0.34)	(0.15)	(0.60)	(0.19)
	[0.314]	[0.244]	[0.304]	[0.209]	[0.734]	[0.884]	[0.552]	[0.852]
Dependent	0.157	0.158	0.169	0.185	0.086	0.120	0.090	0.134
variable, Lagged	(4.49)	(5.68)	(4.79)	(5.86)	(2.36)	(5.90)	(2.16)	(6.03)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.024]	[0.000]	[0.037]	[0.000]
R^2	0.20	0.25	0.21	0.24	0.19	0.22	0.21	0.22

$^+$ indicates significance at the 1% level

Table 3. The effect of R^2 on fund performance: Weekly data, annual frequency

This table presents the same tests as those in Table 2, except that here returns are weekly instead of daily. See the legend in Table 2 for details. The estimation frequency is annual.

Variables lagged one year	FFCR		CPZC	
	alpha	*InfRatio*	*alpha*	*InfRatio*
TR^2	-0.586	-0.020	-0.981	-0.025
	(2.82)	(3.97)	(3.18)	(4.26)
	[0.012]	[0.001]	[0.005]	[0.001]
	Med: -0.413	Med: -0.015	Med: -0.587	Med: -0.021
	Neg: $13/18^+$	Neg: $16/18^+$	Neg: $14/18^+$	Neg: $17/18^+$
Expenses	-0.972	-0.031	-1.060	-0.036
	(3.55)	(4.07)	(3.18)	(4.71)
	[0.002]	[0.001]	[0.005]	[0.000]
Log(TNA)	0.370	-0.004	0.081	-0.004
	(0.97)	(0.43)	(0.23)	(0.47)
	[0.345]	[0.670]	[0.819]	[0.644]
$Log(TNA)^2$	-0.025	0.0003	-0.008	0.0003
	(0.85)	(0.38)	(0.28)	(0.36)
	[0.405]	[0.706]	[0.783]	[0.724]
Turnover	0.001	$-2*10^{-5}$	0.002	$-2*10^{-5}$
	(0.33)	(0.35)	(0.48)	(0.36)
	[0.746]	[0.730]	[0.641]	[0.722]
Log(Fund Age)	-0.290	-0.007	-0.241	-0.006
	(1.35)	(1.29)	(1.34)	(1.06)
	[0.195]	[0.213]	[0.198]	[0.305]
Log(Manager Tenure)	-0.212	-0.003	-0.225	-0.004
	(1.57)	(0.94)	(1.75)	(1.25)
	[0.135]	[0.358]	[0.099]	[0.227]
Dependent variable, Lagged	0.166	0.171	0.161	0.182
	(4.72)	(6.20)	(4.67)	(5.93)
	[0.000]	[0.000]	[0.000]	[0.000]
R^2	0.22	0.25	0.22	0.25

Table 4. The effects of R^2 and Active Share on fund performance

This table presents the same tests as those in Table 2, using daily data with semi-annual frequency of estimation. The model is the same as in Table 2, with the addition of Active Share (*AS*, due to Cremers and Petajisto (2009), the sum of absolute deviations of the fund's stock holdings (weights) from those of its benchmark portfolio. The benchmark indexes are CPZC (Cremers et al. (2010). Data are available for 1990-2006. To save space, the table includes only the estimated coefficients of TR^2 and $TAS = log(AS/(1-AS))$. See the legend in Table 2 for details.
To save space, we do not present the coefficients of all other variables.

Panel B.

Variables lagged one year	CPZC Daily data (semi-annual)		CPZC Weekly data (annual)	
	alpha	*InfRatio*	*alpha*	*InfRatio*
TR^2	-0.717	-0.008	-0.774	-0.014
	(3.40)	(3.49)	(2.91)	(2.27)
	[0.002]	[0.001]	[0.010]	[0.037]
	Med: -0.707	Med: -0.011	Med: -0.333	Med: -0.010
	Neg: 25/35$^+$	Neg: 26/35$^+$	Neg: 15/18$^+$	Neg: 14/18$^+$
TAS	0.710	0.010	0.393	0.017
	(2.62)	(4.33)	(1.60)	(3.32)
	[0.013]	[0.000]	[0.129]	[0.004]
	Med: 0.294	Med: 0.007	Med: 0.139	Med: 0.017
	Neg: 10/35$^+$	Neg: 7/35$^+$	Neg: 7/18	Neg: 3/18$^+$
Other variables (see Table 2)	Yes	Yes	Yes	Yes

Table 5. The effect of R^2 on the Performance of Fama-French 100 Portfolios, Sorted on Size and Book/Market

Regressions of *alpha*, the intercept from a half-year regression of daily excess returns on Fama-French 100 portfolios (sorted on size and book-to-market) on FFCR factor daily returns *RM-Rf*, *SMB*, *HML*. *MOM* (momentum), Russell 2000 (orthogonalized) and their lagged values, and *InfRatio* = *alpha*/*RMSE*, where *RMSE* is the root mean squared error from these regressions. All explanatory variables are as of the end of the previous half-year. $TR^2 = log(\sqrt{R^2}/(1- \sqrt{R^2}))$, where R^2 is obtained from the above regressions. The regression of the 100 Fama-French portfolios includes dummy variables that mimic style. *D-small cap* equals 1 for the 3 smallest size portfolios and zero otherwise, *D-growth* equals 1 for the lowest 3 book/market portfolios and zero otherwise, and *D-value* equals 1 for the highest 3 book/market portfolios and zero otherwise. The estimation is done at semi-annual frequency (36 periods in all) by the Fama-MacBeth method. Presented are the means of the coefficients and their *t*-statistics in parentheses, with the corresponding *p*-values in the square brackets below. The sample period from 1/1990 to 12/2007.

Explanatory variables (lagged)	Performance measures	
	Alpha	*InfRatio*
TR^2	0.006	0.005
	(1.77)	(1.76)
	[0.085]	[0.087]
Alpha	0.029	
	(1.24)	
	[0.222]	
InfRatio		0.066
		(2.91)
		[0.006]
D-small cap	-0.006	-0.002
	(0.61)	(0.24)
	[0.547]	[0.809]
D-growth	-0.017	-0.015
(low	(1.69)	(2.14)
book/mkt)	[0.099]	[0.039]
D-value	-0.004	-0.0002
(high	(0.55)	(0.04)
book/mkt)	[0.583]	[0.968]
R-sqr	0.11	0.12

Table 6: The Effect of Persistence in Fund R^2 on the Prediction of Performance

The table presents the results from the regression model $t(\gamma_t) = b_0 + b_1 * \rho(R^2)_t$, where $t(\gamma_t)$ is the *t*-statistic of the slope coefficient of $TR^2_{j,t-1}$ in the semi-annual cross-sectional Fama-Macbeth regression model (3) of $alpha_{j,t}$ or $InfRatio_{j,t}$, and $\rho(R^2)_t = Corr(R^2_{j,t}, R^2_{j,t-1})$, the cross-fund correlation between $R^2_{j,t}$ of fund *j* in period *t* and $R^2_{j,t-1}$, the value for fund in the previous period. The estimations here are at semi-annual frequency. The table presents the estimated coefficients and their *t*-statistics. The benchmark models are FFCR (Fama-French-Carhart and Russell 2000) or CPZC suggested by Cremers et al. (2010) and augmented by Carhart's momentum factor. *, ** indicates significance at the 0.05 or 0.01 level.

	FFCR		CPZC	
	$t(\gamma_t)$ of *alpha* model	$t(\gamma_t)$ of *InfRatio* model	$t(\gamma_t)$ of *alpha* model	$t(\gamma_t)$ of *InfRatio* model
b_0	0.221 (0.32)	1.395 (1.50)	0.421 (0.46)	0.635 (0.78)
b_1	-2.206 (1.98)*	-5.690 (3.74)**	-4.133 (2.78)**	-5.588 (4.27)**
R-sqr	0.11	0.29	0.19	0.35

Table 7. Fund *alphas*, Sorting on lagged R^2 and *Alpha*: Net Weekly Returns

The table presents the *alpha* of each portfolio. Portfolios are formed by sorting all funds semi-annually into quintiles by R^2 and within that by *alpha* based on the previous half-year (t-1) estimation period. R^2 and *alpha* are estimated for each fund from daily returns by regressing fund returns of the returns of the benchmark factors FFCR or CPZC and their lagged values. FFCR consists of Fama-French-Carhart factors and Russell 2000 (orthogonalized to FFC factors) and CPZC is the set of indexes suggested by Cremers et al. (2009), augmented by Carhart's momentum factor. After sorting, we calculate the equally-weighted weekly net excess return for each of 25 portfolios for the following half-year period t, and regress these weekly returns on the weekly returns of the benchmark factors, FFCR or CPZC for the entire 18-year period. For each portfolio (cell) we present *alpha*, the intercept from this regression, and its *t*-statistic from this regression. The sample period is from 1/1990 to 12/2007. **, * denotes significance at the 5% or 10% level.

Panel A. FFCR $alpha_y$

	R^2_{t-1}						
$Alpha_{t-1}$	Low	2	3	4	High	**All**	**Low-High**
Low	-1.409 (0.97)	-0.487 (0.47)	-0.640 (0.63)	-2.010 (2.29)**	-3.010 (4.53)**	-1.486 (1.90)*	1.601 (1.08)
2	-0.476 (0.52)	-0.472 (0.56)	0.323 (0.42)	-0.487 (0.66)	-1.260 (2.39)**	-0.462 (0.83)	0.784 (0.79)
3	0.436 (0.50)	0.117 (0.15)	-0.751 (1.11)	-0.646 (0.95)	-0.893 (1.82)*	-0.393 (0.85)	1.329 (1.46)
4	0.447 (0.53)	0.392 (0.49)	0.992 (1.25)	-0.758 (1.02)	-1.138 (2.06)**	-0.007 (0.02)	1.585 (1.54)
High	2.457** (2.39)	1.866 (1.60)	2.053* (1.71)	1.754* (1.88)	-0.627 (0.82)	1.544** (2.13)	3.083** (2.26)
All	0.267 (0.36)	0.303 (0.49)	0.390 (0.66)	-0.441 (0.78)	-1.375** (3.07)	-0.161 (-0.38)	1.642 (1.92)*
High-Low	3.865** (2.36)	2.353 (1.52)	2.693* (1.69)	3.764** (3.12)	2.383** (2.83)	3.030** (2.84)	

Panel B. CPZC $alpha_y$

	R^2_{t-1}						
$Alphat_{t-1}$	Low	2	3	4	High	**All**	**Low-High**
Low	-0.738 (0.53)	-1.661 (1.55)	-2.015* (1.66)	-2.174** (2.39)	-2.640** (4.24)	-1.827** (2.16)	1.903 (1.37)
2	-0.078 (0.08)	-0.769 (1.01)	-0.892 (1.02)	-2.313** (2.86)	-1.946** (4.03)	-1.192** (2.02)	1.868* (1.95)
3	1.201 (1.48)	0.890 (1.12)	-0.387 (0.51)	-2.094** (2.80)	-1.554** (3.57)	-0.418 (0.83)	2.755** (3.26)
4	1.957** (2.37)	1.149 (1.36)	0.613 (0.78)	-0.668 (0.92)	-0.874* (1.69)	0.448 (0.90)	2.832** (2.99)
High	2.693** (2.37)	2.495** (2.17)	3.036** (2.73)	1.070 (1.16)	-0.836 (1.26)	1.729** (2.34)	3.530** (2.59)
All	1.011 (1.41)	0.446 (0.67)	0.050 (0.08)	-1.220* (1.94)	-1.554** (3.68)	-0.248 (0.50)	2.564** (3.30)
High-Low	3.431** (1.97)	4.155** (2.85)	5.051** (3.25)	3.244** (2.83)	1.804** (2.53)	3.556** (3.39)	

Table 8. Fund Performance, Sorting on R^2 and *Alpha*: Gross Weekly Returns

The table presents results for a procedure similar to that is Table 7, except that here we use in the final regression the portfolio gross returns, obtained by adding to net returns the average weekly expense ratio. **, * denotes significance at the 5% or 10% level.

Panel A. FFCR $alpha_y$

	R^2_{t-1}						
$Alpha_{t-1}$	Low	2	3	4	High	**All**	**Low-High**
Low	0.004 (0.00)	0.868 (0.83)	0.716 (0.71)	-0.793 (0.90)	-1.844** (2.77)	-0.185 (0.24)	1.848 (1.24)
2	0.821 (0.90)	0.881 (1.05)	1.552** (2.02)	0.657 (0.89)	-0.171 (0.33)	0.761 (1.37)	0.993 (1.01)
3	1.702** (1.96)	1.383* (1.77)	0.464 (0.69)	0.496 (0.73)	0.154 (0.31)	0.795* (1.73)	1.548* (1.71)
4	1.725** (2.04)	1.632** (2.02)	2.197** (2.77)	0.388 (0.52)	-0.121 (0.22)	1.171** (2.60)	1.846* (1.80)
High	3.847** (3.73)	3.165** (2.72)	3.351** (2.79)	2.925** (3.15)	0.425 (0.56)	2.779** (3.82)	3.422** (2.51)
All	1.596** (2.16)	1.602** (2.62)	1.649** (2.79)	0.722 (1.27)	-0.301 (0.67)	1.065** (2.51)	1.897** (2.22)
High-Low	3.843** (2.34)	2.297 (1.48)	2.635* (1.65)	3.718** (3.09)	2.269** (2.69)	2.964** (2.77)	

Panel B. CPZC $alpha_y$

	R^2_{t-1}						
$Alpha_{t-1}$	Low	2	3	4	High	**All**	**Low-High**
Low	0.713 (0.51)	-0.293 (0.27)	-0.674 (0.56)	-0.948 (1.04)	-1.472** (2.37)	-0.516 (0.61)	2.185 (1.57)
2	1.212 (1.24)	0.544 (0.71)	0.376 (0.43)	-1.156 (1.43)	-0.863* (1.79)	0.032 (0.05)	2.075** (2.16)
3	2.470** (3.05)	2.197** (2.77)	0.819 (1.08)	-0.946 (1.26)	-0.501 (1.15)	0.778 (1.55)	2.970** (3.52)
4	3.248** (3.92)	2.432** (2.88)	1.806** (2.29)	0.459 (0.63)	0.114 (0.22)	1.627** (3.26)	3.134** (3.32)
High	4.069** (3.57)	3.808** (3.32)	4.268** (3.83)	2.253** (2.44)	0.199 (0.30)	2.952** (3.99)	3.870** (2.84)
All	2.348** (3.28)	1.762** (2.66)	1.293* (1.93)	-0.051 (0.08)	-0.488 (1.16)	0.978** (2.00)	2.836** (3.65)
High-Low	3.356* (1.93)	4.101** (2.81)	4.942** (3.19)	3.201** (2.79)	1.671** (2.34)	3.468** (3.30)	

Table 9. Determinants of R^2

The table presents Fama-MacBeth regressions of $TR^2 = log(\sqrt{R^2}/(1- \sqrt{R^2}))$, where R^2 is obtained from the semi-annual regression of daily fund excess returns on FFCR or CPZC factors their lagged values. All independent variables are as of the end of the previous half-year. The performance measure *alpha* is the intercept from the above regressions. The Total Net Assets (*TNA*) in $mm, *Expenses* and *Turnover* are as of the end of the half-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 9 style dummy variables. The numbers presented are the means of the coefficients and the Fama-MacBeth *t*-statistics are in parentheses with the corresponding p-values in the square brackets. The sample period is from 1/1990 to 12/2007.

Variables lagged half-year	***FFCR***	**CPZC**
Expenses	-0.207	-0.232
	(5.70) [0.000]	(6.40) [0.000]
Log(TNA)	0.212	0.221
	(6.39) [0.000]	(6.47) [0.000]
Log(TNA)^2	-0.010	-0.011
	(3.64) [0.001]	(3.84) [0.001]
Turnover	0.0002	0.0001
	(1.29) [0.204]	(0.71) [0.483]
Log(Fund Age)	-0.019	-0.025
	(1.45) [0.157]	(2.08) [0.045]
Log(Manager Tenure)	-0.132	-0.137
	(11.61) [0.000]	(11.73) [0.000]
Alpha	-0.002	-0.006
	(2.06) [0.047]	(4.42) [0.000]
Style dummy variables:		
Aggressive Growth	2.128	2.098
	(10.16) [0.000]	(10.51) [0.000]
Equity Income	1.843	1.911
	(6.62) [0.000]	(6.80) [0.000]
Growth	2.472	2.517
	(10.88) [0.000]	(11.60) [0.000]
Long term growth	1.408	1.442
	(7.34) [0.000]	(7.63) [0.000]
Growth and Income	2.723	2.794
	(13.76) [0.000]	(14.72) [0.000]
Mid-Cap	2.041	1.971
	(8.88) [0.000]	(8.97) [0.000]
Micro-Cap	1.288	1.212
	(5.33) [0.000]	(5.35) [0.000]
Small Cap	2.340	2.286
	(11.28) [0.000]	(11.65) [0.000]
Maximum Capital Gains	2.107	2.201
	(11.66) [0.000]	(11.80) [0.000]
R-sqr	0.93	0.93

Table 10. The effect of R^2 on BOND fund performance:

This table presents the same tests as those in Table 2 for bond mutual funds for the years 2001-2007. The estimation frequency is semi-annual. See the legend in Table 2 for details. Included funds have at least 35% of their holdings in corporate bonds, and they do not include Treasury, government or municipal bond funds.

Variables lagged one year	*alpha*	*InfRatio*
TR^2	-0.834	-0.014
	(3.74)	(2.08)
	[0.002]	[0.058]
	Med: -1.014	Med: -0.012
	Neg: 12/14$^+$	Neg: 11/14$^+$
Expenses	-0.962	-0.027
	(2.66)	(2.51)
	[0.020]	[0.026]
Log(TNA)	0.349	0.006
	(0.51)	(0.58)
	[0.616]	[0.572]
$Log(TNA)^2$	-0.022	-0.0005
	(0.36)	(0.53)
	[0.723]	[0.603]
Turnover	0.001	$-4*10^{-6}$
	(1.04)	(0.13)
	[0.319]	[0.900]
Log(Fund Age)	0.006	-0.001
	(0.03)	(0.13)
	[0.977]	[0.897]
Log(Manager Tenure)	0.124	0.002
	(0.80)	(0.42)
	[0.440]	[0.681]
Dependent variable,	0.265	0.247
Lagged	(4.07)	(4.51)
	[0.001]	[0.001]
R^2	0.59	0.58