WORKING PAPER n.03.10

December 2003

Relative Benchmark Rating and Persistence Analysis: Evidence from Italian Equity Funds

Roberto Casarin[1]
Marco Lazzarin[2]
Loriana Pelizzon[3]
Domenico Sartore[1]

[1] University of Venice and GRETA Associati
[2] GRETA Associati
[3] University of Padua

Relative Benchmark Rating and Persistence Analysis: Evidence from Italian Equity Funds

Roberto Casarin [a b]
Marco Lazzarin [b]
Loriana Pelizzon [c]
Domenico Sartore [a b *]

December 2003

a Ca' Foscari University, Venice.
b GRETA, Venice.
c London Business School and University of Padua.
* Contact Author: Domenico Sartore, GRETA, San Marco 3870, I – 30124 Venice (Italy) Phone: +39 041 5238178
Fax: +39 041 5286166, E-mail: sartore@unive.it

Relative Benchmark Rating and Persistence Analysis: Evidence from Italian Equity Funds

Abstract

The recent introduction into the Italian mutual fund market of Morningstar performance rating performed by private institutions gives rise to the question of what is the relation between this relative benchmark measure and the other traditional performance measures. This paper provides a comprehensive analysis of the relative benchmark performance measure (*Morningstar rating*) applied to Italian equity funds. We find that this performance measure is highly correlated with the classical performance measures (*Sharpe ratio*, *Sortino ratio* and *Treynor ratio*) and lowly correlated with the customized benchmark measure (*Information ratio*). Furthermore, performing a persistence analysis, using non parametric methods called *Cross-Product Ratio* and *Chi-Squared test*, we observe that only *Morningstar rating* measure generates a strong degree of persistence. Our results deviate from most European studies that argue that Italian mutual funds display weak persistence.

Introduction

Mutual funds are now the preferred way for individual investors and many institutions to participate in the capital markets, and their popularity has increased demand for evaluations of fund performance. Some private institutions recently introduced a system classification of mutual funds based on performance measure like the *Morningstar rating* measure, and this gives rise to some questions on the relation between the relative benchmark *Morningstar measure* and the traditional performance measures. Previous empirical work on the Italian mutual fund industry (Cesari and Panetta [1998], Grande and Panetta [2002], Bams and Otten [2002], Beltratti and Miraglia [2001] and Casarin, Pelizzon and Piva [2002]) does not take into account this relative "*peer group benchmark*" performance measure, while in Carluccio [1999] the importance of an analysis based also on relative peer group benchmark measure in the Italian market is suggested and only theoretically discussed.

Moreover, several studies analyse the ability of the funds manager to generate positive performance persistently over periods. In the literature, this kind of study is referred to as performance persistence analysis. Grinblatt and Titman [1992], Goetzmann and Ibbotson [1994] and Hendriks, Patel and Zeckhauser [1993] analyse the U.S. funds market and present strong evidence in favour of a "*hot hand*" phenomenon; that is, mutual funds that achieve above average returns continue to enjoy superior performance. It will be interesting to see whether we can confirm the findings of return persistence also in the Italian market. Whereas the previous literature on performance persistence has concentrated mainly on the U.S. funds, while little evidence is available for Italy (see Casarin, Pelizzon and Piva [2000]; Beltratti and Miraglia [2001]; Otten and Bams [2002]; Grande and Panetta [2002]). Furthermore a persistence analysis of the benchmark-relative measures on the Italian market is still lacking.

Therefore, in this paper we propose a comparative analysis of the traditional and of the relative benchmark performance measures, and produce a persistence analysis, following the approach of Malkiel [1995], Brown and Goetzmann [1995] and Agarwal and Naik [2000] based on contingency tables.

We consider a sample of Italian equity mutual funds over the period 1997 through 2000. In particular we consider three categories of performance indicators. We consider measures based on *absolute benchmarks*, measures based on *relative benchmarks* and finally measures based on *customized benchmarks* (see Kritzman [1986][1]).

The first category includes the traditional risk-adjusted measures: Sharpe Ratio, Sortino Ratio and Treynor Ratio. The second category considers a new measure, *Morningstar Risk-adjusted Rating*. This measure, proposed by the rating company Morningstar, is well known among American investors, but Morningstar-like classifications are of recent introduction in the Italian mutual fund market. The last category is very interesting, because from 2000 Italy is the first country in Europe where, by law, the *customized benchmark* must be reported in each mutual fund's prospectus and in its application, to help mutual fund investors by offering market "standards" and in order to be able to evaluate the risk and the return of their own investments. In this case we estimate the information ratio of every fund of the sample. To study the robustness of our analysis we use returns based style analysis to evaluate the homogeneity of our sample.

We compare the fund's ranking produced by every risk-adjusted performance considered above for every period, using *Spearman's rank order correlation* coefficient. Finally, using non-parametric methods, called *Cross-product ratio* and *Chi square test*, we examine the "*hot hand phenomenon*" in the performance of mutual fund managers.

The organization of the paper is as follows: in section one we describe the sample of valuation; in section two we describe and estimate the risk-adjusted performance measures with a particular attention to the *Morningstar rating* measure; in section three we describe all the methodologies used to perform the persistence tests and present the results. Section four concludes.

[1] See also Aldrich [1987]; Mossavar-Rahmani [1987]; Rennie, Cowhey [1990].

1. The data

We collected a data set of weekly[2] returns (from Datastream) for all the 76 funds classified as "Azionari Italia" by Assogestioni[3] from January 1997 to December 2000, a period which covers market ups and downs over stable and turbulent periods. In order to perform the analysis we required at least 48 weeks of data. Thus, the initial sample has been reduced from 76 to 71 elements. We took into consideration any changes in the names of funds. The sample is determined combining these criteria with another restriction: mutual funds must be active at least at December 1999, so in our analysis we consider only funds that have at least a year of sample observations. The data set is obtained on the basis of total return indexes that measure the total returns on the underlying funds, combining both capital performances and reinvested incomes from dividends.

The sample does not present *attrition rate*, defined as the percentage of dead funds in the total number of funds, but is affected, even in a very limited way, by *survivorship bias*[4]. In fact, even if the data set captures changes in the names of mutual funds, the selection procedure does not include funds that changed investment policy. However, it seems rational to say that these operations affect the Italian market only marginally: these are typical operations of a full market, with high levels of competition (Malkiel [1995]).

Furthermore, we take into consideration that if the sample is not homogeneous all the persistence tests could be biased. In particular investment policy changes during the sample period could produce survivorship bias. Moreover, as pointed out by Brown et al. [1999] the existence of different investment strategies, "style factors", across the mutual funds, can lead to reversals in the persistence

---

[2] We collect a dataset of weekly observations, and not monthly observations as in numerous previous studies, to have a better estimate of the performance measures. For example, Cesari and Panetta [1998], Bams and Otten [2002], Casarin, Pelizzon and Piva [2000] evaluate performance measures from a monthly database

[3] The Assogestioni (the Italian mutual funds association) classification is widely used in Italy: it currently includes 24 different categories, based on the prevailing asset classes of investment.

[4] Survivorship bias arises if investors' withdrawals push the poorly performing funds out of the market, so only the superior funds survive. Therefore, samples which exclude funds that perished because of their inferior performance are biased towards finding persistence. See, in particular, Brown, Goetzmann, Ibbotson and Ross [1992], Brown and Goetzmann [1995] and Hendricks, Patel and Zeckhauser [1997].

phenomenon because of the differences in the levels of systematic risk across managers.

For this reason we apply the constrained regression model proposed by Sharpe [1988], [1992] and find that, except for 7 funds, none of all the other funds change their investment policy. We conclude that our sample is rather homogeneous, coherently with the classification provided by Assogestioni. Moreover, we verified also that the exclusion of that funds from the sample does not change the main results on the persistence analysis.

2. Performance Measures and Analysis

In the following we briefly discuss the performance measures used in our work and explain how computations are done. We consider two categories of measures: the simple risk-adjusted measures, the risk-adjusted measures based on the relative benchmark and finally the risk-adjusted measures based on the customized benchmark. Concluding we show the results of the Italian funds performance.

2.1. Risk-adjusted measures based on absolute benchmark

As risk adjusted measures based on absolute benchmark we consider three different ratios: *Sharpe, Treynor* and *Sortino*.

The most commonly used measure of risk-adjusted performance is the *Sharpe ratio* (Sharpe [1966]), which measures the fund's excess return per unit of its risk. The Sharpe ratio is defined as:

$$Sharpe\ Ratio = \frac{fund's\ average\ excess\ return}{standard\ deviation\ of\ \ fund\ excess\ return} \tag{1}$$

The Sharpe ratio allows a direct comparison of the risk-adjusted performance of any two mutual funds, regardless of their volatilities and their correlations with a benchmark.

We calculate the numerator of the Sharpe ratio by averaging the excess return of the mutual fund with respect to the risk free rate. We treat the BOT interest rate as risk free. In particular we use the average interest rate of the closest auction to the reference week.

The *Treynor ratio* (Treynor [1966]) is a risk/return measure similar to the Sharpe Ratio. It measures the return of the fund in excess of the risk-free return, per unit of risk that the fund adds to a well-diversified portfolio. The Sharpe Ratio uses the standard deviation as a measure of risk, while the Treynor ratio uses the fund's beta β , the systematic risk measure.

$$Treynor\ Ratio = \frac{fund's\ average\ excess\ return}{\beta} \tag{2}$$

The quantity, beta, at the denominator is given by the ratio between the covariance of the fund's return and the market's return and the variance of the market's return. We use the return of the COMIT stock index as measure of the market's return.

Standard deviation is sometimes criticized as being an inadequate measure of risk because investors do not dislike variability *per se*. Rather they dislike losses, but are quite happy to receive unexpected gains. One way to meet this objection is to calculate a measure of downside variability, which takes account of losses but not of gains. The downside deviation considers only those returns that fall below a defined target rate, called the *Minimum Acceptable Return* (*MAR*), rather than the arithmetic mean. The *Sortino ratio* (see Sortino and Van De Meer [1991]) measures the return of the fund in excess of the return of the *MAR*, per unit of downside deviation.

$$Sortino\ Ratio = \frac{fund's\ excess\ return}{downside\ deviation} \tag{3}$$

where

$$downside\ deviation = \sqrt{VAR\ \left[\min\left(0, fund's\ excess\ return\right)\right]}$$

The only difference between the Sortino and the Sharpe ratio is the risk measure. In the Sortino ratio we take the standard deviation of the negative variations of the excess return. In our samples many funds exhibit non-normal return's distribution. The downside risk deviation accounts for asymmetry and kurtosis of the excess return distribution.

2.2. Risk adjusted measures based on relative benchmark*: Morningstar Risk-Adjusted Rating.*

Morningstar calculates its own measures of risk-adjusted performance that form the basis of its popular star rating, which is routinely published by the *New York Times*. A recent study (Damato [1996]) reported in both the *Boston Globe* and the *Wall Street Journal* points to the importance of the Morningstar star rating service. This study on American mutual fund market found that 97% of the money flowing into equity funds between January and August 1995 was invested in funds which were rated as 5-star or 4-star funds by Morningstar, while funds with less than 3 stars suffered a net outflow of funds during the same period (see Jaffe [1995] and Damato [1996]). Moreover, the heavy use of Morningstar ratings in mutual fund advertising suggests that mutual fund companies believe that investors care about Morningstar ratings. The relevance of the relative benchmark performance evaluation and in particular of the Morningstar type performance measures is well documented also in the academic literature. (see, for example Blume [1998], Sharpe [1998] and Blake and Morey [1999]).

To calculate its ratings, Morningstar first classifies funds into one of four categories: Domestic Equity, Foreign Equity, Municipal Bond and Taxable Bond. The risk-adjusted return is calculated in the following manner. First they calculate a load-adjusted return for the fund by adjusting the returns for management fees and other costs, and then by adjusting for front-end and deferred loads. Next, they calculate a “Morningstar Return” in which they take the load-adjusted excess return divided by the higher of two variables: the excess average return of the fund category or the average 90-day U.S. T-bill rate:

$$\text{Morningstar Return} = \frac{\text{Load Adjusted Return on the Fund} - \text{T Bill}}{\text{Higher of (Average Category Excess Return or T Bill)}} \qquad (4)$$

Morningstar divides through by one of these two variables to prevent distortions caused by having low or negative average excess returns in the denominator of equation.

A “Morningstar Risk” measure is then computed. This measure is calculated differently from traditional risk measures, such as beta and standard deviation, which both see greater-than and less-than-expected returns as added volatility. Morningstar believes that for most investors their greatest fear is losing money, which they define as under performing the risk-free rate of return an investor can earn from the 90 day Treasury Bill. Hence, their risk measure only focuses on downside risk. To calculate the Morningstar risk, they plot the monthly returns in relation to T-bill returns. They add up the amounts by which the fund trails the T-Bill return each month, and then divide that total by the time horizon’s total number of months. This number, the average monthly underperformance statistic, is then compared with those of other funds in the same broad investment category to assign the risk scores. The resultant Morningstar risk score expresses how risky the fund is relative to the average fund in its category:

$$MorningstarRisk = \frac{Fund's\ \ Average\ Underperformance}{Average\ Underperformance\ of\ \ its\ Category} \qquad (5)$$

To calculate a fund’s summary star-rating, the Morningstar Risk scores are then subtracted from the Morningstar Return scores. From previous measures comes the Morningstar star rating. In our study we apply the Morningstar return and risk measures to the Italian equity funds category, as defined in the Assogestioni mutual funds classification (see Assogestioni [2002]). The homogeneity of this category has been verified statistically through the return based style analysis.

2.3. Risk adjusted measures based on Customized Benchmark: *Information Ratio*

Managers take risk, and potentially add value, by deviating from the benchmark. They may hold fewer securities, and they may weigh them differently from their benchmark weights. They may buy and sell them at different times: in other words, they add value through security selection and market timing decisions.

Customized measures of risk are used to assess the historical magnitude of a portfolio's active bets (security selection, sector weighting, etc.) relative to a customized benchmark. Relative (or customized) measures of risk-adjusted returns are used to assess the portfolio manager's "skill" in making these bets, converting them into higher returns for the client. While the absolute measures described above are suitable for both active and passive portfolios, customized measures are suitable only for actively managed portfolios.

Italy is the first country in Europe where, by law (see also Assogestioni [2000a,b]), the customized benchmark must be contained in the mutual fund's prospectus and in its application. In this way the benchmark can be especially helpful to mutual fund investors by offering market "standards" to help them evaluate the risk and the return of their own investments. In this sense, our work is the first analysis based on customized benchmarks for the Italian market. We consider the customized benchmarks declared by each funds and collected in the Assogestioni's yearly report (see Assogestioni [2000a]). We reconstruct the weekly returns series of the declared customized benchmarks using data on the stock and bond indexes provided by Datastream.

The most widely-used measure of benchmark relative risk is *tracking error* (TE), which is the standard deviation of residual returns (i.e., of the difference between portfolio returns and benchmark returns, also called alpha). Generally, the higher the tracking error, the greater the relative bets the manager has taken (Lee [2000]).

The *information ratio* (IR) is computed by dividing a portfolio's active return relative to the benchmark by its tracking error:

$$Information\ Ratio = \frac{Alpha}{Tracking\ Error} \tag{6}$$

The information ratio measures the quality of the manager's information discounted by the residual risk in the betting process (Goodwin [1998]).

Alpha has been evaluated averaging the excess fund's return with respect the customized benchmark's return over period of four weeks, whereas tracking error is the standard deviation of that excess return.

2.4. Results of performance analysis

To perform a deep analysis of the performance measures we have broken the sample of 176 weeks in 44 subperiods of 4 weeks. For each subsample we evaluate the risk-adjusted performance measures, thus obtaining 71 series, one for each fund.

Due to the lack of data in each subperiod of 4 weeks, risk measures (that is: standard deviation, beta, downside deviation and Morningstar Risk measures) are calculated on the basis of an *exponentially moving average* of 27 historical observations[5] where the latest observations carry the highest weight in the volatility estimate. This approach has two important advantages over the equally weighted model. For example, if we consider the estimation of the standard deviation, σ. The first advantage of using an exponentially weighted moving average is that volatility reacts faster to shocks in the market because recent data carry more weight than data in the distant past. Second, following a shock (a large return), the volatility declines exponentially as the weight of the

[5] 27 observations mean that we used 44 samples (rolling windows) of 27 weeks where 23 weeks are overlapping.

shock observation falls. In contrast, the use of a simple moving average leads to relatively abrupt changes in the standard deviation once the shock falls out of the measurement sample, which, in most cases, can be several weeks after it occurs. For a given set of T returns (with T equal to 27), the formula used to compute exponentially weighted (standard deviation) volatility[6] is:

$$\sigma_\lambda = \sqrt{(1-\lambda)\sum_{t=1}^{T}\lambda^{T-t}(r_t - \bar{r})^2} \qquad (7)$$

Notice that the exponentially weighted moving average depends on the parameter λ ($0 < \lambda < 1$), which is often referred to as the *decay factor*. This parameter determines the relative weights that are applied to the observations. Following the methodology used in RiskMetrics[7] we determined, for the equity mutual fund market, that the optimal decay factor is $\lambda = 0.94$, that is the average of individual optimal decay factors of every funds.

The same weights have been used to determine the other risk measures (beta, downside deviation and Morningstar Risk).

As mentioned above, for each subsample we evaluate the risk-adjusted performance measures thus obtaining 71 series, one for each fund. We conduct a means and standard deviations analysis of that series (see Figure 1) in order to study the effects of the different measures on the mutual funds classification.

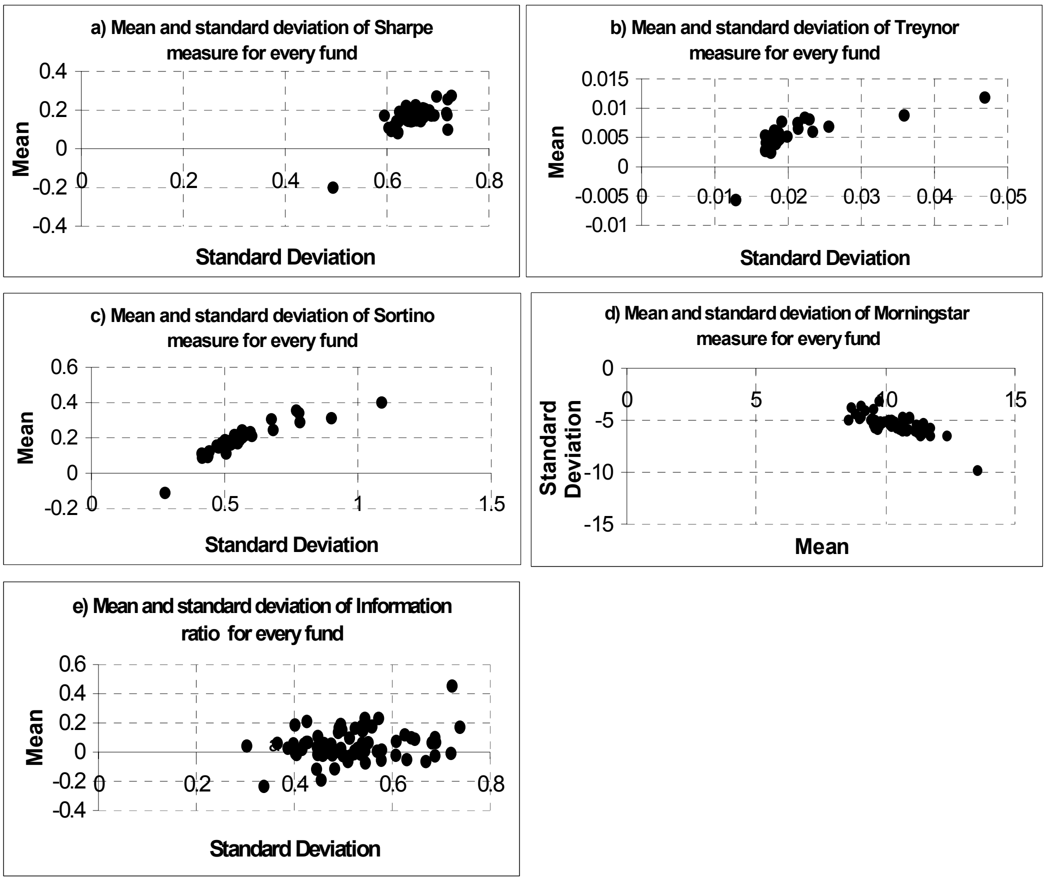
Figure 1 shows that most of the funds behave similarly when evaluated in terms of Sharpe, Treynor, Sortino and Morningstar Ratio measures. On the contrary, funds present different means and standard deviation when we evaluate them through the Information Ratio measure.

---

[6] The *exponentially weighted moving average* model is a particular case of GARCH (1,1) model that can be written as: $\sigma_t^2 = \gamma V + \alpha u_{t-1}^2 + \delta \sigma_{t-1}^2$ where u_t is defined as the continuously compounded return for the period *t* and σ_t^2 is the estimate of the variance rate for period *t.* The *EWMA* model is a particular case of *GARCH(1,1)* where $\gamma = 0$, $\alpha = 1 - \lambda$ and $\delta = \lambda$, see Hull [2000].

[7] See section "Statistics of Financial Market Returns" in J.P.Morgan, Reuter (1996).

Figure 1. – Performance measures mean and the standard deviation



This figure exhibits the mean and the variance of performance measures calculated by breaking the sample period of 176 weeks into 44 sub-periods of 4 weeks. Sharpe, Treynor, Sortino, Morningstar measures and Information ratio are shown respectively in Panel a), b), c), d) and e).

Referring to the performance evaluation, an important aspect is the correlation among different measures. In fact, it is relevant to verify whether various performance measures, with different characteristics, produce analogous rankings of funds. Table 1, which reports the average rank order correlation measured by Spearman's coefficient[8], seems to indicate a very high correlation

[8]The statistics take on values between +1 and –1, where +1 indicates they are identical and –1 indicates the rankings are reversed. Spearman's rank-correlation is computed using the following

between the Sharpe, Sortino, Treynor ratios and Morningstar Risk-adjusted Rating, whereas there exists only a weak relation between these measures and the Information ratio, which appear as a substantially different measure from the others.

From Table 1 we can observe that (traditional) risk-adjusted measures are highly correlated. These results are different with those obtained by Casarin et al. (2000). The reasons of this could be both (i) the different sample used and (ii) the frequency of our observation. Indeed we use a sample of only three years from 1997 till 2000 and Casarin et al. [2000] consider a sample of 11 years from 1988 till 1999. Moreover we use weekly data rather monthly data.

Table 1. – Spearman's rank order correlation coefficient.

	SHARPE	SORTINO	TREYNOR	MORNINGSTAR	INFORMATION
SHARPE	1.00	0.95	0.95	0.90	0.56
SORTINO		1.00	0.93	0.90	0.54
TREYNOR			1.00	0.90	0.55
MORNINGSTAR				1.00	0.57
INFORMATION					1.00

This table exhibits the average of the 44 sub-samples correlation coefficients computed between the rank orders induced by the different performance measures on the 71 funds. This measure of dependence is called Spearman's rank correlation.

Table 1 shows that the Morningstar Risk-adjusted rating is highly correlated with the traditional measures and has a low correlation with the ranking produced by another benchmark based measure: the information ratio. One potential explanation of these results is that, due to the Morningstar's definition of risk, the Morningstar risk-adjusted performance measure is similar to the Sortino ratio. On the other hand the information ratio measures the performance of the

---

formula: $c = 1 - \frac{6\sum_{i=1}^{n}[r(X_i) - r(Y_i)^2]}{n(n^2 - 1)}$, where $r(X_i)$ is the rank of the i[th] fund using one performance measure; $r(Y_i)$ is the rank of the i[th] fund using a different performance measure; *n* number of funds being ranked.

fund with respect to another stochastic variable, i.e. the benchmark, and it is for this reason intrinsically different from the other four measures.

3. Persistence analysis

In our analysis we follow the approach of Malkiel [1995], Brown and Goetzmann [1995] and Agarwal and Naik [2000] to determine the extent of persistence in the performance of mutual fund managers., We examine the persistence of performance measures, mentioned above, following the approach proposed by Brown and Goetzmann [1995], using non-parametric tests based on contingency tables.

The persistence test is structured as follow. We break the sample period of 176 weeks into 44 sub-samples of 4 weeks. For each sub-period and each fund we evaluate all the performance measures and classify each fund in two categories: winner and loser. In particular, at every date, we consider all the funds that were active at the end of previous interval and we construct a contingency table of winners and losers where a fund is a *winner* if the performance measure of that fund is greater than the median performance measure of all the funds in that period, otherwise it is a *loser*. Let the index t=1,…, 43 denote the contingency table number. Persistence in this context relates to the funds that are winners in two consecutive sub-periods denoted by $N_{t,WW}$, or losers in two consecutive periods, denoted by $N_{t,LL}$. Similarly, winners in the first sub-period and losers in the second period are denoted by $N_{t,WL}$, and $N_{t,LW}$ denotes the reverse.

Finally we aggregate the resulting 43 contingency tables and conduct the following test, as in Agarwal and Naik [2000] on the aggregated absolute frequencies denoted by N_{WW}, N_{LL}, N_{WL} and N_{LW}. We denote with N the sum of the absolute frequencies, i.e. N=N_{WW}+N_{LL}+N_{WL}+N_{LW}. We use both *Cross-product ratio* (CPR) and *Chi-square statistic* to detect persistence. CPR, defined as $\frac{N_{WW} \times N_{LL}}{N_{WL} \times N_{LW}}$, captures the ratio of the funds which show persistence in performance to the ones which do not. The null hypothesis in this setting represents lack of persistence for which the CPR equals one. In other words, when there is no persistence, one would expect each of the four categories denoted by N_{WW}, N_{WL}, N_{LW} and N_{LL} to have 25% of the total number of funds. We determine the statistical significance of the CPR by using the standard error of the natural logarithm of the CPR given by (see Christensen [1990]):

$$\sigma_{\ln(CPR)} = \sqrt{\frac{1}{N_{WW}} + \frac{1}{N_{WL}} + \frac{1}{N_{LW}} + \frac{1}{N_{LL}}} \tag{8}$$

In fact, it is possible to demonstrate that the statistic $Z = \frac{\ln(CPR)}{\sigma_{\ln(CPR)}}$, under the null hypothesis of absence of persistence, is asymptotically normally distributed. Thus if the Z-statistic is greater than its critical value, at the significance level of 5%, the null hypothesis is rejected in favour of the presence of persistence.

We also conduct a Chi-square test comparing the observed frequency distribution of N_{WW}, N_{WL}, N_{LW} and N_{LL} for each fund with the expected frequency distribution. In a recent paper, Carpenter and Lynch [1999] study the specification and power of various persistence tests. They find that the Chi-square test based on the number of winners and losers is well specified, powerful and more robust to the presence of survivorship bias compared to other test methodologies. We compute the Chi-square statistic as:

$$\chi^2 = \frac{(N_{WW} - D1)^2}{N} + \frac{(N_{WL} - D2)^2}{N} + \frac{(N_{LW} - D3)^2}{N} + \frac{(N_{LL} - D4)^2}{N} \tag{9}$$

where:

$$D1 = \frac{\tilde{N}_{W.}\tilde{N}_{.W}}{N},\ D2 = \frac{\tilde{N}_{W.}\tilde{N}_{.L}}{N},\ D3 = \frac{\tilde{N}_{\mathrm{L}.}\tilde{N}_{.\mathrm{W}}}{N},\ D4 = \frac{\tilde{N}_{L.}\tilde{N}_{.L}}{N}.$$

$\widetilde{N}$ indicates the theoretical number of funds, under null hypothesis that represents lack of persistence. This statistic, also known as Pearson's statistic, follows a χ^2 distribution with one degree of freedom.

In order to analyze the robustness of our analysis we also performed these two tests by modifying the concept of winner, defining in this way only those funds that exceeded the 75th percentile return. With this approach we are able to verify if a restricted group of funds is able to persistently generate better performance. Indeed, if in the market only few fund managers are able to persistently beat the competitors, when we restrict the winning criterion we could find evidence of persistence. On the other hand, using the median criterion this aspect could be flattened and we may accept the hypothesis of no persistence.

3.1 Results of persistence analysis

Performance persistence results are related to the performance indicator and for this reason we analyse this phenomenon by means of several indicators that present different characteristics.

We show in Table 3 the results of the three statistics described in the previous section. The CPR and chi-square tests confirm the existence of the "*hot hand phenomenon*" for Italian equity mutual funds. The contingence tables for Sharpe, Treynor and Sortino Ratio are very similar for both the definitions of winner (i.e. with respect to median return or 75th return). The measure which generates most persistence is the Morningstar Risk-Adjusted one, as results by p-value of the chi-square test.

For the information ratio, the number of the funds used in calculating persistence is different from the others because some customized benchmarks are not available at the beginning of the sample period. The persistence analysis presents a contradictory behaviour; the CPR confirms the presence of persistence, whereas for chi-square test the null hypothesis of absence of persistence cannot be rejected. This is due to the fact that the information ratio considers the difference between funds returns and customized benchmarks, and this measure orders the funds in a less stable way.

Table 3a. – Persistence analysis based on median performance

	N.	N_{WW}	N_{LL}	N_{WL}	N_{LW}	CPR	Z-Stat.	p-val.	χ^2	p-val.
Sharpe	2932	783	800	676	673	1.377	4.315**	0.000**	4.719	0.029*
Sortino	2932	789	805	670	668	1.419	4.720**	0.000**	5.632	0.017*
Treynor	2932	783	801	676	672	1.381	4.352**	0.000**	4.807	0.028*
Morningstar	2932	883	900	576	573	2.408	11.614**	0.000**	34.324	0.000**
Information	2776	715	747	660	654	1.237	2.801**	0.003**	2.163	0.142

Table 3b. – Persistence analysis based on 75th performance

	N.	N_{WW}	N_{LL}	N_{WL}	N_{LW}	CPR	Z-Stat.	p-val.	χ^2	p-val.
Sharpe	2932	244	1757	468	463	1.978	7.207**	0.000**	10.064	0.002*
Sortino	2932	265	1778	447	442	2.385	9.246**	0.000**	15.493	0.000**
Treynor	2932	255	1768	457	452	2.183	8.279**	0.000**	12.758	0.000**
Morningstar	2932	282	1752	430	425	2.703	10.871**	0.000**	18.169	0.000**
Information	2776	193	1634	479	472	1.395	3.328**	0.000**	3.460	0.063

“ * ” indicates 5% significance whereas “ ** ” indicates 1% significance.

In this Table the following summarizing quantities are related to the 43 contingency tables: “N” indicates the sum of the number of funds in the contingency tables, “N_{WW}” represents the number of persistence cases on “winners”, “N_{LL}” shows the number of persistence cases on “losers”, “N_{LW}” and “N_{WL}” express the number of reversal cases, respectively from “losers” to “winner” and from “winner” to “losers”. In Panel a) and b) winners are defined with respect to the median performance and to the 25th percentile performance respectively.
Moreover, persistence test are exhibited. “CPR” specifies the value of the "Cross-Product Ratio", while the Z-statistic indicates the value of statistic test on “CPR”. “ χ^2 ” indicates the value of the test and “p-value” refers to χ^2 test.

We observe that when we consider winners only those funds that exceeded the 75th percentile return, we find persistence with a higher level of confidence. Therefore, the winning class in this case is more stable than the previous one and consequently there exists a little group of funds that methodically produces better returns and is always at the top of the ranking. When we consider the first winning criterion we included in this class also funds that change rank position very frequently, and so we find less evidence of persistence.

In summary, our results suggest that mutual fund investors can benefit from choosing funds based on past risk-adjusted performance and that persistence is highly related on the performance measure used.

Our findings deviate from most European studies that argue Italian mutual funds display weak persistence (see in particular Grande e Panetta [2002], Bams and Otten [2002]). However, in our study we use (i) different performance measures rather than simply absolute performance, (ii) weekly frequency instead of monthly or quarterly data, and (iii) a more recent sample period. The weak persistence of the Information Ratio suggest to us that differences in the results between our study compared to the previous ones are not mostly related to the sample period or to the frequency of the data considered but particularly to the performance measures used.

Our results highlight other observations. First, the performance measure based on competitors (i.e. on the peer group benchmark) seems to be more persistent. One potential explanation for this result is that this measure partially eliminates the variability due to the market behaviour and is able to capture the peculiarity of the fund with respect to competitors. Clearly, this performance measure is important for investors, since it highlights the "most skilled" fund manager of all the competitors. With our analysis we are able to demonstrate that in the Italian markets there are some managers that persistently perform better than others and that the Morningstar measure allows us to identify these fund managers. This result strengthens the importance of the recent introduction in Italian mutual fund market of Morningstar performance rating performed by private institutions.

4. Conclusion

This paper investigates the extent of performance persistence exhibited by Italian equity mutual funds from January 1997 to December 2000 using the traditional two-period framework. It also examines whether the persistence observed is sensitive to the performance measures used and to the winning criterion. We also compared the fund ranking procedure produced by all risk-adjusted performance measures and above all the ranking produced by the *Morningstar* ones. Our results can be summarized as follows. First, there exists a very high rank correlation between performance measures based on absolute benchmarks and customized benchmarks; on the contrary there is only a weak relation between these measures and the information ratio. Second, the performance measure based on competitors (*Morningstar Risk-Adjusted measure*) is more persistent in both the analyses.

Third, if we consider as winning criterion the returns that exceeded the 75^{th} percentiles even the performance measures based on absolute benchmarks show persistence. The Information ratio exhibits contradictory behaviour.

The results of this paper lead to two clear implications. First, the ranking of the funds, and so the persistence level, are related to the performance indicator chosen. Second, performance measures based on peer group benchmarks allow to distinguish funds that perform persistently better than competitors.

References

Agarwal, V. and Naik, N.Y. (2000) "Multi-period performance persistence analysis of hedge funds", *The Journal of Financial and Quantitative Analysis*, **35(3)**, 327-342.

Agarwal, V. and Naik, N.Y. (2000) "On Taking the 'Alternative' Route: Risks, Rewards, and Performance Persistence of Hedge Funds", *Journal of Alternative Investments*, **2(4)**, 6-23.

Aldrich, P.C. (1987) "Active versus passive: a new look" in *The Journal of Portfolio Management*, **14(1)**, 9–11.

Assogestioni Spa (2000a) "Elenco dei parametri oggettivi di riferimento per i fondi comuni di diritto italiano", available at www.assogestioni.it.

Assogestioni Spa (2000b) "Benchmark e fondi comuni, le regole per interpretare e utilizzare correttamente il parametro oggettivo di riferimento", available at www.assogestioni.it.

Assogestioni Spa (2002) "La classificazione Assogestioni dei fondi di investimento e dei fondi pensione", available at www.assogestioni.it.

Basile, I., Doninelli, N. and Savona, R. (2001) "Management Styles of Italian Equity Mutual Funds", Brescia University, August 2001.

Beltratti, A. and Miraglia, R. (2001) *"I fondi comuni di investimento – Il caso italiano"*, Carocci, Roma.

Blake, C.R., and Morey, M.R. (1999) "Morningstar ratings and mutual fund performance", Working Paper, Fordham University, March 1999.

Blume, E.M. (1998) "An Anatomy of Morningstar Ratings", *Financial Analysts Journal*, March/April 1998, 19-27.

Brown, S.J. and Goetzmann, W.N. (1997) "Mutual fund styles", *The Journal of Financial Economics*, **43**, 373-399.

Brown, S.J. and Goetzmann, W.N. (1995) "Performance persistence", *The Journal of Finance*, **50**, 679-698.

Brown, S.J., Goetzmann, W.N., Ibbotson, R.G. and Ross, A.S. (1992) "Survivorship bias in performance studies", *The Review of Financial Studies*, 553-580.

Brown, S.J., Goetzmann, W.N. and Ibbotson, R.G. (1999) "Offshore hedge funds: Survival & performance 1989-95", *Journal of Business*, **72**, 91-117.

Carhart, M.M. (1997) "On persistence in mutual fund performance", *The Journal of Finance*, **52**, 57-82.

Carluccio, E.M. (1999) "Strategie, benchmarking e performance nell'asset management", *Collana Banca e Mercati*, Bancaria editrice, Roma.

Carpenter, J.N. and Lynch, A.W. (1999) "Survivorship bias and attrition effects in measures of performance persistence", *Journal of Financial Economics*, **54**, 337-374.

Casarin, R., Pellizzon, L. and Piva, A. (2000) "Performances and Performance Persistence of Italian Equity Founds", WP0006 GRETA, June 2000, Venice.

Cesari, R. and Panetta, F. (1998) "Style, fees, and performance of Italian equity funds", *Temi di discussione della Banca d'Italia*, January 1998.

Christensen, R. (1990) "*Log-linear models*", Springer-Verlag, New York.

Coggin, T.D. and Fabozzi, F.J. (1995) "*The handbook of equity style management*", Philadelphia, Fabozzi and Associates Publishing.

Damato, K. (1996) "Morningstar Edges Toward One-Year Ratings." *Wall Street Journal*, April 5th, 1996.

Elton, E.J., Gruber, M.J. and Blake, C.R. (1996) "The persistence of risk adjusted mutual fund performance", *Journal of Business*, **69**, 133-157.

Goetzmann, W.N. and Ibbotson, R.G. (1994) "Do winners repeat?", *The Journal of Portfolio Management*, **20(2)**, 9-18.

Goodwin, T.H. (1998) "The Information Ratio", *Financial Analysts Journal*, July/August 1998, **54(4)**, 34-43.

Grande, G. and Panetta, F. (2002) "Why does performance persist? Evidence from Italian equity funds", Working Paper, 2002.

Grinblatt, M. and Titman, S. (1993) "Performance measurement without benchmarks: an examination of mutual fund returns", *Journal of Business*, **66(1)**, 47-68.

Grinold, R.C. (1999) "Mean–variance and scenario-based approaches to portfolio selection", *The Journal of Portfolio Management*, **25(2)**, 10-22.

Gruber, J.M. (1996) "Another puzzle: the growth in actively managed mutual funds", *Journal of Finance*, **51**,783-810.

Hendricks, D., Patel, J. and Zeckhauser, R. (1993) "Hot hands in mutual funds: short run persistence of relative performance", *Journal of Finance*, **48**, 93-130.

Hull, J.C. (2003) "Option futures and other derivatives", Fifth Edition, Prentice – Hall International, Englewood Cliffs, New Jersey

J. P. Morgan, Reuters (1996) "RiskMetrics™ Technical Document", fourth edition, New York.

Jaffe, C. (1995) "Rating the Raters: Flaws Found in Each Service." *Boston Globe*, August 27th, 1995.

Kent, D., Grinblatt, M., Titman, S. and Wermers, R. (1997) " Measuring mutual fund performance with characteristic-based benchmarks", *Journal of Finance*, **52**, 1035-1058.

Khorana, A. and Nelling, E. (1997) "The Performance, Risk, and Diversification of Sector Funds", *Financial Analysts Journal*, May/June 1997, 62-74.

Kritzman, M. (1986) "How to detect skill in management performance", in *The Journal of Portfolio Management*, **12(2)**, 16-20

Lee, W. (2000) *"Advanced theory and methodology of Tactical Asset Allocation"*, Fabozzi and Associates Publishing, Philadelphia,.

Lucas, L. and Riepe, M.W. (1996) *"The role of returns-based style analysis: understanding, implementing, and interpreting the technique"*, technical document, Ibbotson Associates, Inc., Chicago.

Malkiel, B.G. (1995) "Returns from investing in equity mutual funds 1971 to 1991", *The Journal of Finance*, **50**, 549-572.

Mossavar-Rahmani, S. (1987) "Customized benchmarks in structured management", in *The Journal of Portfolio Management*, **13(4)**, 65-68

Otten, R. and Bams, D. (2002) "European Mutual Fund Performance", *European Financial Management*, **8(1)**, 75-101

Rennie, E.P. and Cowhey, T.J. (1990) “The successful use of benchmark portfolios: a case study”, in *Financial Analysts Journal*, **46(5)**, 18-26.

Roll, R. (1992) “A mean/variance analysis of tracking error”, *The Journal of Portfolio Management*, summer 1992, **18(4)**, 13-23.

Sharpe, W.F. (1992) "Asset allocation: Management style and performance measurement," *The Journal of Portfolio Management*, winter 1992, **18(2)**, 7-19.

Sharpe, W.F. (1988) “Determining a Fund’s Effective Asset Mix,” *Investment Management Review*, 1988, 59-69.

Sharpe, W. F., “Morningstar’s Risk-adjusted ratings, Stanford University”, *Financial Analysts Journal*, July/August 1998, 21-33.

Sharpe, W.F. (1994) “The Sharpe Ratio”, *The Journal of Portfolio Management*, fall 1994, **21(1)**, 49–59.

Simons, K. (1998) “Risk-adjusted performance of mutual funds”, Federal Reserve Bank of Boston, 1998, 33-48.

Sortino, F.A. and Forsey, H.J. (1996) “On the use and misuse of Downside risk”, *Journal of Portfolio Management*, 1996, **22(2)**, 35-43.

Sortino, F.A. and Van De Meer, R. (1991) “Downside risk”, *The Journal of Portfolio Management*, **17(4)**, 27-32.

Tierney, D.E. and Winston, K. (1997) “Using generic benchmarks to present manager styles”, *The Journal of Portfolio Management*, **17(4)**, 33-36.

Treynor, J.L. (1966) “How to rate management investment funds”, *Harvard Business Review*, **43**, 63-75.