
Working Paper No. 09-15

Economics and Finance Working Paper Series

Antonios Antypas, Guglielmo Maria Caporale,
Nikolaos Kourogenis and Nikitas Pittis

**Selectivity, Market Timing and the
Morningstar Star-Rating System**

February 2009

Selectivity, Market Timing and the Morningstar Star-Rating System

Antonios Antypas^(a), Guglielmo Maria Caporale^{(b)*},

Nikolaos Kourogenis^(a) and Nikitas Pittis^(a)

(a) Dept. of Banking and Financial Management,
University of Piraeus[†]

(b) Centre for Empirical Finance, Brunel University, London

Abstract

This paper evaluates the Morningstar mutual fund ranking system. We find that indeed higher Morningstar ratings are associated with higher returns on the portfolios including respectively five-, four-, three-, two- and one-star funds only (STAR5 to STAR1). We then perform an unconditional and conditional portfolio performance evaluation. In both cases the evidence suggests that the better performance of the STAR3, STAR4 and STAR5 categories reflects superior stock selection rather than market timing abilities. Overall, the implication for the Morningstar ranking system is that this is most effective in identifying the worst-performing funds (STAR1 or STAR2) rather than the best-performing ones.

Keywords: Mutual Fund, Morningstar Star-Rating System, CAPM, Conditional and Unconditional Portfolio Performance Evaluation

JEL classification: G23

***Corresponding author:** Professor Guglielmo Maria Caporale, Centre for Empirical Finance, Brunel University, West London, UB8 3PH, UK . Tel.: +44 (0)1895 266713. Fax: +44 (0)1895 269770. Email: Guglielmo-Maria.Caporale@brunel.ac.uk

[†]**Acknowledgements:** This research was co-funded by the European Social Fund and the Greek Ministry of National Education and Religious Affairs: PROJECT "PYTHAGORAS: Funding of research groups in the University of Piraeus".

1 Introduction

Designing appropriate methods to measure mutual fund (more generally, managed portfolio) performance is an unresolved issue despite the existence of a large body of literature on this topic (see, e.g., Hendricks et al., 1993, Goetzmann and Ibbotson, 1994, and Brown and Goetzmann, 1995), as this requires overcoming a number of difficulties. Traditional, unconditional approaches (see, e.g., Jensen, 1972) have been shown to be unreliable, in the sense that they are unable to distinguish between common time-varying risk (premia) and performance of individual portfolios. An alternative approach was suggested by Ferson and Schadt (1996), who put forward a conditional performance evaluation method. They introduce conditioning (public) information variables into the model and are able to estimate time-varying conditional betas. Their key point is that if it is possible to replicate a managed portfolio strategy using publicly available information then such a portfolio cannot be deemed to outperform the others - in other words, superior information and/or market timing ability cannot be invoked as an explanation in the presence of time-varying risk (premia) which cannot be distinguished from average performance. Applying their model to data for 67 mutual funds over the period 1968-1990, Ferson and Schadt (1996) find that the estimated alphas are close to zero (rather than negative as in the unconditional framework of Jensen, 1968 or Elton et al., 1992); also, they find no evidence of the negative market timing performance reported by previous studies (such as Treynor and Mazuy, 1966 and Merton and Henriksson, 1981), which had interpreted it as an indication of poor performance. Overall, a conditional model leads to the conclusion that funds perform much better than it would be inferred on the basis of a traditional, unconditional evaluation, which overlooks a possibly non-zero covariance between the betas and market returns.

The paper by Ferson and Schadt (1996) and most other studies focus on the performance of mutual funds themselves. By contrast, very little attention has been paid to the useful-

ness of the Morningstar star-rating system of mutual funds, which is increasingly used by investors to select mutual funds (and as a predictor of future performance, despite the emphasis put by Morningstar on "achievement"). Its importance was documented, for instance, by a survey reported by Damato (1996) in the Wall Street Journal. A few exceptions are the papers by Blume (1998), Sharpe (1998), Khorana and Nelling (1998), and finally Blake and Morey (2000). The latter is most interesting in that, rather than analysing persistence only, it examines the predictive ability of Morningstar ratings for mutual fund performance. The conclusion of this study is that low ratings are indeed associated with poor future performance, whilst it is not at all clear that very high ratings produce a better future performance than slightly lower or average ratings. All in all, Morningstar ratings by themselves appear to have only a slight advantage over alternative methods to predict future fund performance. An intriguing idea has more recently been put forward by Del Guercio and Tkac (2008), who apply an event-study methodology to analyse more than 10,000 Morningstar star-rating *changes* - their evidence suggests that it is the *change* in the star-rating (as opposed to the rating itself) which affects investment flows into or out of mutual funds.

The present paper focuses on whether Morningstar ratings themselves enable investors to select funds that are likely to exhibit superior performance in the future. Instead of relying on out-of-sample performance measures as in Blake and Money (2000), we conduct a full unconditional as well as conditional performance evaluation of the Morningstar rating system using the framework advocated by Ferson and Schadt (1996). More in detail, we proceed as follows. We create five alternative portfolios of funds (funds-of-funds). The first portfolio, named STAR1, consists of all the funds that in each time period, t , are rated one-star by Morningstar. To be more specific, in period $t=1$ (the first period in our sample) we invest an amount A in a portfolio consisting of all the funds (equally weighted) that have been given one-star by Morningstar in period $t=1$. In period $t=2$, the amount $(1+R_1^{1*}) \times A$ (R_1^{1*}

being the return of the portfolio between periods 1 and 2) is invested again in a portfolio consisting solely of funds that in period 2 were rated one-star by Morningstar. We continue this process until we reach period $t=T$, i.e. the last period of our sample. In this way, we obtain a series of returns $R_1^{1*}, R_2^{1*}, \dots, R_T^{1*}$ generated by investing exclusively in one-star funds. These are interpreted as being a random vector from the process $\{R_t^{1*}\}$ generating one-star portfolio returns. We repeat the same procedure for two-, three-, four- and five-star funds, thus obtaining samples from the returns processes $\{R_t^{2*}\}$, $\{R_t^{3*}\}$, $\{R_t^{4*}\}$, $\{R_t^{5*}\}$, which are supposed to generate returns for the two-, three-, four- and five-star funds respectively. We are interested in examining whether the statistical properties of these five returns processes are different.

To put it differently, we wish to evaluate the following simple investment strategy: if in each time period we create a portfolio consisting only of five-star funds, are the risk-adjusted returns on this portfolio higher than the corresponding ones on a portfolio consisting solely of, say, two-star funds? The idea is that if the better performance of five-star fund is really due to superior management skills then these should be reflected in the returns on a fund including only five-star funds.

Our study does not attempt to evaluate the performance of individual mutual funds. Rather it aims to assess “an evaluation procedure”, namely that of Morningstar. In other words, if we create a portfolio consisting only of those funds that in each time period have received a five-star (or four-, three-, two-, one-) score by Morningstar, and then evaluate the risk-adjusted performance of this portfolio by more traditional portfolio evaluation procedures, what would the result be? Will these traditional evaluation procedures detect any differences in the risk-adjusted performance of the one-, two-, three-, four- and five-star portfolios?

The remainder of the paper is organised as follows. Section 2 describes the data and

analyses the statistical properties of the returns on the five funds-of-funds we construct. Section 3 outlines the standard and conditional CAPM models we adopt to address the issue of whether the better performance of higher-rated funds is in fact attributable to superior management skills. Section 4 discusses the empirical results from both unconditional and conditional portfolio performance evaluation. Section 5 summarises the main findings and offers some concluding remarks.

2 Statistical Properties of the Five Funds-of-Funds Returns

The data used in our study were taken from Morningstar Direct, which provides historical monthly returns of selected mutual funds and their star-based ranking calculated by Morningstar. We focus on the subset represented by equity mutual funds, that is funds that invest at least 90% of their Non-cash Adjusted Total Assets in equity securities. To avoid dealing with currency risk exposure we only consider funds quoted in US Dollars. At present, the Morningstar Direct database contains 21322 equity funds in US Dollars. In order to perform our evaluation we need sufficiently long series, and therefore we have restricted our sample to funds that have been star-rated for at least 10 years. This reduces the sample to 1511 equity funds. For these funds, historical returns and their Morningstar ranking are available since 01/1998.

We begin our analysis of the returns of the five funds-of-funds, STAR1, STAR2, STAR3, STAR4 and STAR5 defined in the previous section by examining the univariate properties of $\{R_t^{1*}\}$, $\{R_t^{2*}\}$, $\{R_t^{3*}\}$, $\{R_t^{4*}\}$, and $\{R_t^{5*}\}$. Table 1A reports descriptive statistics for each of them, together with their first-order autocorrelation. Table 1B presents estimates of the

correlation matrix. The results can be summarised as follows:

(i) The sample mean of returns is an increasing function of the number of stars. The average monthly returns of STAR1 and STAR5 are 0.31% and 0.51% respectively.

(ii) The pattern is less clear for sample standard deviations: the smallest is exhibited by STAR3, followed by STAR4 and STAR2. The standard deviation of STAR5 is the second highest, after that of STAR1.

(iii) The distributions of all the five returns exhibit negative skewness, this being highest for STAR2 and lowest for STAR5.

(iv) All the five returns series are serially uncorrelated, i.e. their degree of persistence is zero.

(v) The estimated correlation coefficients are very large. They range from 0.87 (the coefficient between STAR1 and STAR5) and 0.98 (the coefficient between STAR1 and STAR2, STAR2 and STAR3 and also STAR3 and STAR4). With correlation coefficients so close to unity, it is rather unlikely that any differences in the risk-adjusted performance of these funds-of-funds will be detected whatever the definition of “risk”.

The greatest difference in mean returns is between STAR5 and STAR1, for which the smallest correlation coefficient (0.87) is also obtained. To examine whether the mean return of STAR5 is statistically different from the mean return of the other four portfolios, we generate the return-differential series, $\Delta R_t^{5i*} = R_t^{5*} - R_t^{i*}$, $i = 1, 2, 3, 4$, and test, by means of a t-test, whether the means, μ_i , of ΔR_t^{5i*} are different from zero. This can be done by running a regression of ΔR_t^{5i*} on a constant term, c_i , that is $\Delta R_t^{5i*} = c + \nu_{it}$ and testing the significance of the coefficient c_i , $i = 1, 2, 3, 4$. The results are reported in Table 1C, together with a series of misspecification tests, in order to provide some information on the time series properties of the four return-differential series, and establish whether the conditions are met for the employed t-test to have good properties. Specifically, in addition to the

usual tests of serial correlation in ν_{it} , we also report the results from testing for the presence of non-linear temporal dependence in ν_{it} , i.e. the first-order autocorrelation of the squared residuals ν_{it}^2 along with the Ljung-Box $Q(l)$ -statistic for testing the hypothesis that the first l autocorrelations are equal to zero. Moreover, we report the well-known BDS test proposed by Brock, Dechert, Scheinkman and LeBaron (1996), which is designed to test the stronger assumption that the noise series, ν_{it} , is independent and identically distributed (i.i.d.). The results can be summarised as follows:

(i) The null hypothesis that the mean of return-differentials is equal to zero is not rejected for any of the four series. Nevertheless, it might be worth noting that the point estimates of the mean as well as the value of the t-statistic increase with the “star-difference” (that is, they are bigger for ΔR_t^{51*} followed by ΔR_t^{52*} , ΔR_t^{53*} , ΔR_t^{54*}).

(ii) The noise series ν_{it} , and hence the series ΔR_t^{5i*} , are not serially correlated but they exhibit strong second-order temporal dependence. In particular, the BDS test strongly rejects the null hypothesis that the series ΔR_t^{5i*} are i.i.d.

On the basis of the above results, we examine whether the inability to reject the null hypothesis of zero return-differentials by the t-tests might be due to the presence of non-linear dependence in ν_{it} , which has not been taken into account. Specifically, we re-estimate the models $\Delta R_t^{5i*} = c + \nu_{it}$, assuming that the errors ν_{it} follow GARCH(1,1) processes, that is $\nu_{it} = h_{it}\varepsilon_{it}$, $h_{it}^2 = d_i + a_i h_{it-1}^2 + b_i \varepsilon_{it-1}^2$. The results, reported in Table 1C(ii), suggest the following:

(iii) When second-order dependence is taken into account, the statistical inference on the existence of differentials in the average returns among the star-rated funds changes drastically. The null hypothesis that the mean return-differential is zero is rejected for ΔR_t^{51*} , ΔR_t^{52*} , and ΔR_t^{53*} at the 5% level, and even for ΔR_t^{54*} at the 10% level. This means that the star-rating system of Morningstar does produce a classification of funds which exhibit

some significant differences in terms of their average monthly returns.

(iv) The results reported above seem to be trustworthy since the hypothesis that the standardised noise series, $\varepsilon_{it} = \nu_{it}/h_{it}$ are i.i.d. is not rejected by the BDS test. Moreover, additional tests (not reported) for the presence of structural breaks within the sample seem to support the hypothesis that the standardised error process $\{\varepsilon_{it}\}$ is identically distributed.

3 Risk-Adjusted Returns: The Models

In this section we investigate whether the higher average returns of the five-star funds are the reward for the additional risk that the managers of these funds bear relative to the risk incurred by the managers of, say, three-star funds or reflect instead superior management skills of the managers of the five-star funds.

For this purpose, we consider an asset pricing model that describes the relationship between the expected return and risk of the various portfolios under consideration. Specifically, we adopt the conditional CAPM model of Ferson and Schadt (1996) and Shanken (1990) in which the level of the systematic portfolio risk is a function of the observed variables (see also, Lettau and Ludvigson 2001). This in turn implies that the relationship between the excess returns of the portfolio j and the excess returns of the market factor is given by the following relationships:

$$r_{j,t+1} = b_j(Z_t)r_{m,t+1} + \varepsilon_{j,t+1} \quad (1)$$

$$E(\varepsilon_{j,t+1} | Z_t) = 0 \quad (2)$$

$$E(\varepsilon_{j,t+1}r_{m,t+1} | Z_t) = 0 \quad (3)$$

where $r_{jt} = R_t^{j*} - R_{ft}$, $i = 1, 2, \dots, 5$, R_{ft} is the return of a one-month Treasury bill, $Z_t =$

$[Z_{1,t}, Z_{2,t}, \dots, Z_{n,t}]$ is an n -vector of state variables observable by the managers at time t , and $r_{mt} = R_t^m - R_{ft}$ with R_t^m stands for the returns of the market factor. This specification implies that the systematic risk of the portfolio j , as measured by $b_j(Z_t)$, changes with time. The time-varying nature of beta is due to the fact that the portfolio manager receives at time t an “information signal”, contained in the state variables Z_t , on the basis of which he changes the beta of his portfolio. If the signal is “correct” and the manager succeeds in “receiving” it, then the changes in the beta of the portfolio at time t will be consistent with the realized returns $r_{m,t+1}$ at time $t+1$. To put it differently, if $r_{m,t+1} > 0$ then the correct interpretation of the signal implies that the manager will shift the portfolio towards including stocks with high betas. The preceding discussion implies that the ability of the fund manager to “time” the market depends on the extent to which he/she can translate the information content of Z_t into predictions on the future behavior of $r_{m,t+1}$. This does not necessarily mean that “everybody” in the market can “read” the information contained in Z_t . In other words, although the variables Z_t are indeed publicly available, the information content of Z_t might be available only to a “skilful” fund manager.

The next question concerns the specification of the function $b_j(Z_t)$. Since the true functional form is unknown, we shall approximate it by using a first or second-order Taylor series expansion. We begin by applying a first order approximation of β_j , in which case equation (1) becomes

$$r_{j,t+1} = \beta_{j,0} + \sum_{i=1}^n \beta_{j,i} Z_{i,t} r_{m,t+1} + \varepsilon_{j,t+1} \quad (4)$$

It is quite natural to assume that the dependent variable, $r_{j,t+1}$, in (4) and $r_{m,t+1}$ are $I(0)$ variables. However, the stationarity of $Z_{1,t}, Z_{2,t}, \dots, Z_{n,t}$ cannot be assumed a priori. In the absence of stationarity of those variables, we may face the problem that (4) is an “unbalanced” regression. As a result, before we proceed any further we must analyse in detail the alternative models (all based on (4)) that arise depending on the statistical properties of the

variables $Z_{1,t}, Z_{2,t}, \dots, Z_{n,t}$. Specifically, we need to distinguish three cases:

(i) All the variables $Z_{1,t}, Z_{2,t}, \dots, Z_{n,t}$ are $I(0)$. In this case, the new variables $Z_{i,t}r_{m,t+1}$ will also be $I(0)$ (given that the market returns variable $r_{m,t+1}$ is quite naturally $I(0)$), and equation (4) is legitimate since $r_{j,t+1}$ is also quite naturally $I(0)$.

(ii) Some (or all) of the variables $Z_{1,t}, Z_{2,t}, \dots, Z_{n_0,t}$, $n_0 \leq n$ are $I(1)$ and not cointegrated. In this case, the product variables $Z_{i,t}r_{m,t+1}$, $i = 1, 2, \dots, n_0$ will have an asymptotically unbounded unconditional variance and will not be $I(0)$. In such a case, we have the problem of an “unbalanced regression” since the dependent variable, $r_{j,t+1}$ is $I(0)$.

(iii) Some (or all) of the variables $Z_{1,t}, Z_{2,t}, \dots, Z_{n_0,t}$, $n_0 \leq n$ are $I(1)$ and cointegrated. In this case, we proceed as follows: First, we rewrite (4) as

$$r_{j,t+1} = \beta_{j,0} + r_{m,t+1} \sum_{i=1}^n \beta_{j,i} Z_{i,t} + u_{j,t+1}. \quad (5)$$

Equation (5) shows that under the assumption that the returns processes $\{r_{j,t}\}_{t \geq 1}$ and $\{r_{m,t}\}_{t \geq 1}$ are $I(0)$, the only case where some of the $Z_{i,t}$ can be $I(1)$ is when the corresponding coefficients are such that only multiples of the cointegration relations between those $Z_{i,t}$ s that are cointegrated are left on the right-hand side of (5).

To see the difference between cases (ii) and (iii) more clearly, let us examine the following example:

$$r_{j,t+1} = \beta_{j,0} + (\beta_{j,1}Z_{1,t} + \beta_{j,2}Z_{2,t})r_{m,t+1} + u_{j,t+1} \quad (6)$$

If $Z_{1,t}$ and $Z_{2,t}$ are not cointegrated and $\beta_{j,1}\beta_{j,2} \neq 0$, the unconditional variance of the right-hand side will grow to infinity as $t \rightarrow \infty$, violating our initial assumption that the unconditional variance of $r_{j,t+1}$ is bounded. Therefore the estimated values of $\beta_{j,1}$ and $\beta_{j,2}$ will be very close to 0 when the sample is large. If, on the other hand, $Z_{1,t}$ and $Z_{2,t}$ are

cointegrated and satisfy

$$Z_{1,t} = a_0 + a_1 Z_{2,t} + w_t$$

where $\{w_t\}_{t \geq 1}$ is $I(0)$, the only way for the unconditional variance of $\beta_{j,1}Z_{1,t} + \beta_{j,2}Z_{2,t}$ to remain asymptotically bounded, with nonzero $\beta_{j,1}$ and $\beta_{j,2}$, is the case where

$$\beta_{j,1}Z_{1,t} + \beta_{j,2}Z_{2,t} = \lambda (Z_{1,t} - a_1 Z_{2,t}) = \lambda (a_0 + w_t), \lambda \in \mathbb{R}.$$

The last equation implies that

$$\beta_{j,1} = \lambda$$

and

$$\beta_{j,2} = -a_1 \beta_{j,1}.$$

If we identify the cointegrating relationship between $Z_{1,t}$ and $Z_{2,t}$, then we can rewrite (6) as

$$r_{j,t+1} = \beta_{j,0} + \beta_{j,1} q_{1,t} r_{m,t+1} + u_{j,t+1} \quad (7)$$

where $q_{1,t} = Z_{1,t} - a_1 Z_{2,t} = a_0 + w_t$.

The previous example demonstrates that it is necessary for the sum $\sum_{i=1}^n \beta_{j,i} Z_{i,t}$ in (5) to be $I(0)$ in order for (5) to be a legitimate regression. It also suggests how to treat the initial set of candidate state variables, in order to obtain in (5) a well-balanced regression. Specifically, the following steps must be taken:

First, we identify all those state variables (elements of Z_t) that are $I(1)$. Assume that the number of such $I(1)$ variables is n_0 . If $n_0 > 0$, without loss of generality, reordering the variables if necessary, we can make sure that, for $i \leq n_0$, $\{Z_{i,t}\}_{t \geq 1}$ are $I(1)$ and, for $n_0 < i \leq n$, $\{Z_{i,t}\}_{t \geq 1}$ are $I(0)$. Second, we identify any cointegrating relationships between the processes $\{Z_{i,t}\}_{t \geq 1}$, $1 \leq i \leq n_0$. Let $k < n_0$ be the rank of the cointegrating system.

This means that we can find a $(k \times n_0)$ matrix A of order k , such that

$$A \begin{bmatrix} Z_{1,t} \\ Z_{2,t} \\ \vdots \\ Z_{n_0,t} \end{bmatrix} = U_t, \quad (8)$$

where $\{U_t\}_{t \geq 0}$ is $I(0)$ with nontrivial coordinates $U_{i,t}$, $1 \leq i \leq k$. Again, without any loss of generality, we can reorder the variables $Z_{i,t}$, $1 \leq i \leq n_0$ in (5), so that the first k columns of A are linearly independent. Therefore, we can write $A = [A_1, A_2]$, where the $k \times k$ matrix A_1 is invertible. Then, left multiplication of (8) by A_1^{-1} yields

$$[I_k, A_1^{-1}A_2] Z_t = A_1^{-1}U_t,$$

which in turn gives

$$\begin{bmatrix} Z_{1,t} \\ Z_{2,t} \\ \vdots \\ Z_{k,t} \end{bmatrix} = -A_1^{-1}A_2 \begin{bmatrix} Z_{k+1,t} \\ Z_{k+2,t} \\ \vdots \\ Z_{n_0,t} \end{bmatrix} + A_1^{-1}U_t. \quad (9)$$

The last equation is the first part of Phillips's (1991) triangular system. The second part of this system is

$$\begin{bmatrix} \Delta Z_{k+1,t} \\ \Delta Z_{k+2,t} \\ \vdots \\ \Delta Z_{n_0,t} \end{bmatrix} = V_t, \quad (10)$$

where V_t is also $I(0)$. Equations (8) and (10) provide us with n_0 $I(0)$ processes that can

be considered as state variables in a new regression, replacing the $Z_{i,t}$, $1 \leq i \leq n_0$. Third, having defined the appropriate set, W_t , of $I(0)$ state variables,

$$W_t = (W_{i,t})_{1 \leq i \leq n} = [U_{1,t}, U_{2,t}, \dots, U_{k,t}, \Delta Z_{k+1,t}, \dots, \Delta Z_{n_0,t}, Z_{n_0+1,t}, \dots, Z_{n,t}]'$$

we can define the following regression

$$r_{j,t+1} = \beta_{j,0}^* + \sum_{i=1}^k \beta_{j,i}^* U_{i,t} r_{m,t+1} + \sum_{i=k+1}^{n_0} \beta_{j,i}^* \Delta Z_{i,t} r_{m,t+1} + \sum_{i=n_0+1}^n \beta_{j,i}^* Z_{i,t} r_{m,t+1} + u_{j,t+1}. \quad (11)$$

This regression can be rewritten in a more compact form in terms of W_t as

$$r_{j,t+1} = \beta_{j,0}^* + \sum_{i=1}^n \beta_{j,i}^* W_{i,t} r_{m,t+1} + u_{j,t+1}, \quad (12)$$

or simply

$$r_{j,t} = \beta_{j,0}^* + \sum_{i=1}^d \beta_{j,i}^* X_{i,t} + \zeta_{j,t}, \quad (13)$$

where $X_{i,t} = W_{i,t-1} r_{m,t}$ for $1 \leq i \leq n$. The regression defined in (12) can be considered a first order approximation of the general model given by

$$r_{j,t+1} = \beta_{j,0}^* + \beta_j^*(W_t) r_{m,t+1} + \varepsilon_{i,t+1}^*, \quad (14)$$

which involves only $I(0)$ processes.

The preceding discussion is based on approximating the unknown function $b_j(Z_t)$, or equivalently $\beta_j^*(W_t)$, by using a first-order Taylor series expansion. Alternatively, we can

approximate $\beta_j^*(W_t)$ by using a second-order Taylor expansion. In such a case we have

$$r_{j,t+1} = \beta_{j,0}^* + \sum_{i=1}^n \beta_{j,i}^* W_{i,t} r_{m,t+1} + \sum_{1 \leq i \leq k \leq n} \beta_{j,i,k}^* W_{i,t} W_{k,t} r_{m,t+1} + \zeta_{j,t+1}. \quad (15)$$

The last equation involves $n + \binom{n}{2} = \frac{n(n+1)}{2} \triangleq d$ explanatory variables of the form $W_{i,t} r_{m,t+1}$ or $W_{i,t} W_{k,t} r_{m,t+1}$, $1 \leq i \leq k \leq n$, which can be denoted as $X_{l,t+1}$, $1 \leq l \leq d$.

We can rewrite (15) as:

$$r_{j,t} = b_{j,0} + \sum_{i=1}^d b_{j,i} X_{i,t} + \zeta_{j,t} \quad (16)$$

where $b_{j,0} = \beta_{j,0}^*$, $b_{j,i} = \beta_{j,i}^*$ and $X_{i,t} = W_{i,t-1} r_{m,t}$ for $1 \leq i \leq n$, $b_{j,i} = \beta_{j,g,h}^*$ and $X_{i,t}$ is of the form $W_{g,t-1} W_{h,t-1} r_{m,t}$, when $n+1 \leq i \leq d$, for some $1 \leq g, h \leq n$.

The use of polynomial approximations of (14) may simplify its treatment, but if the order of approximation is underspecified, the estimation residuals are likely to be affected by the missing part of β_j^* . This is true even in the simple case where β_j^* is a second-order polynomial. For example, let the true model be:

$$r_{j,t+1} = \beta_{j,0}^* + \beta_{j,1}^* W_{1,t} r_{m,t+1} + \beta_{j,2}^* W_{2,t} r_{m,t+1} + \beta_{1,2,k}^* W_{1,t} W_{2,t} r_{m,t+1} + \zeta_{j,t+1}$$

and assume that a first-order approximation is used:

$$r_{j,t+1} = \beta_{j,0}^* + \beta_{j,1}^* W_{1,t} r_{m,t+1} + \beta_{j,2}^* W_{2,t} r_{m,t+1} + u_{j,t+1}.$$

One can see that any estimation of $\beta_{j,0}^*$, $\beta_{j,1}^*$, and $\beta_{j,2}^*$ using the first-order approximation will have to compensate for the missing expected value of $\beta_{1,2,k}^* W_{1,t} W_{2,t} r_{m,t+1}$. A non-zero expectation of $W_{1,t} W_{2,t} r_{m,t+1}$ will result in biased estimates of some or all of $\beta_{j,0}^*$, $\beta_{j,1}^*$, and

$\beta_{j,2}^*$. Another implication concerns the second-order properties of $u_{j,t+1}$: it is clear that the behavior of the estimated $u'_{j,t+1}$ s will be related to the missing term $\beta_{1,2,k}^* W_{1,t} W_{2,t} r_{m,t+1}$, and therefore these will exhibit conditional heteroscedasticity.

Finally, note that all the above models can be augmented by the market timing term, $\gamma_j r_{m,t+1}^2$, proposed by Traynor and Mazui (1966). A positive (negative) timing coefficient γ_j is interpreted as evidence suggesting superior (inferior) market timing abilities of the corresponding fund manager.

3.1 Empirical Results

3.1.1 Unconditional Portfolio Performance Evaluation

We begin our empirical analysis by considering the so-called unconditional evaluation of the star-rated funds-of-funds under consideration, which is based on the standard version of CAPM. The latter assumes that $b_j(Z_t) = b_j \forall t$. Under this hypothesis, the fund managers do not attempt to “time” the market, so they do not actively change the betas of their portfolios. Moreover this assumption also implies that the betas of the assets forming the portfolios do not change over time, or if they change the changes in the beta of one asset are exactly offset by those in the beta of another asset. Table 2A reports the results from the OLS estimation of equation

$$r_{j,t} = b_j r_{m,t} + \gamma_j r_{m,t}^2 + \varepsilon_{j,t+1} \quad (17)$$

Given that the error terms in all the five regressions exhibit conditional heteroscedasticity, we report the results based on heteroscedasticity-consistent (HC) standard errors in Table 2B, and those from explicitly assuming that the errors are GARCH(1,1) processes in Table 2C. The results can be summarised as follows:

(i) The OLS and HC estimates are quite similar, suggesting that the unconditional estimates of a for all the portfolios except STAR1 are positive and significantly different from zero. Interestingly, the highest a is achieved by STAR5. On the contrary, the estimates of the market timing coefficient γ appears to be insignificantly different from zero for all portfolios.

(ii) When conditional heteroscedasticity is taken into account, the results change significantly. In particular, the stock selection coefficient, a , now appears to be significantly positive only for STAR3, STAR4 and STAR5. Moreover, the GARCH-based estimates of a for all five portfolios appear to be smaller than the corresponding OLS-based estimates. For example, the OLS and GARCH estimates of a for STAR5 are 0.52 and 0.31, respectively. The GARCH estimates of γ (similarly to the corresponding OLS estimates) are insignificantly different from zero. The overall picture emerging from the GARCH estimates suggests that the five portfolios under consideration can be classified into two groups. The first one consists of STAR1 and STAR2, and is characterised by neither stock selection nor market timing abilities. The second one includes STAR3, STAR4 and STAR5, and exhibits some stock selection ability (which is almost identical among the three portfolios belonging to this group) but no market timing ability.

(iii) It is worth noting that, despite the constant beta assumption in this unconditional fund performance evaluation, the estimates of a are generally positive. This is in contrast with the early results of Jensen (1968) and the subsequent results of Elton et. al (1992) (among others) who report negative estimates of a , which may be caused by (unaccounted) time variation in the betas. This in turn implies that, despite the bias in the estimates of a , caused by the possible time variation of the betas, a positive stock selection ability can be inferred for STAR3, STAR4 and STAR5.

(iv) Related to (iii): Diagnostic tests (not reported) for parameter stability of equation

(17) estimated by OLS indicate the presence of significant time variation in the parameters of this model. However, this instability may be the result of omitted conditional heteroscedasticity (i.e. it may come from time variation of the standard error of the regression). When the GARCH effects are taken into account the observed instability is reduced, though not entirely eliminated. The possible time variation of beta, in particular, which may come from a response of the fund manager to changing economic conditions, is the focus of the next subsection.

3.1.2 Conditional Portfolio Performance Evaluation

The next issue is the selection of the variables in Z_t . Ferson and Schadt (1996) suggest including the one-month Treasury bill yield, z_{1t} , the term spread, z_{2t} , defined as the difference between the constant-maturity 10-year Treasury bond yield and the 3-month Treasury bill, and the quality spread, z_{3t} , in the corporate bond market defined as the Moody's BAA-rated corporate bond yield minus the AAA-rated corporate bond yield. In addition, we include variables that are usually considered important indicators by the financial community such as the weighted average of the foreign exchange value of the US dollar against a subset of the broad index currencies, z_{4t} , the Consumer Sentiment Index of the University of Michigan, z_{5t} , the price of oil, z_{6t} , and the Chicago Board Options Exchange volatility index (VIX), z_{7t} .

As explained in the previous section, the choice of the appropriate model for conditional portfolio evaluation depends on the statistical properties of the state variables z_{1t} , z_{2t} , ..., z_{7t} . The results from a variety of unit root tests, reported in Table 3A, unambiguously indicate that the first six series are I(1) while the last one is I(0). However, the tests on the cointegration properties of z_{1t} , z_{2t} , ..., z_{6t} , reported in Table 3B, lead to less clear-cut conclusions. In particular, when the lag-length, l , of the Vector Autoregressive model,

VAR(l), on which the two tests are based, is relatively large, both test statistics suggest a cointegration rank, k , of at least one, and occasionally two. On the contrary, when l is relatively small, both tests are unable to reject the null hypothesis of no cointegration. As a result, we run three alternative conditional regressions assuming $k = 0$, $k = 1$ and $k = 2$, with the results (assuming GARCH(1,1) errors and including the market timing term $\gamma_j r_{m,t+1}^2$) being reported in Table 4A, 4B and 4C, respectively. The results can be summarised as follows:

(i) The results are very robust across the three alternative cointegration rank assumptions. Indeed, the information criteria for $k = 0$ are very close to those for $k = 1$ or $k = 2$ for all five portfolios under consideration.

(ii) Despite the significance of (some of) the state variables, the results on stock selection (a) and market timing (γ) abilities from the conditional evaluation are similar to those obtained from the unconditional evaluation. In particular, for $k = 0$ the stock selection coefficient, a , appears to be significantly positive for STAR2, STAR3, STAR4 and STAR5, whereas for $k = 1$ and $k = 2$ it appears to be significantly positive for STAR3, STAR4 and STAR5. Concerning the latter portfolio, the highest estimate of (conditional) a is 0.36, obtained for $k = 2$, whereas the lowest is equal to 0.30, obtained for $k = 1$. Concerning market timing ability, no portfolio for any value of k produces a significantly positive estimate of γ_j . On the contrary the estimates of γ_j are negative and in many cases significantly so.

(iii) The results in (ii) suggest that the positive excess returns produced by STAR3, STAR4 and STAR5 should be thought of as the result of superior stock selection rather than market timing abilities.

4 Conclusions

This paper adds to the rather limited number of studies which to date have attempted to evaluate the Morningstar star-based system for ranking mutual funds (see, e.g., Blake and Morey, 2000). Its aim is to provide evidence on whether portfolios assembled using higher-rated mutual funds consistently outperform those made up of funds with lower Morningstar star-ratings. In particular, we are interested in examining whether a higher rating reflects superior management skills of the managers of those funds, and therefore a simple investment strategy could be adopted which would systematically result in higher average returns if the Morningstar ranking system is indeed informative about fund performance, this strategy consisting of always selecting the highest-rated funds when creating a portfolio. For this purpose, first we examine the statistical properties of the five funds-of-funds return series, i.e. the returns on the portfolios including respectively five-, four-, three-, two- and one-star funds only (STAR5 to STAR1). We show that, provided second-order dependence is taken into account, statistically significant return differentials can indeed be found, the higher Morningstar rating being associated with higher returns. In order to establish whether this is in fact due to superior management skills, we estimate appropriate asset pricing models for risk-adjusted returns. Specifically, we consider both a standard version of the CAPM model for unconditional portfolio performance evaluation, and a conditional CAPM (see Ferson and Schadt, 1996, and Shanken, 1990) in which portfolio risk is a function of observed variables in order to carry out a conditional evaluation as well. The results based on the former specification (when allowing for conditional heteroscedasticity) indicate that only the three highest-rated categories of funds are characterised by some stock selection ability, whilst none of the five categories exhibit market timing ability. Similar results are obtained for the conditional portfolio evaluation, the evidence suggesting that the better performance of the STAR3, STAR4 and STAR5 categories reflects superior stock selection rather than market

timing abilities. Overall, the implication for the Morningstar ranking system is that this is most effective in identifying the worst-performing funds (those to which one or two stars are assigned) rather than the best-performing ones: it can be used as a guide to avoid one- and two-star rated funds, but it is not really able to discriminate between three-, four- and five-star funds (although this does not rule out that mutual fund investors are more sensitive to *changes* in the ratings compared with the ratings themselves, as highlighted by Del Guercio and Tkac, 2008).

References

- [1] Blake, C.R and M.R. Morey (2000), "Morningstar Ratings and Mutual Fund Performance", *Journal of Financial and Quantitative Analysis*, 35, 3, 451-483.
- [2] Blume, M.(1998), "An anatomy of Morningstar ratings", *Financial Analysts journal*, March/April, 19-27.
- [3] Brown, S.J. and W. N. Goetzmann (1995), "Performance persistence", *Journal of Finance*, 50, 679-698.
- [4] Damato, K. (1996), "Morningstar edges toward one-year ratings", *Wall Street Journal*, April 5th, 1996.
- [5] Del Guercio, D. and P.A. Tkac (2008), "The effect of Morningstar ratings on mutual fund flow", *Journal of Financial and Quantitative Analysis*, 43, 4, 907-936.
- [6] Elton, E., Gruber, M., Das, S. and M. Hlavka (1992), "Efficiency with costly information: a reinterpretation of evidence for managed portfolio", *Review of Financial studies*, 6, 1-22.
- [7] Ferson, W.E. and R.W. Schadt (1996), "Measuring fund strategy and performance in changing economic conditions", *Journal of Finance*, 51, 2, 425-461.
- [8] Goetzmann, W.N. and R.G. Ibbotson (1994), "Do winners repeat?", *Journal of Portfolio Management*, Winter, 9-18.
- [9] Hendricks, D., Patel, J. and R. Zeckhauser (1993), "Hot hands in mutual funds: short-run persistence of relative performance, 1974-1988", *Journal of Finance*, 48, 93-130.

- [10] Jensen, M.C. (1968), “The performance of mutual funds in the period 1945-1964”, *Journal of Finance*, 23, 389-416.
- [11] Jensen, M.C. (1972), “Optimal utilization of market forecasts and the evaluation of investment performance”, in G.P. Szego and K. Shell, eds., *Mathematical Methods in Investment and Finance*, North-Holland, Amsterdam.
- [12] Khorana, A. and E. Nelling (1998), “The determinants and predictive ability of mutual fund ratings”, *The Journal of Investing*, November, 61-66.
- [13] Lettau, M. and S. Ludvigson (2001), “Resurrecting the (C)CAPM: a cross-sectional test when risk premia are time varying”, *Journal of Political Economy*, 109, 1238-1287.
- [14] Merton, R.C. and R.D. Henriksson (1981), “On market timing and investment performance II: statistical procedures for evaluating forecasting skills”, *Journal of Business*, 54, 513-534.
- [15] Phillips, P.C.B. (1991), “Optimal inference in cointegrated systems”, *Econometrica*, 59, 283-306.
- [16] Shanken, J. (1990), “Intertemporal asset pricing: an empirical investigation”, *Journal of Econometrics*, 45, 99-120.
- [17] Sharpe, W.F. (1998), “Morningstar performance measures”, *Financial Analysts Journal*, July/August, 21-33.
- [18] Treynor, J. and K. Mazui (1966), “Can mutual funds outguess the market?”, *Harvard Business Review*, 44, 131-136.

TABLES

Table 1
Time Series Properties of Monthly Returns of Star-Rated Funds-of-Funds

A. Univariate Characteristics									
Fund-of-Funds	mean	s.d.	skewn.	kurt.	J-B*	min.	max.	$\hat{\rho}_1$	Q(12)*
STAR1	0.31	5.35	-0.90	5.24	0.00	-23.26	12.60	0.053	0.93
STAR2	0.39	4.70	-0.97	4.92	0.00	-19.90	8.96	0.091	0.97
STAR3	0.42	4.32	-0.95	4.47	0.00	-17.01	7.42	0.095	0.99
STAR4	0.45	4.35	-0.87	4.06	0.00	-16.15	8.71	0.094	0.95
STAR5	0.51	4.87	-0.70	4.08	0.00	-16.58	12.98	0.082	0.78

*:p-val

B. Correlation Matrix						
	STAR1	STAR2	STAR3	STAR4	STAR5	
STAR1	1	0.98	0.95	0.93	0.87	
STAR2	0.98	1	0.98	0.97	0.90	
STAR3	0.95	0.98	1	0.98	0.92	
STAR4	0.93	0.97	0.98	1	0.96	
STAR5	0.87	0.90	0.92	0.96	1	

C. Testing for Zero Mean in Return-Differentials

(i) OLS: $\Delta R_t^{5i*} = c_i + \nu_{it}$.

$$H_0 : c_i = 0$$

Return-Differential	\hat{c}_i	s.e. (\hat{c}_i)	t-stat.	Serial Correlation		Non-Linear Dependence		
				$\hat{\rho}_1(\nu_{it})$	Q(12)*	$\hat{\rho}_1(\nu_{it}^2)$	Q(12)*	BDS*
STAR5-1	0.19	0.23	0.82	0.051	0.38	0.47	0.00	0.00
STAR5-2	0.11	0.18	0.61	0.069	0.09	0.55	0.00	0.00
STAR5-3	0.09	0.16	0.53	0.029	0.08	0.45	0.00	0.00
STAR5-4	0.05	0.11	0.42	-0.009	0.11	0.52	0.00	0.00

*:p-val

(ii) GARCH(1,1): $\Delta R_t^{5i*} = c_i + \nu_{it}$, $\nu_{it} = h_{it}\varepsilon_{it}$, $h_{it}^2 = d_i + a_i h_{it-1}^2 + b_i \varepsilon_{it-1}^2$

$$H_0 : c_i = 0$$

Return-Differential	\hat{c}_i	s.e. (\hat{c}_i)	t-stat.	i.i.d. for standard. residuals	
				BDS	
STAR5-1	0.27	0.09	2.91		0.43
STAR5-2	0.14	0.06	2.13		0.70
STAR5-3	0.11	0.05	2.08		0.12
STAR5-4	0.05	0.03	1.80		0.52

Table 2
Unconditional Evaluation of Star-Rated Funds-of-Funds

A. OLS							
Fund-of-Funds	a	$t(a)$	b	$t(b)$	γ	$t(\gamma)$	\overline{R}^2
STAR1	0.42	1.63	1.05	18.20	-0.01	-1.84	0.79
STAR2	0.40	2.28	0.98	25.07	-0.01	-1.78	0.87
STAR3	0.37	3.20	0.95	36.47	-0.01	-1.91	0.93
STAR4	0.39	3.31	0.96	35.85	-0.00	-1.58	0.93
STAR5	0.52	2.47	0.99	20.86	-0.01	-1.56	0.83
	AIC	SIC	Serial Cor. Q(12)	Non-Lin. Depend. Q(12)	BDS		
STAR1	4.653	4.720	0.59	0.00	0.00		
STAR2	3.884	3.951	0.12	0.00	0.00		
STAR3	3.049	3.116	0.15	0.02	0.04		
STAR4	3.107	3.173	0.11	0.00	0.01		
STAR5	4.268	4.334	0.13	0.00	0.00		

B. Heteroscedasticity-Consistent s.e.'s							
Fund-of-Funds	a	$t(a)$	b	$t(b)$	γ	$t(\gamma)$	
STAR1	0.42	1.52	1.05	18.94	-0.01	-1.32	
STAR2	0.40	2.10	0.98	25.52	-0.01	-1.13	
STAR3	0.37	3.02	0.95	37.09	-0.01	-1.36	
STAR4	0.39	3.19	0.96	35.03	-0.00	-1.31	
STAR5	0.52	2.55	0.99	18.14	-0.01	-1.47	

C. GARCH(1,1)								
Fund-of-Funds	a	$t(a)$	b	$t(b)$	γ	$t(\gamma)$	AIC	SIC
STAR1	-0.05	-0.36	1.05	24.39	0.00	0.07	4185	4.319
STAR2	0.15	1.22	0.99	32.82	0.00	0.25	3.547	3.681
STAR3	0.29	2.47	0.95	39.22	-0.00	-0.62	3.044	3.177
STAR4	0.31	2.88	0.96	38.21	-0.00	-0.86	3.068	3.202
STAR5	0.31	2.66	0.99	38.76	-0.00	-0.81	3.625	3.759

Table 3
Statistical Properties of the State Variables

A. Unit Root Tests					
Variable	ADF	PP	DF-GLS	MZ_a	MZ_t
z_{1t}	-1.05	-1.25	-1.22	-4.71	-1.42
z_{2t}	-1.54	-2.06	-1.16	-2.66	-1.14
z_{3t}	-0.91	-0.65	-0.63	-3.59	-0.79
z_{4t}	-0.89	-0.60	-0.88	-2.32	-0.87
z_{5t}	-1.97	-1.75	-1.96	-6.35	-1.76
z_{6t}	0.76	1.40	1.47	2.89	1.50
z_{7t}	-3.67	-3.38	-2.21	-12.15	-2.12
5% c.v.'s	-2.87	-2.87	-1.94	-8.10	-1.98

B. Testing for Cointegration Among $z_{1t}, z_{2t}, \dots, z_{6t}$						
	$l = 1$		$l = 6$		5% c.v.'s	
	$\lambda - \max$	TR	$\lambda - \max$	TR	$\lambda - \max$	TR
$k = 0$	38.14	89.49	43.11	116.27	40.07	95.75
$k = 1$	26.81	51.35	32.49	73.15	33.87	69.81
$k = 2$	10.87	24.53	17.23	40.66	27.58	47.85
$k = 3$	9.21	13.65	14.14	23.42	21.13	29.79
$k = 4$	3.35	4.44	5.37	9.28	14.26	15.49
$k = 5$	1.08	1.08	3.91	3.91	3.84	3.84

Notes:

1) z_{1t} =one-month treasury bill yield, z_{2t} =constant-maturity 10-year Treasury bond yield minus 3-month Treasury bill, z_{3t} = Moody's BAA-rated corporate bond yield minus AAA-rated corporate bond yield, z_{4t} =the exchange rate of the dollar, z_{5t} =the consumer confidence index, z_{6t} =the price of oil, z_{7t} =the CBOE's VIX volatility index.

2) ADF and PP refer to the standard Augmented Dickey-Fuller (1979) and Phillips-Perron (1988) tests respectively for the null hypothesis of a unit root. The lag-length and the bandwidth parameter in ADF and PP respectively, were selected by the Schwarz information criterion and the Newey and West (1994) procedure respectively. DF-GLS refers to the unit root test proposed by Elliot, Rothenberg and Stock (1996) based on GLS detrended series. MZ_a and MZ_t are two of the four tests proposed by Ng and Perron (2001).

3) l denotes the lag length of the unrestricted Vector Autoregressive Model (VAR) based on which the Johansen (1991) maximum eigenvalue ($\lambda - \max$) and trace (TR) statistics were calculated.

Table 4
 Conditional Evaluation of Star-Rated Funds-of-Funds
 (GARCH(1,1) Error Specification)

**A: Cointegration Rank among $z_{1t}, z_{2t}, \dots, z_{6t}$ is Equal to Zero.
 $\Delta z_{1t}, \Delta z_{2t}, \dots, \Delta z_{6t}$ and z_{7t} are Employed as State Variables.**

Fund-of-Funds	a	$t(a)$	γ	$t(\gamma)$	Significant State	AIC	SIC
					Variables		
STAR1	-0.01	-0.02	-0.008	-1.12	$\Delta z_{4t}, \Delta z_{6t}$	4.175	4.353
STAR2	0.26	2.32	-0.008	-1.92	$\Delta z_{1t}, \Delta z_{2t}, \Delta z_{4t}, \Delta z_{6t}$	3.505	3.728
STAR3	0.33	3.42	-0.007	-2.71	$\Delta z_{4t}, \Delta z_{6t}$	3.004	3.183
STAR4	0.26	2.58	-0.005	-1.98	$\Delta z_{6t}, z_{7t}$	3.016	3.194
STAR5	0.34	2.73	-0.010	-3.36	$\Delta z_{3t}, \Delta z_{6t}, z_{7t}$	3.505	3.706

**B: Cointegration Rank among $z_{1t}, z_{2t}, \dots, z_{6t}$ is Equal to One.
 Cointegration Relation, u_{1t} , Together with z_{7t} are Employed as State Variables.**

Fund-of-Funds	a	$t(a)$	γ	$t(\gamma)$	AIC	SIC
STAR1	-0.07	-0.50	-0.001	-0.23	4.179	4.357
STAR2	0.14	1.24	0.001	0.04	3.545	3.723
STAR3	0.23	2.23	-0.003	-0.85	3.043	3.222
STAR4	0.28	2.66	-0.004	-1.17	3.024	3.202
STAR5	0.30	2.47	-0.005	-1.57	3.531	3.709

**C: Cointegration Rank among $z_{1t}, z_{2t}, \dots, z_{6t}$ is Equal to Two.
 Cointegration Relations, u_{1t} , and u_{2t} Together with z_{7t} are Employed as State Variables.**

Fund-of-Funds	a	$t(a)$	γ	$t(\gamma)$	AIC	SIC
STAR1	-0.09	-0.61	-0.002	-0.326	4.193	4.394
STAR2	0.14	1.16	0.000	0.051	3.562	3.762
STAR3	0.23	2.07	-0.003	-0.961	3.056	3.256
STAR4	0.29	2.68	-0.003	-0.949	3.033	3.233
STAR5	0.36	3.10	-0.010	-3.01	3.544	3.745