Working Paper Series, N. 9-2008, May 2008

Department of Statistical Sciences
University of Padua
Italy

Dicing with the market: randomized procedures for evaluation of mutual funds

Francesco Lisi
Department of Statistical Sciences
University of Padua
Italy

Abstract: A statistically correct procedure for evaluating management skill in the mutual fund industry is described. The approach is based on random and semi-random portfolios and is completely data-driven, allowing reference to a specific benchmark or to peer-groups to be avoided. Rather, it produces a new benchmark, which represents the minimum level of performance that *skilled* managament should reach. An application to some Italian equity funds is also given.

Keywords: performance evaluation, randomized procedures, random portfolios, mutual funds, benchmark.

UNIVERSITÀ DEGLI STUDI DI PADOVA
DIPARTIMENTO DI SCIENZE STATISTICHE

Contents

1 Introduction **1**

2 A randomized procedure for performance evaluation **3**

3 Evaluation of procedure **6**

4 Empirical analyses **7**

4.1 Performance of some Italian mutual funds 7

4.2 Semi-random portfolios 10

5 Conclusions **11**

Department of Statistical Sciences
Via Cesare Battisti, 241
35121 Padova
Italy

tel: +39 049 8274168
fax: +39 049 8274170
http://www.stat.unipd.it

Corresponding author:
Francesco Lisi
tel: +39 049 827 4182
lisif@stat.unipd.it

Dicing with the market: randomized procedures for evaluation of mutual funds

Francesco Lisi
Department of Statistical Sciences
University of Padua
Italy

Abstract: A statistically correct procedure for evaluating management skill in the mutual fund industry is described. The approach is based on random and semi-random portfolios and is completely data-driven, allowing reference to a specific benchmark or to peer-groups to be avoided. Rather, it produces a new benchmark, which represents the minimum level of performance that *skilled* managament should reach. An application to some Italian equity funds is also given.

Keywords: performance evaluation, randomized procedures, random portfolios, mutual funds, benchmark.

1 Introduction

The main interest of subscribers to mutual funds is to know how much their wealth grows and, thus, the whole performance of a fund manager in a fixed interval of time. Closely connected with these points, subscribers are also interested in being able to evaluate the opportunity of investing in funds and spending money to reward the manager's expertise.

The problem of performance evaluation and the methods to decompose portfolio performance has been widely considered in the literature (Lai and Li, 2007; Keswani and Stolin, 2006; Silli, 2006; Kothari and Warner, 2002; Christopherson *et al.*, 1999; Carhart, 1997; Daniel *et al.*, 1997; Chen and Knetz, 1996; Griblatt and Titman, 1993; Grinblatt and Titman, 1989). In particular, performance decomposition is useful in detecting the part of porfolio return to be ascribed to the manager's skill (Wermers, 2000). Unfortunately, in spite of numerous studies, the effective skill of fund managers in generating yields is again a controversial point in the fund management literature (Cuthbertson *et al.*, 2008; Rodriguez, 2008; Barras *et al.*, 2008; Kosowski *et al.*, 2006; Otten and Bams, 2004, ter Horst *et al.*, 2001).

Several studies of fund performance refer to an equilibrium model which, in a quite general formulation, may be written as:

$$r_{p,t} - r_{f,t} = \alpha_t(Z_{t-1}) + \sum_{i=1}^{k} \beta_{i,t}(Z_{t-1})\lambda_{i,t}(Z_{t-1}) + \varepsilon_t. \qquad (1)$$

where $r_{p,t}$ is the fund return at time t, $r_{f,t}$ is the risk-free return, λ_i are risk factors that determine expected returns in financial markets, and ε_t is an error term, often assumed to be Gaussian and independent. Z_t are variables on which the model parameters depend.

Referring to such a model, parameter α is commonly interpreted as a measure of over- or under-performance relative to the market proxy used. If model (1) is correct, positive deviations of α from zero may be interpreted as representing the selection skill of managers. Examples of works applying model (1) for performance evaluation are Cuthbertson *et al.* (2008), Barras *et al.* (2008), Ferson *et al.* (2006), Silli (2006), Capocci and Hüber (2004), Otten and Bams (2004).

Another 'classical' way of looking at mutual fund performance is based on examination of persistence of the results, that is, the degree of positive serial correlation in returns of managed portfolios (Christophersen *et al.*, 1998; Carhart, 1997; Goetzmann and Ibbotson, 1994; Brown and Goetzmann, 1995; Hendricks *et al.*, 1993; Griblatt and Titman, 1992, among others). Lastly, fund performance has also been studied with respect to a benchmark (e.g., Daniel *et al.*, 1997; Lehmann and Modest, 1987), peer-groups (Ankrim, 1998), to some specific rating (Casarin *et al.*, 2005) or by style analysis (Dor *et al.*, 2008; ter Horst *et al.*, 2004; Bogle, 1998).

In this work evaluation of managers' skill is faced through a completely different, statistical, simulative approach. The underlying idea is that of comparing results of a random or semi-random, managerial strategy with a real, or reasoned, one. This comparison is not aimed at establishing if a random investment strategy is to be preferred to a reasoned one, but at verifying if results obtained by well informed managers are consistent with those generated by unskilled ones. A second goal is to define a minimum return level, under which results cannot be ascribed to management skill and, thus, cannot justify management fees.

The idea of using random portfolios for mutual fund evaluation is not completely new. It originates in the celebrated book by Malkiel (1973), *A random walk down Wall Street*, and in the now historic Investment Dartboard column, published by the Wall Street Journal. For 14 years, starting from 1988, this gave rise to a challenge, called the *Dartboard Contest*[1], among professional management and portfolios chosen at random. However, this challenge, which attracted much interest and produced a number of studies from academics and professionals, shows several questionable issues in its regulation, from both statistical and financial points of view (e.g., Heron *et al.*, 2003). Some of these are the too low number of assets included in the portfolios and the number of of portfolios themselves, the effect of price pressures in the stocks picked by professionals, which may favor managers and comparison between risk-unadjusted portfolios.

This work aims to apply random portfolio logic following a correct statistical approach, which requires a mutual fund return to be compared with the whole distribution of random portfolio returns. To achieve this goal, in a correct statistical framework means that two antithetical hypotheses must be compared: that of no management skill, versus an alternative of some kind of real skill. In order to dis-

[1]In Italy, a similar contest, called *Bull Hunt (Caccia al Toro)*, is proposed annually by the financial and economic newspaper *Il Sole 24 ore*

criminate between these two hypotheses starting from a net asset value time series, we need a performance measure, the distribution of no-skill performance and to verify if fund performance is consistent with this distribution. The no-skill performance distribution may be obtained by considering random portfolios, i.e., those composed of randomly selected assets.

The resulting procedure differs from others proposed in the literature (e.g., Surz, 2005) and shows several interesting features: first, it refers to a very intuitive concept of skill, in line with that expected by the investor. In addition, being a model-free method, it does not require any parameter estimation, thus avoiding the need for long time series and reducing bias-survival problems (Elton *et al.*, 1995). Due to the nonparametric nature of the procedure, distributional assumptions, particularly that of the Gaussianity of returns, are not necessary. Also, the proposed method is very flexible and can be applied in quite different contexts, allowing different return and risk measures to be used and costs and managerial constraints to be examined. Last but not least, it is independent both of the classical idea of benchmark and of a peer-group comparison.

The rest of the paper is organized as follows. Section 2 described the basic form of the procedure and some generalizations are discussed. Section 3 considers the problem of summarising results and proposes some possible statistics. An application of the procedure to some Italian funds, with relative results, is described in Section 4. The same Section also shows how to use semi-random portfolios for performance evaluation. Section 5 concludes.

2 A randomized procedure for performance evaluation

Given a fund F, we focus on evaluating the hypothesis system

$$\begin{cases} H_0 : F \text{ show skill} \\ H_1 : F \text{ does not show skill.} \end{cases} \tag{2}$$

Before going into the merits of testing system (2), it should be stressed that the term *skill* referring to a fund might be ambiguous and not precise. In fact, a fund, which is able to refund management costs but which does not lead to any other extra-return for the investor, strictly speaking does show all some kind of skill: that of rewarding the manager's work. It is clear, however, that this skill is far from being attractive for the investor. Instead, the notion of skill to which it refers in the following is, instead, the ability of a fund to yield returns, net of the manager's reward, significantly higher than that which could be obtained by a manager without any specific skill.

A correct statistical evaluation of system (2) requires a performance measure $r(F)$ and its distribution under H_0. Testing system (2), thus, consists of comparing observed performance $r(F)$ with distribution under the null hypothesis.

Defining the distribution of $r(F)$ under H_0 means considering all possible results that can be obtained without ability, and estimating, for each result, the probability (density). The exact distribution is very difficult to obtain analytically, but a hopefully good approximation, can be obtained through random portofolios, that is

portfolios composed of randomly selected assets. Since this kind of selection does not require any specific ability, its results may be considered as representative of the situation defined by the null hypothesis.
In this context, let NAV_t, $(t = 1, ..., T)$, be the Net Asset Value time series of a fund. As a performance measure, we consider the holding period return, $r_t(F) = (NAV_t/NAV_1) \times 100$.
The procedure for testing hypothesis system (2) starting from time series NAV_t, is the following.

1. Let $\mathbf{S} = \{S_i\}$, $i = 1, ..., N_S$ be the manager's investment universe, that is the basket of all the financial assets to which the manager refers for his investments. It represents all the possible portfolio components that a fund manager might have.

2. Generate M portfolios by random sampling k assets from $\mathbf{S}$.

3. For each of the M portfolios F_i, $(i = 1, ..., M)$, calculate the holding period return $r_t(F_i)$, $(i = 1, ..., M)$ and $(t = 1, ..., T)$. For each time t, the set of M measures $r_t(F_i)$ may be used to estimate the distribution under the null by means of the empirical distribution.

4. For each time t, compare the observed return, $r_t(F)$, with the distribution of returns obtained through random portfolios. Let PV_t be the portion of random portfolios which led to a peformance greater than that of the fund. It is intuitive that, the higher PV_t is, the more we are inclined to conclude for the no-skill hypothesis.

5. Let α be the fraction which discriminates between the two hypotheses. Thus, for each given t, if $PV_t > \alpha$, the null hypothesis is accepted at the $(1 - \alpha)$ confidence level (see Figure 1).
   Alternatively, we can consider return level $r_t(F_{1-\alpha})$ corresponding to the $(1 - \alpha)-$th quantile of the distribution of $r_t(F_i)$: it represents the best performance that can be reached, at the $(1-\alpha)$ confidence level, without skill. Equivalently, $r_t(F_{1-\alpha})$ can also be viewed as the minimum performance level that a *skilled* manager should reach.

Note that, unlike the usual situation, the comparison is not made between the fund return and the mean of the random results [2], but with a suitable quantile. The logic is that all results referring to simulated portfolios are obtained completely at random but a fraction equal to α is discarded because it is considered too lucky.

The procedure now described is in its basic form. However, to be applied in a real context it must be generalized in several aspects, which are potentially critical in applications. The first and most important issue is how to account for the degree of risk inherent in various portfolios. Clearly, it makes sense to compare the performance of random portfolios with that of real ones only if risks are equal because, otherwise, one might conclude that higher performance is only due to a higher risk.

[2]Even, in the Dartboard Contest and the Bull Hunt challenge, each comparison is made with only one portfolio!

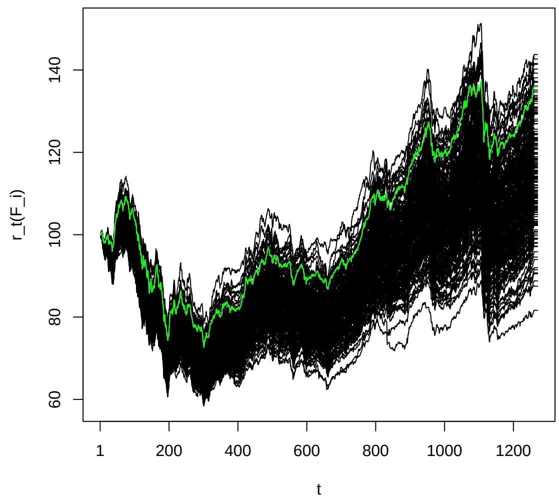

Figure 1: Example of distribution of $r_t(F_i)$. White line is quantile, $r_t(F_{0.95})$.

Analogously, one could argue that minimizing risk may be preferred to maximizing performance and also that the ability to keep down risk is not negligible. In order to account for the level of risk there are two possible approaches. The first consists of checking *ex post* the riskiness of random portfolios: if the percentage with risk level lower than or equal to that of the observed fund is very high, then the comparison may be considered fair.
An alternative is adjusting performance with a penalty connected to the risk level. For example, a possible risk-adjusted performance measure (Modigliani and Modigliani, 1997; Rodriguez and Shapiro, 2007) is:

$$R_t(F_i) = \frac{V_F}{V_{F_i}} r_t(F_i), \tag{3}$$

with V_{F_i} and V_F risk measures, respectively, of the $i-$th random portfolio and of the fund. In this way, a given performance level reached by higher risk would be penalized by factor V_F/V_{F_i}. In the application of Section 4.1 measure (3) will be used.
In the first step of the procedure knowledge of $\mathbf{S}$, the manager's investment universe, is assumed. This piece of information is not always available, as only the manager knows exactly his reference basket. However, a reasonable approximation may be given by the universe implied by the benchmark declared in the fund prospectus.
As regards the parameters of the procedure, the number M of portfolios should be high, so that particularly lucky or unlucky situations can compensate each other. The number k of stocks in the portfolio should not be too low in order to have a sufficiently differentiated portfolio. Statman (1987) points out that, for a random

portfolio, 30-40 stocks are enough to have a neglectable unsystematic risk[3].
Weights w_j of the k assets may be equal ($w_j = 1/k, \forall j$) or also randomly selected. In this work, equally-weighted portfolios are considered because it is the simplest solution among those managed by someone without skill.
As managers react to market changes and phases, the composition of portfolios should be updated periodically. As a consequence, for a fair comparison, transaction costs, as well as all other kinds of costs, must be considered.

3 Evaluation of procedure

Given time interval $[t_1, t_T]$, the above method allows us to declare - within a confidence level - whether the null hypothesis of system (2) can be accepted or not. In practice, however, it is not obvious to which temporal interval one should refer as a time which, naturally, should be considered *final time* does not exist. If fund performance $r_t(F)$ is always higher or always lower than quantile $R_t(F_{1-\alpha})$, there are fewer problems because the conclusion is the same each time. Unfortunately, this situation occurs only rarely, whereas, the fund return more often out- or underperforms $R_t(F_{1-\alpha})$ only in some periods.
The problem is thus how to evaluate the procedure not only in a given time interval, but also with respect to several periods, for example defining the starting time but not the final time. Unfortunately when, instead of considering a precise final time, several times are considered, the test level is barely controllable, all the more because the sequence of tests is not independent. The first consequence is that it is difficult to achieve a well defined confidence level and one must be content with having merely indicators of the situation. Choosing them is not univocal and is important, in order to to have statistics able to give reliable indications on fund performance with respect to that of random portfolios. Among the several statistics available, the following are particularly informative:

- Daily P-value (PV_t):

$$PV_t = \frac{\sum_{i=1}^{M} I(r_t(F_i) > r_t(F))}{M} \tag{4}$$

  with $I(\cdot)$ as indicator function. It represents the fraction of M random portfolios that outperform the fund at time t. The graph of PV_t gives a direct description of the relation existing between fund performance and M random portfolio performance. In addition, it enables us to consider the persistence of results over time.

- Percentage of Outperformance (PO_t), i.e, the percentage of days, witin a given period, in which the fund outperforms the quantile:

$$PO_t = \frac{\sum_{t=1}^{T} I(r_t(F) > r_t(F_{1-\alpha}))}{T} \cdot 100 \tag{5}$$

[3]Thus, portfolios composed of 4 (as in the case of the Dartboard Contest) or 5 (as in the case of the Bull Hunt challenge) also contain a significant amount of unsystematic risk and show a higher volatility level than a well-diversified portfolio.

with $I(\cdot)$ indicator function and $r_t(F_{1-\alpha})$ performance of the $(1-\alpha)$-quantile. Note that the so-called Shortfall Probability (the probability that the fund reaches a result worse than the quantile), is only $1 - PO_t$.

- Quantile Tracking Error:

$$QTE = \frac{\sum_{t=1}^{T}(r_t(F) - r_t(F_{1-\alpha}))}{T}. \tag{6}$$

  This tells by how much, on average, fund performance is higher than quantile performance.

- Quantile Tracking Error Volatilty:

$$QTEV = \frac{\sum_{t=1}^{T}(r_t(F) - r_t(F_{1-\alpha}))^2}{T}. \tag{7}$$

  This gives indications on how much variability there is around QTE.

Of course, for risk-adjusted comparisons, return measures of random portfolios must be adjusted for risk and, thus, $R_t(F_i)$ must be used instead of $r_t(F_i)$. Since the use of a risk-adjusted return implies using a measure of risk, the above indicators also depend on the volatility measure in question.

4 Empirical analyses

4.1 Performance of some Italian mutual funds

In order to give an example of how the above method works, it was applied to 23 Italian funds (see Tables 1 and 2), with an Italian equity component of at least 90% and whose investment universe, deducible from the declared benchmark, is represented by all stocks quoted on the Milan stock exchange, with a small monetary component when necessary. Analyses were carried out for the period January 2002 - January 2007.

For each fund, $M = 1000$ random portfolios, composed of $k = 30$ stocks, were generated. All portfolios were updated every 35 working days. At each update, it was assumed that the whole portfolio was re-allocated and that the entire value was reinvested in new stocks (some of which might be the same as before). Some preliminary analyses showed that the choice of k and of the updating period are critical only for very low or very high values, whereas for $25 \leq k \leq 50$ and for refreshing intervals between 25 and 50 working days, it is not so crucial.
Transaction costs were fixed at 0.1% of the invested capital and taxes, following Italian law, were set at 12.5% of the net return.
Incoming and outcoming cash flows, caused by subscription requirements and quote repayments, were ignored.
As a performance measure, the holding period return, adjusted for risk following expression (3) was considered. The risk measures used for adjustment were standard deviation, semi-variance and 1% Value-at-Risk, calculated by a nonparametric

approach.
Figure 2 displays three examples of fund returns (broken line) and 95% quantile returns (full line), the latter adjusted through standard deviation. The graphs of Figure 2 should be analysed together with those of the corresponding daily p-values (Figure 3). It is quite clear that the first fund tends to outperform the quantile, showing discrete management skill and that the performance of the second is not very different from that of the quantile, whereas the performance of the third is less than that of the quantile, suggesting quite poor management skill. Moreover, note that in bearish market periods, the portfolio of the real fund tended to to lose competitive edge with respect to random portfolios.
Statistics useful for comparative performance evaluation are listed, for all analysed funds, in Table 1 in the case of standard deviation-adjusted returns and in Table 2 for VaR-adjusted returns. The results obtained by adjusting returns with semi-variance are very similar and are therefore not reported. In 15 cases out of 23, adjusting for standard deviation the fund outperforms the quantile less than 60% of days (in 12 cases out 23, in the VaR-adjusted case). In addition, only for 8 funds (7 in the VaR-adjusted case) was the QTE indicator higher than 2: i.e. in most cases, the fund and the quantile, on average, lead to the same adjusted return.
It is worth highlighting the fact that the percentage of random portfolios with risk, as measured by standard deviation or VaR, lower than that of the fund was generally very high (98% or higher) and that the risk levels were very similar. Thus, it seems that the adjustment only has a limited effect.
On the whole, the difference between kinds of adjustments is not dramatic, although there are cases in which the two risks penalize or reward funds in different ways. For example, the Euromobiliare Azioni Italiane fund outperforms the quantile in 56.38% of days in the SD-adjusted case but only in 47.4% of days with VaR-adjustment. This shows that the fund is particularly affected by extreme risks. The opposite situation occurs for the Azimut Trend Italia fund, whose percentage of the quantile outperforming days is larger when the return is adjusted for VaR, which indicates a better attitude towards attempting to avoid high losses.
Tables 1 and 2 also show, for each fund, the Morningstar evaluation based on a rating system which assigns from 1 to 5 stars according to the method described in Morningstar (2006), which accounts for return and risk levels and management costs. With respect to the Morningstar rating, the proposed evaluation statistics often appears consistent. However, it is interesting to note that, while funds with a high PO index generally also have high ratings, some differences in ratings are difficult to explain in the light of analyses based on random portfolios. For example, the Azimut Trend Italia and Euromobiliare Azioni Italiane funds receive a five-year Morningstar evaluation, respectively, of 5 and 1 stars, whereas according to the statistics in Table 1,the two funds do not appear very different.

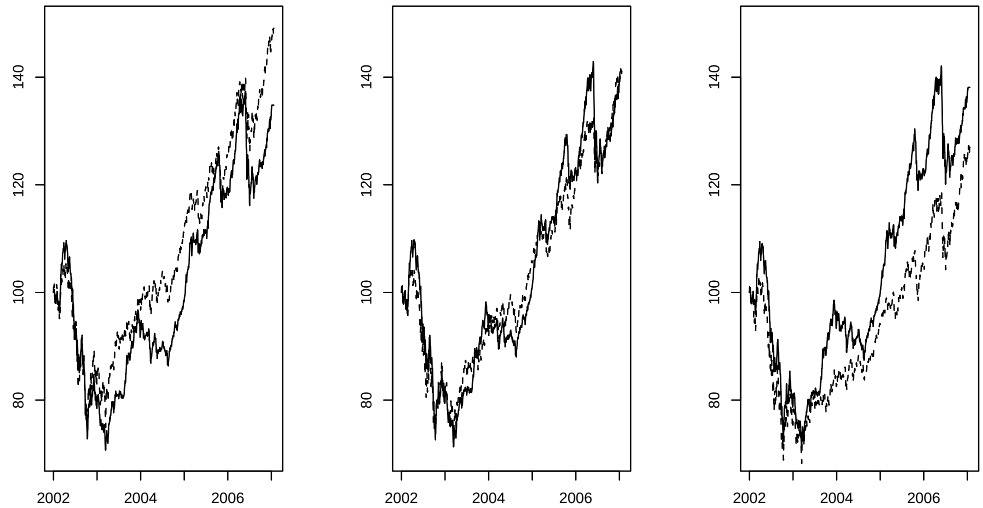

Figure 2: Some examples of holding period returns of a fund (broken line) and of quantile (full line). Left to right: Fondersel Italia, Nextra Azioni Italia, Gestielle Italia.

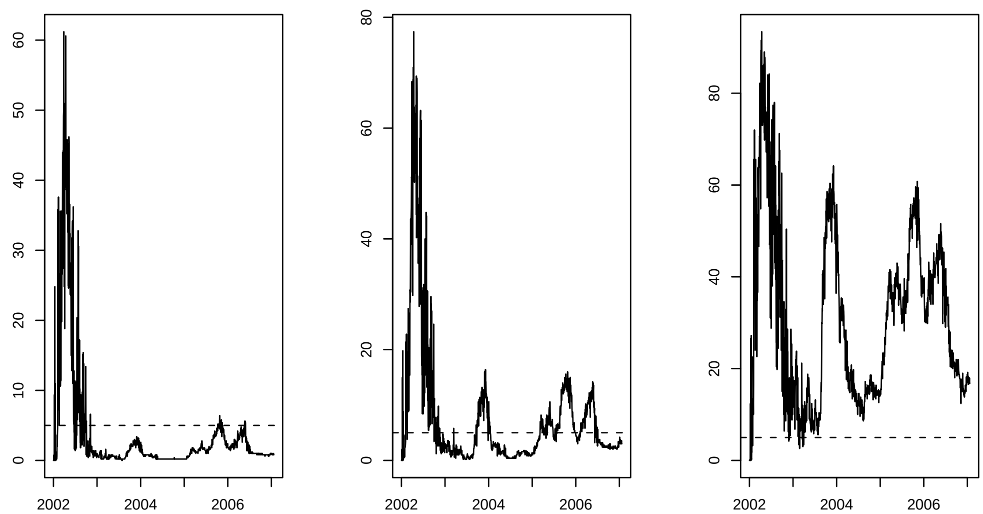

Figure 3: PV_t ($\times$ 100) for Fondersel Italia, Nextra Azioni Italia and Gestielle Italia funds. Broken line set at 5%.

4.2 Semi-random portfolios

To better appreciate and evaluate the results of the previous section, it is useful to consider some criticisms which may be made at this approach. It is often said that this kind of comparison is not fair because it is made between a single yield, that observed for the fund, and a number of possible yields among which, necessarily, there will also be the best ones. A second argument in favour of the manager is that, if the performance lies near the quantile, the manager is able to reach high-level performance with respect to all possible results, avoiding most negative situations.
Actually, neither statement is correct because they do not take into account the fact that the distribution on which the quantile is calculated is generated by fully random portfolios. Working in this way, particularly good portfolios are highly inprobable. Excellent portfolios would lead to results much better that those which could be reached by randomly selecting stocks. For example, for the same data and period, let us suppose we can find, at each updating, the best 30 stocks on the market. At each refresh, the probability of selecting such stocks at random is of order 10^{-42}, more or less the same probability that heads will turn up at the toss of a coin 137 times consecutively. Instead, this optimal and quite unrealistic choice would lead to a mean return over five years of about 16000%. Even relaxing the constraint and chosing 30 stocks at random among the top 50% (that is among the best 155), the probability of this happening, at each refresh, would be of the order of 10^{-10} and the mean period return around 1000%, far higher that the 95% quantile return of the random portfolios, which is $40-50\%$. If we relax the constraint further, we can assume that we can choose 20 out of the 30 stocks among the top 50% and 10 among the worse 50%. Now, at each updating, the probability that this occurs only by a lucky chance is about 0.025 but, again, the final mean return is around 150%. Even selecting the stocks among the top 90%, which means excluding the worst 10%, it would reach a five-year return of almost 100%. A performance level similar to that of the 95% quantile is obtained when, at each refresh, the worst 7% stocks are excluded, which would not seem a particularly demanding task. Figure 4 displays graphs of the *median* holding period return (over 100 simulations) for some semi-random portfolios. Lastly, Table 3 lists the summary statistics when the comparison is made between portfolios composed of stocks selected among the best 90% and others composed entirely at random. Clearly, these statistics are quite different from those obtained with real funds.
These analyses allow us to make some interesting remarks. First of all, it is not true that the distribution of random portfolios includes particularly lucky situations. In addition, if managers were truly able to select the best stocks, even in an ample sense, performance would be far higher and very different from that of the random portfolios. These analyses also show that even the ability to exclude a small fraction of the worst stocks would be sufficient for a fundamental increment in performance. This in turn suggests that, instead of focusing on stock selection, it might be enough to focus on stock exclusion.

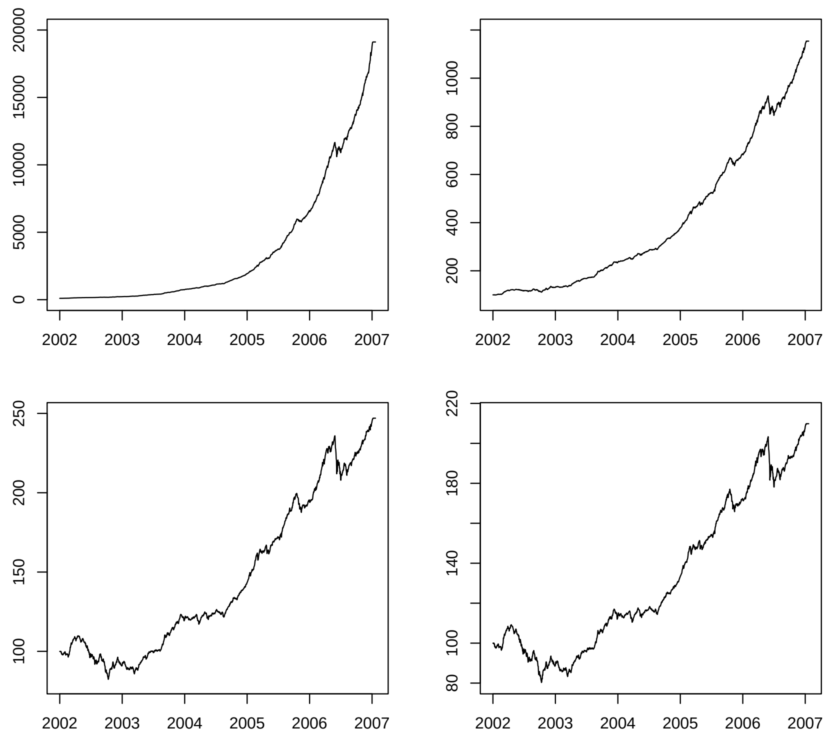

Figure 4: Median of holding period returns when stocks are selected at random among (clockwise) a) top 15% stocks; b) top 50% stocks; c) 20 stocks among top 50% and 10 among worst 50% stocks; d) among top 90% stocks. Note change of scale.

5 Conclusions

In this work a randomized procedure for evaluating the hypothesis that a fund may exhibit management skill is proposed. It is a statistically correct method, based on the comparison between the observed performance of a fund and the distribution of performance under the no-skill hypothesis, the latter estimated by random portfolios. The procedure allows us to compare performances which are risk-equivalent and, thanks to its flexibility, can be applied to any measure of return and risk and can also easily be adapted to complex investment strategies. The approach is simple and intuitive and avoids reference to a benchmark. Rather, it produces a new benchmark, represented by the quantile of the distribution of the performance of random portfolios, which defines the minimum level of performance that a *skilled* manager should reach. Another attractive feature is that it is completely data-driven.

A statistical advantage is that, being a nonparametric method, it does not require the estimation of a model or assumptions on return distribution. The proposed

algorithm can also be applied to short periods, avoiding the need for very long time series.
The main statistical limitation is that when the fund is evaluated considering several time intervals, the level of the test can no longer be controlled. A second limitation is the assumption that the investment universe of the manager is known, whereas in practice this piece of information may not be known to the analyst. However, it is definitely known by the manager, who can use this approach for self-assessment. Lastly, it should be highlighted that this is a case of ex-post evaluation so that the procedure has no perspective meaning. However, it is interesting to note that this evaluation does not relate to the performance of other managers, but it is an absolute one, avoiding any peer-group comparison.
The empirical results from applying this method to 23 Italian mutual funds point out that only a small part of them ahow clear management skills. Also, the use of semi-random portfolios shows that, even when the fund return outperforms the quantile, the results are less than those which could be attained if a small percentage of the worst stocks could be excluded.

References

Ankrim E.M. (1998), Peer-relative active portfolio performance: it's even worse than we thought, *The Journal of Performance Measurement*, Summer issue, 6-11. Barras L., Scaillet O., Wermers R. (2008), False discoveries in mutual fund performance: measuring luck in the estimated alphas, Working Paper n. RHS-06-043, Robert H. Smith School of Business, University of Maryland.
Bogle J.C. (1998), The implication of Style Analysis for mutual fund performance evaluation, *Journal of Portfolio Management*, 24, 34-42.
Brown S.J., Goetzmann W. (1995), Performance Persistence, *Journal of Finance*, 50, 679-698
Capocci D., Hüber G. (2004), Analysis of hedge fund performance?, *Journal of Empirical Finance*, 11, 55-89.
Carhart M.M. (1997), On persistence in mutual fund performance, *Journal of Finance*, 52, 57-82.
Casarin R., Lazzarin M., Pelizzon L., Sartore D. (2005), Relative benchmark rating and persistence analysis: evidence from italian equity funds, *European Journal of Finance*, 11, 297-308.
Chen Z., Knez P.J. (1996), Portfolio Performance Measurement: Theory and Applications, *Review of Financial Studies*, 9, 511-556.
Christopherson J.A., Ferson W.E., Turner A.L. (1999), Performance evaluation using conditional alphas and betas, *Journal of portfolio management*, 26, 59-72.
Christopherson J.A., Ferson W.E., Turner A.L. (1998), Conditioning manager alphas on economic information: another look at persistence of performance, *Review of financial studies*, 11, 511-556.
Cuthbertson K., Nitzsche D., O'Sullivan N. (2008), UK mutual fund performance: skill or luck?, *Journal of Empirical Finance*, doi:10.1016/j.jempfin.2007.09.005.
Daniel K., Grinblatt M., Titman S., Wermers R. (1997), Measuring Mutual Fund

Performance with Characteristic-Based Benchmarks, *Journal of Finance*, 52, 1035-1058.

Dor A.B., Budinger V., Dynkin L., Leech K. (2008), Constructing peer benchmarks for mutual funds: a style analysis-based approach, *Journal of Portfolio Management*, 34, n.2.

Elton E.J., Gruber M.J., Blacke C.R. (1995), Survivor bias and mutual fund performanceSurvivor bias and mutual fund performance *The Review of Financial Studies*, 9, 1097-1120.

Ferson W.E., Sarkissian S., Simin T. (2006), Asset Pricing models with conditional betas and alphas: the effectf of data snooping and spurious regression, *Journal of Financial and Quantitative Analysis*, forthcoming

Goetzmann W., Ibbotson R. (1994), Do winner repeat? Patterns in mutual fund performance, *Journal of Portfolio Management*, 20, 9-17.

Grinblatt M., Titman S. (1993), Performance Measurement without benchmark: An Examination of Mutual Fund Returns, *Journal of Business*, 6, 47-68

Grinblatt M., Titman S. (1992), The Persistence of Mutual Fund Performance, *Journal of Finance*, 47, 1977-1984

Grinblatt M., Titman S. (1989), Portfolio Performance Evaluation: Old Issues and New Insights, *The Review of Financial Studies*, 2, 393-421

Hendricks D., Patel J., Zeckhauser R. (1993), Hot hands in mutual funds: short-run persistence of relative performance, 1974-1988, *Journal of Finance*, 48, 93-130.

Heron R.A., Larsen G.A., Resnick B.G. (2003), The improper use of dartboard portfolios as performance benchmarks, *Financial Decisions*, Spring 2003, article 4.

Lai S, Li H. (2007), The performance evaluation for fund of funds by comparing asset allocation of mean-variance model or genetic algorithms to that of fund managers, *Applied Financial Economics*, 18, 485-501.

Lehmann B.N., Modest D. (1987), Mutual fund performance evaluation: a comparison of bechmarks and benchmarks comparison, *Journal of Finance*, 42, 233-266.

Keswani A., Stolin D. (2006), Mutual funds performance persistence and competition: a cross-sector analysis, *The Journal of Financial Research*, XXIX, 349-366.

Kosowski R., Timmermann A., Wermers R., White H. (2006), Can Mutual Fund "Stars" Really Pick Stocks? New Evidence from a Bootstrap Analysis, *Journal of Finance*, 61, 2551-2595.

Kothari S.P., Warner J.B. (2001), Evaluating mutual fund performance, *The Journal of Finance*, LVI, 5, 1985-2010.

Malkiel B. (1973), *A random walk down wall street*, W. W. Norton & Company.

Modigliani F., Modigliani L. (1997), Risk-adjusted performance, *Journal of Portfolio Management*,23, 45-54.

Morningstar (2006), The Morningstar rating methodology, *Morningstar Methodology Paper*.

Otten R., Bams D. (2004), How to measure mutual fund performance: economic versus statistical relevance, *Accounting and Finance*, 44, 203-222.

Rodriguez J. (2008) European mutual funds and portfolio's country exposure: does active management add value?, *Applied Financial Economics*, forthcoming.

Rodriguez A.E., Shapiro S.J. (2007) Risk-ajusted performance as a rigorous approach to removing subjectivity from expert assesments of suitability, *Journal of*

Business Valuation and Economic Loss Analysis,2, article 3.
Silli B. (2006), Modern approaches in the mutual funds evaluation of management skill in the mutual fund industry, Working paper, Universitat Pompeu Fabra, Barcelona.
Statman M. (1987), How many stocks make a diversified portfolio?, *Journal of Financial and Quantitative Analysis*, 22, 353-363.
Surz R.J. (2005), Testing the hypothesis of "hedge fund performance is good", *Journal of Wealth Management*, spring issue.
ter Horst J.R, Nijman T.E., de Roon F.A. (2004), Evaluating style analysis, *Journal of Empirical Finance*, 11, 29-53.
ter Horst J.R, Nijman T.E., Verbeek M. (2001), Eliminating look-ahead bias in evaluating persistence in mutual fund performance, *Journal of Empirical Finance*, 8, 345-373.
Wermers R. (2000), Mutual Fund Performance: An Empirical Decomposition into Stock-Picking Talent, Style, Transaction Costs, and Expenses, *Journal of Finance*, 55, 1655-1695.

Fund	MRO	MR5	PO	MPV	QTE	QTEV
Apulia Azioni Italia	3	2	53.7	8.31	0.06	15.28
Arca Azioni Italia	3	3	53.94	7.81	0.52	16.79
Aureo Azioni Italia	3	3	45.2	10.92	-0.93	14.33
Azimut Trend Italia	5	5	58.35	7.37	0.87	22.11
Bipiemme Italia	4	4	92.52	1.57	7.63	90.57
Bipitalia Azioni Italia	3	2	43.94	9.8	-1.03	19.25
BNL Azioni Italia	3	2	52.52	8.63	-0.02	15.99
CA-AM MIDA Azionario Italia	3	3	53.23	8.07	¡0.01	18.97
Capitalgest Italia	2	2	41.81	11.77	-2.26	32.24
Capitalia Az Italia	3	2	73.31	5.07	2.12	15.79
Credit S Az Italia B	3	3	50.24	10.1	-0.46	17.82
Ducato Geo Italia	3	4	60	8.79	1.12	17.41
Euromobiliare Azioni Italiane	2	1	56.38	7.18	0.94	26.98
FondErsel Italia	4	4	89.06	3.44	6.47	66.15
Generali Capital	3	3	83.62	4.09	4.7	45.12
Gestielle Italia	2	1	1.1	33.33	-9.46	126.89
Gestnord Azioni Italia	3	3	54.65	8.02	0.41	16.14
Imi Italy	4	4	73.54	7.96	3.66	43.31
Leonardo Azionario Italia	4	4	82.52	4.44	4.56	42.49
Nextra Azioni Italia	3	3	48.98	9.37	-0.51	15.56
Nextra Azioni Italia Din	4	4	79.21	4.86	4.09	36.07
Optima Azionario Italia	4	4	57.64	7.19	1.2	16.92
Pioneer Azionario Crescita A	4	4	80.55	4.14	4.35	40.56

Table 1: Performance of the Italian funds: summary statistics for standard deviation-adjusted returns. MRO=Overall Morningstar Rating, MR5=five-year Morningstar Rating, PO=percentage of outperformance, MPV=mean daily p-value, QTE=quantile tracking error, QTEV=quantile tracking error volatility.

Fund	MRO	MR5	PO	MPV	QTE	QTEV
Apulia Azioni Italia	3	2	52.05	8.06	0.1	15.12
Arca Azioni Italia	3	3	54.8	8.02	0.56	15.93
Aureo Azioni Italia	3	3	45.28	10.52	-0.56	12.59
Azimut Trend Italia	5	5	62.68	6.99	1.32	26.04
Bipiemme Italia	4	4	91.97	1.61	7.46	89.84
Bipitalia Azioni Italia	3	2	46.69	9.89	-1.17	21.02
BNL Azioni Italia	3	2	63.78	7.92	1.49	16.05
CA-AM MIDA Azionario Italia	3	3	54.49	7.8	0.5	14.94
Capitalgest Italia	2	2	24.88	14.32	-3.37	29.51
Capitalia Az Italia	3	2	71.73	5	1.95	12.88
Credit S Az Italia B	3	3	49.84	9.15	0.07	16.48
Ducato Geo Italia	3	4	46.69	11.05	0.03	11.88
Euromobiliare Azioni Italiane	2	1	47.4	8.25	-0.41	37.58
FondErsel Italia	4	4	86.85	3.37	5.63	54.12
Generali Capital	3	3	83.46	4.05	5.21	54.31
Gestielle Italia	2	1	1.02	36.46	-8.33	84.72
Gestnord Azioni Italia	3	3	52.83	8.66	0.09	13.99
Imi Italy	4	4	71.26	9.02	3.53	49.43
Leonardo Azionario Italia	4	4	81.34	5.48	4.6	45.45
Nextra Azioni Italia	3	3	55.28	8.26	0.55	15.08
Nextra Azioni Italia Din	4	4	74.49	6.3	2.92	26.29
Optima Azionario Italia	4	4	61.34	6.83	1.41	17.89
Pioneer Azionario Crescita A	4	4	81.81	3.61	4.11	33.19

Table 2: Performance of the Italian funds: summary statistics for VaR-adjusted returns. MRO=Overall Morningstar Rating, MR5=five-year Morningstar Rating, PO=percentage of outperformance, MPV=mean daily p-value, QTE=quantile tracking error, QTEV=quantile tracking error volatility.

Fund	PO	QTE	QTEV
Top 90%	93.34	30.4	1407.06

Table 3: Performance of the median of 100 portfolios composed of stocks selected at random among top 90%.

Working Paper Series

Department of Statistical Sciences, University of Padua

You may order paper copies of the working papers by emailing wp@stat.unipd.it

Most of the working papers can also be found at the following url: http://wp.stat.unipd.it

UNIVERSITÀ DEGLI STUDI DI PADOVA
DIPARTIMENTO DI SCIENZE STATISTICHE

Department of Statistical Sciences
University of Padua
Italy