Backward/forward optimal combination of performance measures for equity screening

Monica Billio[a], Massimiliano Caporin[b,*], Michele Costola[a]

[a] *University Ca' Foscari Venezia, Department of Economics*
[b] *University of Padova, Department of Economics and Management "Marco Fanno"*

Abstract

We introduce a novel criterion for performance measure combination designed to be used as an equity screening algorithm. The proposed approach follows the general idea of linearly combining selected performance measures with positive weights and combination weights are determined by means of an optimisation step. The underlying criterion function takes into account the risk-return trade-off potentially associated with the equity screens, evaluated on a historical and rolling basis. By construction, performance combination weights can vary over time, allowing for changes in preferences across performance measures. An empirical example shows the benefits of our approach compared to naive screening rules based on the Sharpe ratio.
Keywords: performance measures, combining performance measures, portfolio allocation, equity screening, differential evolution.
JEL codes: C44, C58, C61, G11, G17.

*Corresponding author: University of Padova, Department of Economics and Management "Marco Fanno", Via del Santo, 33, 35123 Padova, ph. +39-049-827-4258

Email addresses: `billio@unive.it` (Monica Billio), `massimiliano.caporin@unipd.it` (Massimiliano Caporin), `michele.costola@unive.it` (Michele Costola)

Authors acknowledge financial support from the project MIUR PRIN MISURA - Multivariate models for risk assessment and GRI in Financial Services for financial support. The second author acknowledges financial support from the University of Padova grant CPDA073598.

1. Introduction

The investment process is described by the complete set of actions taken by a portfolio manager, including the definition of the investment objectives and the associated strategic allocation, the construction of tactical asset allocation and stock selection, and general rules for portfolio monitoring (see for example Grinold and Kahn 1999). The security selection step focuses on identifying the most promising investment opportunities, represented by specific assets. Different approaches might be employed at this stage, inspired by technical analysis or based on a more fundamental analysis. In general, security selection methodologies can be classified as qualitative or quantitative. The latter presumes the existence and the use of some quantitative tools.

The broad class of quantitative security selection instruments includes the so-called equity screening rules, methodologies whose purpose is to rank a large set of assets in order to focus attention on the best ones or to exclude the worst ones. Screening rules can be used directly as security selection tools or might simply represent a first step in a security selection procedure; in fact, they permit to restrict the investment universe to a reasonably limited set of assets, to be analysed in greater detail by analysts. We stress that screening rules should, when used as asset allocation tools (for instance by directly investing in the best assets) might turn out to be suboptimal, since they do not control for the correlation across assets.

Relevant and relatively simple examples of screening rules are given by performance measures; these are quantities that, in most cases, represent a remuneration per unit of risk, or risk adjusted returns. In the last decade, the financial economics literature has discussed a large number of alternative performance measures; see the surveys by Aftalion and Poncet (2003), Le Sourd (2007), Bacon (2008), Cogneau and Hubner (2009a,b) and Caporin et al. (2014). The available performance measures can be classified into four large families, as suggested by Caporin et al. (2014), to highlight and understand their differences: relative performance measures (rewards per unit of risk), absolute performance measures (risk-adjusted measures referred to a benchmark or to a set of risk factors), measures derived from utility functions and measures expressed as functions of return distribution features. It is also important to note that performance measures belonging to the same class are heterogeneous since they can be based on different quantities (such as utility functions, moments, partial moments or quantiles) or different information sets (different selections of risk factors). Furthermore, if performance measures are used to order assets (as equity screening rules), the ranks they produce for a common set of assets might be sensibly different; see Caporin and Lisi (2011). The last finding confirms that alternative measures have different views over assets, and the construction of an 'optimal' equity screening tool should take those different viewpoints into account.[2]

Differently, within a performance evaluation framework, several authors have considered the problem of determining the optimal portfolio weights by maximizing different performance measures. They aim at finding the 'best' performance measure; see for example Farinelli et al. (2008, 2009), among others. The outcomes of these studies are not completely conclusive, since different performance measures provide superior results over different samples and different assets.

A possible solution to the above-mentioned limitations is the construction of a composite performance index to be used within an equity screening program, or to guide the allocation of a portfolios without taking into account a single performance measure. To our best knowledge, Hwang and Salmon (2003) is the first and unique paper proposing a combination of performance measures. The authors propose the construction of a combined index by resorting to a copula function. However, given a number of issues, including the need of recovering by simulation the performance measure densities, they do not provide an empirical analysis supporting their proposal. We follow the spirit of Hwang and Salmon (2003), and contribute to this strand of the quantitative finance literature by introducing a new approach for the construction of a composite performance index. Differently from the cited paper, our approach is computationally feasible, thanks to the adoption of a linear combination criterion. Moreover, we highlight a number of computational and implementation issues, suggesting methods that allow overcoming most of them. Finally, we provide an extensive empirical example.

The combination criterion we propose, called Backward/Forward, follows the general idea of linearly combining existing performance measures with positive weights. These weights are determined by means of an optimisation step. The underlying criterion function explicitly takes into account the risk-return trade-off evaluated on a historical and rolling basis. By construction, and due to the rolling window evaluation approach,

[2] We do not consider performance measures based on portfolio holdings as their computation requires access to detailed portfolio composition over time; see Wermers (2006).

the Backward/Forward method provides performance combination weights that can vary over time, thus allowing for changes in preferences across performance measures. Backward/Forward is implicitly robust to dynamic features of returns densities, as these will affect the evaluation of performance measures that are the inputs of our screening algorithm. The final product of the linear combination of performance measures is a composite performance index, which can then be used to generate asset screens.

Apart from introducing the Backward/Forward composite index, we discuss several implementation issues that further detail and clarify the methodology. These include the selection of performance measures, their evaluation and the optimisation of the objective function with respect to the combination weights. All those elements have a relevant role in the evaluation of the composite index and clearly illustrate the flexibility and the features of the proposed approach.

The Backward/Forward composite index is determined within a pre-specified equally weighted (1/N) asset allocation scheme. Such allocation choice is only a tool simplifying the identification of the composite performance index. The use of an equally weighted strategy might be questionable, but represents a reasonable compromise. In fact, the 1/N strategy, also known as Talmud strategy (Duchin and Levy, 2009), assigns an equal weight to each asset in the portfolio. Several studies in the literature show that the 1/N strategy outperforms in-sample and in particular, out-of-sample the Markovitz rule (Jobson and Korkie, 1980; Michaud and Michaud, 2008; Duchin and Levy, 2009; De Miguel et al., 2009). The reason for this essentially relates to the estimation errors on the expected returns and covariance matrix which lead to poor estimates. Consequently, not only the naive 1/N investment rule is at least statistically equivalent to the Markowitz optimised portfolio but also other optimising models presented in the literature do not consistently and statistically provide a Sharpe ratio (or a certain equivalent return) higher than the 1/N portfolio (De Miguel et al., 2009). Recently, Tu and Zhou (2012) propose a combination of the 1/N rule with four sophisticated strategies including the Markowitz rule and empirical findings on this theory-based combination show an improvement in terms of performances with respect to the naive 1/N rule. Such an extension goes beyond the purpose of this paper but it can be easily implemented as a further step.

We present an empirical application that illustrates the use of the Backward/Forward index in portfolio allocation. We show how the Backward/Forward composite performance index can be used as an equity asset allocation screener.

The remainder of the paper proceeds as follows. In Section 2 we define the investment objective and introduce the Backward/Forward composite index. In Section 3, we discuss several implementation aspects. Section 4 contains the empirical application on, and Section 5 concludes.

2. The investment problem and the objective function

Our main purpose is to deal with the security selection problem faced by an investor (a portfolio manager), who is willing to allocate her portfolio over a subset of assets included in her investment universe, and wants to select them by using a combination of different performance measures. The selection takes place by means of what we call an equity screening tool, also referred to as preselection in the large scale portfolio literature; see, among others, Ortobelli et al. (2011).

We assume the investor follows a one–step allocation rule and thus she chooses at time t assets to form the portfolio with an investment horizon of one period, ending at $t+1$. To be consistent, the equity screening is based on a criterion function depending on a set of performance measures evaluated using the information available at time t. Therefore, given the information set at time t, the investor first determines the performance measures and then computes the composite performance index. This index is used as an equity screening tool, and helps the investor to identify the most interesting assets - those with higher composite performance index value. The investor computes the performance measures starting from asset returns and asset-related information. We assume that the asset returns densities have finite moments allowing the computation of most performance measures. Finally, the selected assets are introduced in the portfolio, with weights to be determined by the investor. We stress we do not make a distributional assumption for assets returns for a number of reasons. First, because asset returns might be characterized by different densities that are likely to be non-Gaussian. Second, the choice of a specific distribution is limitative, opposite to the empirical evidence, and could lead to the identification of an optimal performance measure. This is the case for Gaussian returns, agents with negative exponential utility and the Sharpe ratio.

We underline how our procedure does not consider the allocation problem, but focuses on the asset screening, with the final purpose to construct a ranking of assets in the investment universe, able to identify a subset which might be considered optimal for the subsequent construction of the portfolio. Our approach might be of interest for a portfolio investment process where a small number of analysts might start from a very large investment universe and selects a-priori, by means of a screening algorithm, the assets to be later analysed from a fundamental perspective. Alternatively, again when dealing with a very large investment universe, the application of modern portfolio theory, based on the evaluation of the assets covariance, might become complex due to the lack of the necessary degrees of freedom required to estimate the whole covariance matrix. One might resort to a variety of computational and statistical techniques to overcome the problem, but those will be anyway subject to model errors and estimation errors. A screening algorithm might thus represent an alternative possibility. Finally, a screening algorithm might also be of interest within a Core-Satellite allocation strategy and could represent the rational for the selection of the Satellite part of the portfolio.

We thus assume that the investor includes in her portfolio M assets chosen from a larger group containing N assets. Note that $M << N$ in order to avoid excessive transaction and rebalancing costs, also if M should not be too small, otherwise diversification benefits will tend to vanish. In this study, we fix $M = 25$, 50, or 100. Those values are reasonable in small and medium-sized managed portfolios, and will allow us to verify if changes in the number of assets will provide relevant variations in the portfolio turnover and, as a consequence, on rebalancing costs.

Our main contribution is the peculiar screening rule we propose, the Backward/Forward index, which is based on an optimised linear convex combination of performance measures. Our claim is that the combination will take advantage of different views on the assets, or, similarly, of different information, including the asset returns density, the relationship between asset returns and risk factors and the use of alternative utility functions. According, an equity screening procedure based on our composite index should be more efficient than a screening rule based, for instance, on a single performance measure. The intuition behind such a claim stems from empirical evidences on asset returns. In fact, the Sharpe ratio is an optimal measure under a normality assumption. However, when return densities deviate from normality, higher order moments, partial moments or quantiles may have additional informative content, see for instance Farinelli and Tibiletti (2008).

Given the choice of M and the investment universe of N assets, the objective of the investor is to select the M assets to be included in her portfolio according to an optimality criterion. The optimality criterion leading to the assets choice is thus crucial, and is described in the following. We stress here that it must be distinguished from the optimality criterion chosen by the investor to allocate her portfolio over the M assets and it is based on a combination of performance measures.

Let us first introduce some notations. We define for each asset j at time t a composite performance index, $CI_{j,t}$, which is a function of Q performance indices $p_{i,j,t}$, where $i = 1, 2, ...Q$, and $j = 1, 2, ...N$. Note that this index is computed using the information set up to time t, I^t, and is used to allocate the portfolio in time t with investment horizon $t + 1$. We impose the simplifying assumption that the set of performance measures is fixed and known a-priori: the value of Q is fixed over time, and the Q performance measures used in the combination do not change over time.

We suggest the following composite index for asset j at time t based on a linear combination of the Q performance measures:

$$CI_{j,t}\left(w_1, w_2, ..., w_Q\right) = \sum_{i=1}^{Q} w_i p_{i,j,t}, \quad j = 1, 2, ..., N, \tag{1}$$

$$w_i \geq 0, \quad i = 1, 2, ..., Q, \qquad \sum_{i=1}^{Q} w_i = 1,$$

where we impose weights to be positive and to sum up to one. Note that the weights are the same for all assets and are time-invariant. To simplify the notation, the composite index appears as a function of only the weights while in reality it is also a function of the performance measures. Nevertheless, given that the purpose of the analysis is to estimate the combination weights given a collection of performance measures, we suppressed the latter from the compact representation of the composite index.

If the composite index is known, the best M assets are those with the M highest values of $CI_{j,t}$. Those can be later used to allocate a portfolio across the M assets following a given allocation strategy/approach. However, the performance combination weights $\mathbf{w} = \{w_1, w_2, ...w_Q\}$ have to be estimated and we propose to determine them by maximizing the following criterion function:

$$\max_{\mathbf{w}} f\left(\mathbf{w}\right) = \frac{1}{m} \sum_{l=t-m+1}^{t} r_{p,l} - \lambda \frac{1}{m} \sum_{l=t-m+1}^{t} \left(r_{p,l} - \mu_p\right)^2, \tag{2}$$

$$r_{p,l} = \frac{1}{M} \sum_{j \in \mathcal{A}_t(\mathbf{w})} r_{j,l}, \tag{3}$$

$$\mathcal{A}_t(\mathbf{w}) = \left\{ j \in \Omega | CI_{j,t}(\mathbf{w}) \geq \widetilde{CI}_{M,t}(\mathbf{w}) \right\}, \tag{4}$$

where Ω is the set of all assets, $\widetilde{CI}_{k,t}(\mathbf{w})$, $k = 1, 2, \ldots, N$ is the sequence of ordered values of the composite index, $\widetilde{CI}_{M,t}(\mathbf{w})$ is the M-th largest value of the composite index, $\mathcal{A}_t(\mathbf{w})$ is the set of the M assets with the highest score of the $CI_{j,t}(\mathbf{w})$ index (note this set depends on the choice of the weights' vector); $r_{p,l}$ is the time l return of the equally weighted portfolio over the assets included in $\mathcal{A}_t(\mathbf{w})$; and μ_p is a calibrated return level. Note that the criterion function is optimised with respect to the weights by means of a numerical algorithm, as discussed in a following section. Moreover, the composite index, which is a function of the weights, is also included in the criterion function, whose maximization follows an iterative approach (see Section 3). The Backward/Forward composite index is obtained from equation (1) with weight recovered from the maximization of (2). Therefore, the optimal $CI_{j,t}$ is a by-product of the maximization step (as the index is computed for the evaluation of the criterion function). The Backward/Forward composite index can then be used as a screener to rank assets, and thus lead to a Backward/Forward equity screening rule.

The composite index has weights constant across assets. We are aware of the strength and relevance of such an assumption, which we motivate in two ways. First, the estimation of asset-specific combination weights would sensibly increase the computational complexity, making the evaluation of the criterion function extremely complex and time consuming. Second, the change in the combination weights across assets would make difficult the cross-sectional comparison of the composite index. In fact, the composite index would assigning different weights to different measures, leading to an increase in the heterogeneity across assets.

It is easily seen that the first term of the criterion function is the average return of the portfolio over the last m observations and the second term is similar to a risk measure, weighted by a risk aversion coefficient λ. The risk measure depends on the choice of μ_p, which we set either equal to the average portfolio return, thus making the second term equivalent to the portfolio variance; alternatively, we fix μ_p equal to the average return of a benchmark over the last m observations, making the second term equivalent to a variance tracking error. Finally, we suggest to set the risk aversion coefficient between 1 and 50, mimicking the standard choices in the mean-variance framework.

The overall criterion function is thus similar to a mean-variance utility function. However, (2) is not optimised with respect to the portfolio weights, which are fixed, but with respect to the performance measure weights. The intuition behind this criterion function is that we are determining the weights which would have maximized, using up-to-date information, the difference between the return and the risk of the allocated portfolio (where the risk is weighed by a risk aversion coefficient). The risk is either monitored in absolute terms, by using the portfolio variance, or in relative terms, by comparing the portfolio returns to those of a benchmark index. This second option is a reasonable choice if the investor is an investment manager. The criterion function is thus a backward evaluated mean-variance function which is used to forward allocate the portfolio with $1/M$ weights. We stress that the use of an equally-weighted portfolio allocation at this stage is just a device to estimate the performance combination weights and is not imposing that an equally weighted strategy has to be used to allocate the portfolio across the M assets with highest composite index.

The proposed approach to define the Backward/Forward composite performance index entails a number of implicit assumptions. From a statistical point of view, the construction of performance measures at the single asset level implies a focus on the marginal distributions of each asset included in the analysis. According, the composite index in (1) is an equity screening tool since it does not provide optimal asset allocation. In fact, a relevant aspect is not taken directly into account, i.e. the correlation across assets. Dependence among assets has only an implicit role in the portfolio return and risk in (2) and (3), but it is not an explicit element considered in the criterion function. Anyhow given that (2) penalises excessive risks and that the portfolio is equally weighted, the effect of asset correlation is partially sterilised. It might be possible that highly correlated assets with relatively good performances are included in the equally weighted portfolio, thus reducing the diversification benefits, and for this reason, we suggest the use of (1) within an investment process, but not directly as an asset allocation tool.

A second aspect not directly covered by the Backward/Forward composite index is the dependence between performance measures. To reduce the potential negative impact of highly correlated performance measures (which would limit the benefits of the composite index), we suggest to include performance measures with low rank correlation in (1), following Eling and Shumacher (2007), Eling (2008), Eling et al. (2011) and Caporin and Lisi (2011).

The function $f(\mathbf{w})$ includes fixed performance weights, but when we estimate the weights over rolling samples the performance combination weights change. In fact, they update with respect to the changing relevance of the underlying performance measures. We first relate such a feature to the evidence provided in Caporin and Lisi (2011) that show time-variation in the rank correlation across performance measures, suggesting that their informative content is not stable over time. The change in the performance measure relevance is also associated with a change over time of the asset return densities, or equivalently, of their moments and quantiles. In fact, if these elements vary over time, performance measurements vary over time and their views over competing assets change over time.

To capture potential changes in the asset returns densities, and in particular on their moments or quantiles, we consider a rolling evaluation of the performance measures, as in Biglova et al. (2004). In that way we can capture time-varying features of the asset return moments and quantiles, which are the constituents of most performance measures. If the conditional values of the cited quantities were time-varying, a rolling methodology would take that into account. On the contrary, if they were time-invariant, we would only have an effect coming from sampling errors, which could be controlled by changing the size of the rolling window. We are thus implicitly assuming that the sample estimators of moments, quantiles, and quantities obtained as transformations of sample data (such as utility functions) are consistent and unbiased estimators of the corresponding conditional quantities. The time-change in return densities might also be controlled by working on relatively low data frequencies, such as monthly.

3. Implementation issues

In the following, we discuss a number of issues that should be considered in the implementation of the Backward/Forward composite performance index. These elements clarify first the definition of two elements which are pre-requisites of the equity screening methodology: the investment universe and the benchmark. Later, we highlight the flexibility of the equity screening approach, widening the concept of performance measures, which might include other indicators such as the asset market value. We then move to the evaluation of performance measures and to the possibility of standardizing their values. Finally, we deal with the optimisation of the criterion function $f(\mathbf{w})$, suggesting the use of genetic algorithms. Given its relevance, a dedicated subsection discusses the latter aspect.

Investment universe. When the allocation is performed over time for different t, our approach does not require the portfolio cardinality M and the number of assets N to be fixed. These two quantities could change over time, thus allowing changes in the universe of available assets (companies may die, or might be involved in mergers and acquisitions, or new companies can be included in the investment universe) as well as changes in the portfolio strategies (increasing/decreasing the diversification).

Benchmark. If a benchmark is included in the criterion function, its choice also has to be carefully considered. In fact, the benchmark has to be chosen such that it is representative of the N assets included in the analysis. This is required to evaluate an appropriate tracking error. The benchmark and the assets should thus include the effect of dead companies. In fact, the use of a specific equity market index as benchmark, together with N currently traded assets exposes the equity screening to a survivorship bias. An alternative approach that overcomes the bias and excludes dead companies is to create a synthetic benchmark using a set of N selected assets and their market values. We follow this last approach for simplicity.

Definition of performance measure. Our approach is flexible, and the term "performance measures" could be interpreted in a wider sense. In fact, we can optimally combine a set of indicators associated with listed companies. These indicators could be performance measures, but could also be liquidity measures, technical analysis indicators or company-specific variables (revenues, employees, balance sheet ratios). From a different viewpoint, the Backward/Forward composite index might be separately evaluated for a set of risk measures, as well as for a set of reward measures. Our approach is thus potentially close to a multi-criteria methodology, similar in some respects, to Ballestro et al. (2007).

Companies' market value. Liquidity is one of the possible market constraints that could affect our equity screening strategy. In fact, the selected assets can differ in terms of market value and thus liquidity, making the allocation of the optimal portfolio problematic. In extreme cases, Backward/Forward composite index, and the equity screen it provides, could suggest to invest in companies with small market value, whose shares might be characterised by limited liquidity. As a result, the implementation of the portfolio could be characterised by large costs (transaction costs as well as large deviations in the price due to the liquidity problems or the impossibility of creating the portfolio because some trades could not be executed in the market due to the absence of a counterpart). To mitigate this aspect, and thus force the optimal portfolio to invest in small caps only if their performances are really relevant, we suggest to introduce the market value as a further performance measure. This would capture the liquidity effect: higher the market value, higher the liquidity. Clearly, other measures of stock liquidity can also be considered.

Evaluation of performance measures. The performance indicators chosen to build the Backward/Forward composite index are generally computed on a given sample. In order to follow the evolution over time of the asset return densities, we suggest to evaluate performance measures over a rolling window of m observations. The value of m depends on the time frequency of observations and on the total sample length; some examples could be 60 or more months, about 50 weeks, or 40 or more days. In general terms, we suggest to use between 40 and 60 observations to avoid excessive volatility in performance measure values that might induces relevant changes in the construction of the composite index, in the assets included in $\mathcal{A}_t(\mathbf{w})$ and consequently, a large turnover in the portfolio. On the contrary, longer samples could significantly smooth performance measures sequences, leading to a very low turnover, but would not capture local (medium period) changes in performance measure relative rankings. With respect to the data frequency, we suggest the use of monthly data. Higher frequencies will induce relevant and frequent changes on the portfolio combination weights and on the asset rankings.

Standardisation. Given a list of Q performance measures, our final purpose is the construction of a composite index. However, we must recognise that different performance measures could have different ranges, thus making their combination dependent on the scale of the chosen performance measures. For this reason, we suggest to consider standardised performance measures as inputs of the composite index. Let $p_{i,j,t}$ be a given performance measure, we suggest to compute the composite index $CI_{j,t}$ using the following quantities as inputs:

$$\bar{p}_{i,j,t} = \frac{p_{i,j,t} - \min \left\{p_{i,j,t}\right\}_{j=1}^{N}}{\max \left\{p_{i,j,t}\right\}_{j=1}^{N} - \min \left\{p_{i,j,t}\right\}_{j=1}^{N}}. \tag{5}$$

Such a standardisation makes the performance indices to vary between 0 and 1, thus avoiding the scale effect, and ideally putting all performance measures on the same playing field.

3.1. Objective function optimisation

The determination of the Backward/Forward composite index requires the solution of a non-trivial optimisation problem. For each point in time, the evaluation of $f(\mathbf{w})$ in (2) conditional to a vector of weights $\mathbf{w}$ requires the following steps:

- Evaluate the performance measures $p_{i,j,t}$;
- Compute the standardized performance measures $\bar{p}_{i,j,t}$;
- Determine for each asset the composite index $CI_{j,t}(\mathbf{w}) = \sum_{i=1}^{Q} w_i \bar{p}_{i,j,t}$;
- Identify the set $\mathcal{A}_t(\mathbf{w})$;
- Obtain the ex-post return of the allocation $r_{p,l} = \frac{1}{M} \sum_{j \in \mathcal{A}_t(\mathbf{w})} r_{j,l}$ and the objective function $f(\mathbf{w})$.

The criterion function $f(\mathbf{w})$ is a non-linear and non-differentiable function of performance measure weights $\mathbf{w}$. In fact, these enter only in the construction of the set $\mathcal{A}_t(\mathbf{w})$ that contains the assets with the highest values of the index $CI_{j,t}(\mathbf{w})$. Furthermore, different values of the weights could provide the same set of 'best' assets, thus making the optimisation of $f(\mathbf{w})$ computationally demanding.We first provide a graphical example to clarify the computational problems that might arise in the optimisation of the objective function detailed in Section 2. Let us assume we have three performance measures and thus two weights to be estimated (the third one is obtained through the constraint). We report in Figure (1) the value of the criterion function for all possible weight combinations. Notably, the surface has many flat areas and thus local maxima.

Consequently, optimisation methods based on derivatives of the function $f(\mathbf{w})$ do not guarantee the identification of the global optima.

To enforce the identification of the global optima, we thus suggest the use of genetic algorithms, in particular the Differential Evolution algorithm (DE) developed by Storn and Price (1997), which is a population-based optimiser. One particular feature of the DE algorithm is that it encodes every type of parameter as floating-point numbers. As reported in Price et al. (2005), this provides different advantages with respect to the bit–flipping algorithm of traditional Genetic Algorithms implementations. For instance, it induces better scales on large problems and a faster convergence. In turn, this implies a reduced computational effort.

In the DE, the starting point is determined by sampling the objective function at different random initial points of the parameters, i.e. the Q performance combination weights. Moreover, each parameter in our objective function is bounded in [0,1] since it represents the coefficient of a convex combination. Initial points (also called population members in the DE framework) are thus sampled from a $Q-$dimensional domain. In the empirical application we used the DE optimisation package for Matlab developed by Markus Buehren and available in MatlabCentral. This package is based on the code of Storn and Price (1997). Note that a single evaluation of the objective function took on average 15 seconds using an Intel 3.4 GHZ Intel Core 7 processor machine. The execution time can however be reduced using the parallel processing on multiple cores. We set the number of population members as suggested by the author equal to 10 by the number of parameters, thus $10Q$. The algorithm stops when one of the following conditions is met: the maximum number of iterations is reached (we set the maximum at 100); the function evaluation lasts for a maximum of 60 seconds and all possible combination of parameters have been tested. For a detailed description of the algorithm, see Storn and Price (1997), Maringer (2005) and Price et al. (2004). For applications of the DE algorithm in finance, see Maringer (2005), Gilli et al. (2008), Hagstromer and Binner (2009), Krink et al. (2009), Krink and Paterlini (2011) and Gilli and Schumann (2012), among others. For other applications of genetic algorithms in finance see Mohr et al. (2013), among others.

To further support our suggestion, we perform a comparison of different solvers (global and local optimisers) when facing an optimisation that admits several local optima. We consider the following solvers: the differential evolution algorithm (DE), a genetic algorithm (GA) and interior point algorithm (IP).[3]

We start by simulating 1000 different starting values for the parameters of interest (the performance measure weights), and feed with them the solvers in the estimation of the objective function on the same data. We expect that global optimisers, DE and GA, will identify the global optima with a much larger frequency compared to the IP approach. Figure 2 contains the frequency histogram of the maximum objective function values across the 1000 replications. The distribution for IP (white) highlights the presence of several local optima, while the DE (dark grey) and GA (light grey) distributions are centered around the same value, but with a different dispersion. This clearly emerges from (1). The averages of the objective function values for 1000 optimisations are quite similar for the DE and GA, but DE shows a lower standard deviation. On the other hand, the IP algorithm provides the highest dispersion, reflecting the shortfall of the gradient optimisation when dealing with multiple local optima (note also that this corresponds most likely to lower values of the objective function). Overall, this small simulation experiment confirms the appropriateness of the DE algorithm for our optimisation problem.

4. Equity screening with composite indices on the US market

We consider the US stock Market for an empirical assessment of the Backward/Forward composite index within an asset allocation framework. The composite index is applied as an equity screening rule, and our purpose is to verify its advantages in terms of portfolio returns. We first select the performance measures to take into account, and later describe the data. Moreover, we introduce alternative naive equity screening rules that are compared to our proposal. The empirical results are reported in a fourth subsection.

4.1. Selected performance measures

The results of our approach clearly depend on the choice of performance measures combined in the index $CI_{j,t}(\mathbf{w})$. The following surveys might be used to select among the large set of performance measures proposed in the financial economics literature: Aftalion and Poncet (2003), Le Sourd (2007), Bacon (2008), Cogneau and Hubner (2009a,b) and Caporin et al. (2014). In addition, Eling and Shumacher (2007), Eling

[3]For details on the solvers we consider, see Storn and Price (1997) for Differential Evolution, Goldberg (2000) for Genetic Algorithm, and Bonnans et al. (2006) for the interior point algorithm.

(2008), Eling et al. (2011) and Caporin and Lisi (2011) report comparisons among alternative performance measures. The measures we consider have been selected in order to include well-known quantities, like the Sharpe, Sortino, Treynor, and Appraisal ratio, as well as measures based on partial moments, quantiles and drawdowns, which are not much common. Our purpose is to provide an empirical application showing the benefits associated with the combination of different measures. In the application, we evaluate the performance measures at time $t-1$ and use the equity screening outcome for selecting assets to be hold up to time t. The on-line appendix reports a detailed list of performance measures, including also their equations, while here we simply list the measures we use in the empirical analyses together with the symbol used to identify them. Our selection includes traditional performance measures, the Sharpe ratio (Sh) of Sharpe (1966 and 1994), the expected return over the Mean Absolute Deviation (ERMAD) introduced by Konno (1990) and Konno and Yamazaki (1991), the Appraisal ratio (AR), the Treynor index (Treynor, 1965), or Risk Adjusted Return (RaR), and the M2 index by Modigliani and Modigliani (1997). Those measures provide risk-adjusted returns differing in the way they measure the asset risk, ranging from volatility, to systematic and idiosyncratic risk, up to the benchmark risk. We also include measures based on the drawdowns: the Calmar ratio (CR) of Young (1991), the Sterling ratio (SR) of Kestner (1996), and the ratio (BR) by Burke (1994). These three measures provide risk-adjusted returns with a focus on the extreme risks as monitored by the drawdowns. We then include measures based on partial moments: the Sortino ratio (Sr) by Sortino and Van der Meer (1991), and the Kappa 3 (K3) measures by Kaplan and Knowles (2004). This third group of measures modifies both the return and risk measures extrapolating these elements form the returns empirical density, and using the entire density support. Finally, we consider a measure based on quantiles, the expected return over absolute Value-at-Risk (VR) of Dowd (2000). This last measure considers again the returns distribution to recover a risk measure, but focuses only on the lower tail. In the empirical analyses, we consider a rolling evaluation of the performance measures over a window of $m = 60$ months. As mentioned in the previous section, the Market Value (MV) of each company will be included as an additional performance measure to penalise smaller companies.

4.2. Dataset description and benchmark construction

Our dataset is based on the constituents of the S&P Composite 1500 (as of February 22nd, 2012). The time series were downloaded from Datastream at a monthly frequency from the 31st of January 1990, to the 31st of January 2012, for a total of 265 observations. We also recovered a proxy of the risk free asset, the JP Morgan 1 Month Cash bond index. To cope with survivorship bias, we restricted the dataset to a collection of assets constantly available in the analysed sample. Following this criterion, we restricted our attention to 695 assets.[4]

The S&P1500 market index cannot be used as a benchmark to evaluate the performances of the equity screening approach based on the Backward/Forward composite index. In fact, our selected sample does not include a relevant part of the assets composing the S&P1500. Moreover, the S&P1500 composition changes over time.[5] Therefore, we build a benchmark that is coherent with the selected assets. This index corresponds to the value-weighted index composed of the 695 selected assets.[6]

4.3. Portfolio allocation and naive equity screening

We apply our equity screening approach, based on the Backward/Forward composite index, to the selected assets by estimating performance measures on rolling windows of 60 months. Starting from the end of January 1995 (the first month where 60 monthly returns are available), we identify, across the 695 assets, the 50 assets that maximise the criterion function in (2). We used two specifications for the Backward/Forward composite index discussed in Section 2, differing in the risk component on the criterion function (2): the portfolio variance (VO) and the tracking error volatility (TE).[7] Moreover, we make use of two different values for the risk aversion parameter, 1 and 20. The first corresponds to a mild penalisation of the risk, while the second mimics the choices of a more risk-averse investor. Finally, we compare our screening algorithm to a naive equity screening rule based on the Sharpe ratio. Therefore, with a rolling procedure similar to that outlined before, we select the 50 assets that have higher Sharpe ratios.

[4]The list of assets is available upon request.

[5]The time series of the S&P 1500 constituents is not available to us through Datastream.

[6]Additional details on the benchmark are available in the on-line supplementary material.

[7]We set the term μ_p equal to the average portfolio return in the first case, while μ_p is equal to the benchmark return in the second case.

To compare the obtained asset screens, we use them to allocate the wealth of alternative investors, each using a specific equity screening rule. For the portfolio allocation of the 50 selected assets, we opt for an equally weighted portfolio rule where each asset has an equal weight of 2%. As already noticed, equally weighted portfolios have been shown to have performances comparable to, if not better than, optimised portfolios; see De Miguel et al. (2009). Moreover, when optimised portfolios are used, differences among their returns/risk/performances might be due both to the screening rules, and to the estimation error associated with the portfolio weights, and thus the estimation error might hide any discrepancy among screening rules. We compute the monthly realised returns of equally weighted portfolios based on different screening rules. The portfolio composition is modified on a monthly basis, where our criterion function (2) is optimised each month, and Sharpe ratios are estimated every month. At the end, we have a total of 205 monthly portfolio returns. These returns are compared by means of: standard descriptive analyses, including the computation of some risk measures; a horse-race over the range February 1995 to January 2012; the weights associated with the different performance measures; the associated turnover of portfolios.

We define the turnover as the difference in the portfolio's composition in the period t with respect to the period $t-1$. This indicator highlights the variability in selection in terms of assets for a given strategy,

$$TO_i = 1 - \frac{\mid I_{i,t} \cap I_{i,t-1} \mid}{M} \in [0, 1], \tag{6}$$

where $I_{i,t}$, I_{t-1} are the set of selected assets in the two period for the strategy i, $\mid \cdot \mid$ indicates the cardinality of the set, and M the number of selected assets. Moreover, we compare the performances also to those of the benchmark, computed as described in the previous subsection.

4.4. Performance results

Table (2) includes the descriptive analysis of the portfolios, while Figure (3) shows the cumulated returns from 1995 to 2012. Compared to the benchmark, all equity screening-based portfolios provide higher cumulated returns. If we consider an investor with an initial wealth of 1, the portfolio with the highest cumulative return (8.53) is given by the criterion function which considers the tracking error volatility and a risk aversion coefficient equal to 1. The second highest portfolio in terms of cumulative returns (6.69) is based on the criterion function which depends on the portfolio variance. The risk-aversion coefficient has a relevant impact; in fact, the TE and VO portfolios with risk aversion set to 20 are less profitable than the Sharpe-based portfolio. Respectively, (3.26) and (3.68) compared with the Sharpe-based (5.36). The result is even stronger if we analyse the turnover, which is sensibly higher for higher values of the risk aversion. In the TE strategy, it changes from 0.1822 to 0.2457. Similarly, it increases from 0.1862 to 0.2208 in the VO strategy. Figure (4) reports the corresponding boxplots of the distribution returns for the various strategies. In the considered case, the strategies with and without the lower market value bound are very similar to each other. As we might expect, the strategies with a higher risk aversion provides a lower dispersion compared to the others. In particular, the VO strategy with the lower bound in the market value is the best in terms of volatility.

Comparing the Sharpe index of the portfolios, the Sharpe-based portfolio seems to be the preferred choice (0.2288), except for the VO case with risk aversion coefficient set to 20 which provides a Sharpe ratio equal to 0.2314. Such a result is a consequence of the criterion used for portfolio construction, and might be expected.

In terms of risk measures, we observe that all portfolios based on screening rules are more risky than the benchmark, which gives an annualized volatility equal to 0.1538. However, we stress that all screening-based portfolios have not been optimised to reduce the risk, but are simply based on an equally weighted allocation scheme. As a result, risk reductions might be achieved by optimising portfolio weights across the selected assets, or by generalizing the criterion function and introducing weights in equation (3). The latter choice would require the joint optimisation of the criterion function in (2) with respect to both the portfolio weights and the performance combination weights. Such a generalization is left to future research as we focus here on the introduction of the composite performance index. Finally, we emphasise that risk measures decrease with increasing risk aversion, as expected. Moreover, for a risk aversion coefficient equal to 20, risk measures they are better than the benchmark and also preferred to those of the Sharpe-based portfolio, with the exception of the TE case where the annualized volatility is higher than that of the benchmark (for the other risk measures the strategy provides better results than the benchmark). As an example, by increasing the risk aversion, the VaR at 5% in the TE strategy changes from 0.0935 to 0.0668 while in the VO strategy from 0.0872 to 0.0573.

Overall, the introduction of screening rules provides higher returns than the benchmark, with a preference for the Backward/Forward algorithm compared to simpler screening based only on the Sharpe ratio. Risk measures are different across screening strategies, but this is a consequence of the portfolio construction, that is not optimised. The turnover induced by the screening rules is influenced by the degree of risk aversion, and becomes higher and more volatile with increasing risk aversion; see Figure (5). The Figure shows that the turnover induced by a Sharpe-based screening is oscillating between 10% and 30% on a monthly basis. Similar values are provided by the TE screening with low risk aversion in a large part of the sample. Deviations are observed during periods of high volatility (from 2007) and with higher values of the risk aversion. In these two cases, the turnover induced by the TE screening is higher than the turnover of the Sharpe based screening, and it reaches values close to 50% with low risk aversion and up to 90% with high risk aversion.

In the Backward/Forward composite index, the weights assigned to the different performance measures have a relevant role; moreover, they change over time, and react to the different features of returns time series. Table (3) includes the descriptive statistics for the weights in the tracking error- and volatility-based screening.[8] We first point out that our optimisation algorithm generally assigns a very small weight - which is close to zero - to the Sharpe ratio, independent of form of the criterion function and of the risk aversion coefficient. Both the tracking error and portfolio volatility objective functions provide similar performance measure weights when the risk aversion coefficient is set to 1; in particular, we observe that the Modigliani-Modigliani index receives the largest weight. Respectively, 0.56 in the TE strategy and 0.44 in the VO strategy. Other performance measures receiving a high weight are the Appraisal Ratio and the Excess Return over Mean Absolute Deviation. The other measures are characterised by very small average weights and limited standard deviations, which signal that they receive a relevant weight only occasionally as shown in Table (3) . When the risk aversion coefficient is increased to 20, the difference between the two forms of the criterion function leads to different average weights assigned to the performance measures. When we focus on the tracking error-based function, the Market Value, the Appraisal Ratio and the Excess Return over Mean Absolute Deviation receive a weight larger than 10%.

[8]A graphical example on the time-varying evolution of weights is included in the on-line appendix.

In contrast, in the second implementation based on portfolio volatility, the Burke and RAR measures increase over 10%, while the Market Value falls below 10%. Even for a large risk aversion level, the performance measures with small average weight have a large standard deviation, thus confirming their limited relevance. Results support thus our expectation of variability in the informative content of performance measures, and might also be seen as a confirmation of the potential interest in performance measures able to go beyond the Sharpe ratio.

A important element to emphasise is the limited weight assigned to the Market Value. As a consequence, the selected assets might be characterised by small market value and thus small liquidity, possibly creating difficulties in the implementation of portfolios based on these assets. This is confirmed by Figure (6) in which we see a sharp decrease in the average market value of the selected companies in the second half of the sample. Such a behavior is common across the different implementations of the screening algorithm. To force the impact of the Market Value, we run a second set of evaluations where, in the solution of (2), we constrain its weight with a lower bound of 10%.

Table (2) includes the descriptive statistics and shows the impact of the Market Value constraint in terms of cumulated returns, risk measures, and Sharpe ratios. Overall, to impose a minimum relevance to the companies' market value leads to a slight risk reduction: Value-at-Risk, Expected Shortfall, returns volatility and range all improve. However, the total and average returns, and the Sharpe ratio decrease, except in the case with tracking error objective function and high level of risk aversion. In addition, the turnover shows a decrease, which is larger for the cases where the risk aversion is set to 20. It decreases from 0.2457 to 0.1972 in the TE and from 0.2208 to 0.2013 in the VO strategy.

To investigate the degree of similarity among the strategies (and the associated equity screens), we compute a concordance index which highlights the common selection in terms of assets,

$$\texttt{ConcIdx} = \frac{\mid I_i \cap I_j \mid}{M} \in [0, 1], \tag{7}$$

where I_i, I_j are the set of selected assets for the strategy i, j with $i \neq j$, $\mid \cdot \mid$ indicates the cardinality of the set, and M the number of selected assets. In Table (2), we also report the average of the concordance index to compare each strategy with the Sharpe

Portfolio rule. The concordance index is close to 0.6 for both the tracking error volatility and the portfolio variance criteria. Such a result, on the one side highlights the different signals coming from the strategies obtained through the performance measures. On the other side, it is a bit surprising given the small weight assigned to the Sharpe ratio in the composite index. This is a by-product of the correlation across single performance index ranks, which can be controlled by selecting measures as suggested by Caporin and Lisi (2011), but cannot be fully annihilated.

Looking at the Market Value of the selected companies, we note an increase in the second part of the sample compared to the previous cases; see Figure (6). The introduction of a lower bound for the Market Value leads to the selection of equities with an average market value generally higher than the average market value of the benchmark. Weights assigned to the performance measures are thus partially affected by the constraint imposed on MV; see Table (4). However, the sets of the most influential performance measures remain unchanged.

4.5. Robustness Checks

As already mentioned, the number of assets identified by our screening algorithm can be easily modified. The on-line supplementary material contains the Figures and Tables associated with the comments here reported. We first consider the selection of either 25 or 100 assets. The results for these strategies are very similar to the ones with $k = 50$ reported in Table (4). The higher the risk aversion for the investor, the lower the variance of the given returns. By comparing the results across different values of M, we note that screening algorithms always beat the benchmark in terms of cumulated returns but not in terms of risk measures. Nevertheless, we observe a general reduction in the risk measures for increasing M, and an improvement in the Sharpe ratios for $M = 100$. Such a finding depends on the possibility of identifying profitable investment opportunities (that is assets) the performance of which might be changing over time, leading to assets being "above average quality" but not necessarily "top performers". If a small number of assets is included, the selected equities are subject to more frequent changes, as shown by the average turnover (decreasing for increasing M). As a result, when the number of selected assets increases, the performances improve.

Moreover, we observe that the Sharpe ratios of the portfolios based on Backward/Forward equity screens are better than the naive approach in a few cases only, but they are associated with different objective functions: when $M = 25$ the use of the tracking error-based objective function, with a large risk aversion and bounded weight of the Market Value, provides the best results.In contrast, when $M = 100$, the results are slightly better for the objective function using the variance of the selected portfolio and a large risk aversion level, dependent from the presence of a bound on the weight of MV. The case where $M = 50$ is in the middle, with the Sharpe ratios of the naive strategy and our screening rule being very close to each other. Overall, the empirical application on the US stock market shows that the screening algorithm based on the Backward/Forward composite index is able to identify profitable investment opportunities.

In addition, it is pretty clear that the concordance index depends also on the number of selected assets in the market: the higher the M, the higher the concordance index. In fact, the concordance indices for $M = 25$ assets are lower with respect to the ones associated with either 50 or 100 assets. The same effect appears also for the turnover, where the higher is M, the less frequent is the change in the composition of the portfolio.

Finally, looking at Tables (3) and (4) it clearly emerges a weight dominance in the composite index by AR and M2 performance measures. Our approach assigns to $M2$ and AR measures the higher weights in all the different criteria when the risk aversion coefficient is set equal to 1. The result holds even if we change the number of selected assets, M. Conversely, when the investor is more risk adverse (risk aversion coefficient equal to 20), the weights are more re-distributed on the different performance measures. We consider at this purpose the realized portfolio returns based on the two strategies mentioned above. The AR strategy provides a lower dispersion of the returns with respect to the M2 strategy. Furthermore, increasing the number of assets included in the portfolio reduces the volatility of the strategy towards the market index.

In our empirical application, the M2 strategy outperforms all the other strategies in terms of cumulative and annual returns. This result holds for each strategy and for all the M selected assets. The annual volatility for the different M is lower than the volatility of the benchmark. The realized returns of the AR strategy is in line with the other portfolio strategies obtained with the composite index.

It is worth noting the ability of the Backward/Forward methodology to capture the market “momentum” through the best performance measures according to the specified criteria. This can be seen also by looking at concordance indices of the

Backward/Forward-based allocations with respect to the Sharpe, AR and M2 portfolios. If the risk aversion coefficient is set to 1, TE and VO strategies have an high degree of similarity with the $M2$ strategy and the concordance index is on average higher than 0.70 - third and sixth column for no bounded and bounded MV weights in the criterion function, respectively. Alternatively, when the risk aversion coefficient is set to 20, the concordance index among $M2$ and TE/VO strategy falls down. Instead, it remains quite stable for Sharpe (SH) and Appraisal Ratio (AR) in TE and VO criteria for different levels of the risk aversion coefficient.

5. Conclusions

We introduce a new composite index of performance measures, the Backward/Forward index, to be used as a screening algorithm that selects within an investment universe a subset of assets. Such an index linearly combines different performance measures where the combination weights are derived from an optimisation problem that takes into account past performances associated with the "optimal" weights and the subsequent asset ranking. Accordingly, past performances lead to the asset selection for future allocations, suggesting the Backward/Forward equity screening name. We discuss several implementation issues of our screening algorithm and then apply it on the US stock market. Results show advantages of our composite performance index in a simplified asset allocation framework. In fact, by comparing simulated equally weighed portfolio strategies, the proposed composite performance index provides superior results in terms of realised results.

Several aspects of our analysis might be further extended. Within the proposed framework it would be interesting to estimate the optimal number of assets M which might be expressed as a fraction of the total number of assets N, or the optimal weights of the portfolios, going beyond the use of a simplified equally weighted allocation. In addition, several constraints can be added to the criterion function, such as limits on the risk (maximum variances, VaR constraints), maximum transaction costs, turnover constraints, just to cite some possibilities. From a practical point of view, our empirical study might be improved by the introduction of "dead" companies, if a suitable database is available.

Acknowledgements.

Authors acknowledge financial support from the project MIUR PRIN MISURA - Multivariate models for risk assessment. The second author acknowledges financial support from the University of Padova grant CPDA073598. We thank the Editor, two anonymous referees, Francesco Lisi and the participants of the Computational and Financial Econometrics 2009 Workshop held in Cyprus for the comments we received on a previous version of this paper. The usual disclaimers apply.

[1] Aftalion, F., and Poncet, P. (2003). Les techniques de mesure de performance. Paris: Economica.

[2] Bacon, C.R. (2008). Practical Portfolio Performance Measurement and Attribution. Wiley, Chichester.

[3] Ballestro, E., Gunther, M., Pla-Santamaria, D., and Stummer, C. (2007). Portfolio selection under strict uncertainty: A multi-criteria methodology and its application to the Frankfurt and Vienna Stock Exchanges, European Journal of Operational Research, 181, 1476-1487.

[4] Biglova, A., Ortobelli, S., Rachev, S. and Stoyanov, S. (2004), Different approaches to risk estimation in portfolio theory, Journal of Portfolio Management, 31, 103-112.

[5] Bonnans, J. F., Gilbert, J. C., Lemarechal, C., and Sagastizabal, C. A. (2006). Numerical optimization: theoretical and practical aspects. Springer Science and Business Media.

[6] Burke, G. (1994). A Sharper Sharpe Ratio. Futures 23, 56.

[7] Caporin, M., G. Jannin, F. Lisi, and B. Maillet (2014), A survey of the four families of performance measures, Journal of Economic Surveys, 28, 917-942.

[8] Caporin, M., and Lisi, F. (2011), Comparing and Selecting Performance Measures for Ranking Assets, Economics: The Open-Acces, Open-Assessment E-journal 5, 31 pages.

[9] Cogneau, P., and Hübner, G. (2009a), The (more than) 100 ways to measure portfolio performance - part 1: standardized risk-adjusted measures, Journal of Portfolio Measurement, 13, 56-71.

[10] Cogneau, P., and Hübner, G. (2009b). The (more than) 100 Ways to measure portfolio performance - part 2: special measures and comparison, Journal of Portfolio Measurement, 14, 56-69.

[11] Constantinides, G. (1979). A Note on the Suboptimality of Dollar-Cost Averaging as an Investment Policy, Journal of Financial and Quantitative Analysis, 14, 443-450.

[12] De Miguel, V., Garlappi, L., and Uppal, R. (2009). Optimal versus naïve diversification: how inefficient is the 1/N portfolio strategy?, Review of Financial Studies, 22, 1915-1953.

[13] Dowd, K. (2000). Adjusting for Risk: An Improved Sharpe Ratio, International Review of Economics and Finance 9, 209-222.

[14] Duchin, R. and Levy, H. (2009). Markowitz versus the Talmudic portfolio diversification strategies. The Journal of Portfolio Management, 35(2), 71-74.

[15] Eling, M. (2008). Does the measure matters in the mutual fund industry?, Financial Analyst Journal, 64, 54-66.

[16] Eling, M. and Schuhmacher, F. (2007). Does the choice of performance measure influence the evaluation of hedge funds? Journal of Banking and Finance, 31, 2632-2647.

[17] Eling, M., Farinelli, S., Rossello, D., and Tibiletti, L., (2011). One-size or tailor-made performance ratios for ranking hedge funds, Journal of Derivatives and Hedge Funds, 16, 267-277.

[18] Farinelli, S., Ferreira, M., Rossello, D., Thoeny, M. and Tibiletti, L. (2008). Beyond Sharpe ratio: optimal asset allocation using different performance ratios, Journal of Banking and Finance, 32, 2057-2063.

[19] Farinelli, S., Ferreira, M., Rossello, D., Thoeny, M. and Tibiletti, L. (2009). Optimal asset allocation aid system: from "one-size" vs "taylor-made" performance ratio, European Journal of Operational Research, 192, 209-215.

[20] Farinelli, S. and Tibiletti, L. (2008). Sharpe thinking in asset ranking with one-sided measures, European Journal of Operational Research, 185, 1542-1547.

[21] Gilli, M., and Schumann, E. (2012). Heuristic optimisation in financial modelling, Annals of operations research, 193, 129-158.

[22] Gilli, M., Maringer, D., and Winker, P. (2008). Applications of heuristics in finance, in Schlottmann, D., Weinhardt, C., and Schlottmann F., eds., Handbook on information technology in finance, Ch. 26, Springer-Verlag, Berlin, Heidelberg, 2008.

[23] Goldberg, D. E. (2000). Genetic AlgorithmsIn Search, Optimization and Machine Learning, Revised ed.

[24] Grinold, R., and Kahn, R. (1999). Active Portfolio Management: A Quantitative Approach for Producing Superior Returns and Controlling Risk. McGraw-Hill New York.

[25] Hagstromer, B., and Binner, J.M. (2009). Stock portfolio selection with full-scale optimization and differential evolution, Applied Financial Economics, 19, 1559-1571.

[26] Hwang, S. and Salmon, M. (2003). An analysis of performance measures using copulae, in: Knigth, J., and Satchell, S. (eds), Performance measurement in finance: firms, funds and managers, Butterworth-Heinemann Finance, Quantitative Finance Series.

[27] Jobson, J. David, and Bob Korkie. (1980). Estimation for Markowitz efficient portfolios, Journal of the American Statistical Association 75.371, 544-554.

[28] Kaplan, P.D. and Knowles, J.A. (2004). Kappa: A Generalized Downside Risk-adjusted Performance Measure, Journal of Performance Measurement 8, 42-54.

[29] Kestner, L.N. (1996). Getting a Handle on True Performance, Futures 25, 44-47.

[30] Konno, H. (1990). Piecewise Linear Risk Functions and Portfolio Optimization, Journal of the Operations Research Society of Japan 33, 139-156.

[31] Konno, H. and Yamazaki, H. (1991). Mean-Absolute Deviation Portfolio Optimization Model and its Application to Tokyo Stock Market, Management Science 37, 519-531.

[32] Krink, T., Mittnik, S., and Paterlini, S. (2009). Differential evolution and combinatorial search for constrained index tracking, Annals of Operations Research, 172, 153-176.

[33] Krink, T., and Paterlini, S. (2011). Multiobjective optimization using differential evolution for real-world portfolio optimization, Computational Management Science, 8, 157-179.

[34] Le Sourd, V. (2007). Performance Measurement for Traditional Investment, Financial Analysts Journal, 58, 36-52.

[35] Maringer, D. (2005). Portfolio management with heuristic optimization, Advances in Computational Management Science, Volume 8, Springer.

[36] Michaud, Richard O., and Robert O. Michaud (2008). Efficient asset management: a practical guide to stock portfolio optimization and asset allocation. OUP Catalogue.

[37] Modigliani, F. and Modigliani, L. (1997). Risk-adjusted Performance: How to Measure it and Why, Journal of Portfolio Management 23, 45-54.

[38] Mohr, E., Schmidt, G. and Jansen, S. (2013). A Comparative Study of Heuristic Conversion Algorithms, Genetic Programming and Return Predictability on the German Market; In E. Tantar, A.-A. Tantar, P. Bouvry, P. Del Moral, P. Legrand, C.A. Coello Coello, and O. Schtze, (Eds.), EVOLVE - A bridge between Probability, Set Oriented Numerics and Evolutionary Computation, Volume 447 of Studies in Computational Intelligence (SCI), pp. 393-414, Springer.

[39] Ortobelli, S., Angelelli, E., and Toninelli, D., (2011). Set-portfolio selection with the use of market stochastic bounds., Emerging Markets Finance and Trade 47, 5-24

[40] Price, K., Storn, R., and Lampinen, J. (2004). Differential evolution - A practical to global optimization. Springer Berlin.

[41] Sharpe, W.F. (1966). Mutual Fund Performance, Journal of Business, 39, 119-138.

[42] Sharpe, W.F. (1994). The Sharpe Ratio, The Journal of Portfolio Management, 20, 49-59.

[43] Sortino, F.A. and van der Meer, R. (1991). Downside Risk. Capturing What's at Stake in Investments Situations, Journal of Portfolio Management 17, 27-31.

[44] Storn, R., and Price, K. (1997). Differential Evolution: a simple and efficient adaptive scheme for global optimization over continuous spaces, Journal of Global Optimization, 11, 341-359.

[45] Treynor, J.L. (1965). How to Rate Management of Invested Funds, Harvard Business Review 43, 63-75.

[46] Tu, J., and Zhou, G. (2011). Markowitz meets Talmud: a combination of sophisticated and naive diversification strategies, Journal of Financial Economics, 99, 204-215.

[47] Young, T.W. (1991). Calmar Ratio: A Smoother Tool, Futures 20, 40.

[48] Wermers, R. (2006). Performance evaluation with portfolio holdings information, North American Journal of Economics and Finance, 17, 207-230.

Table 1: Results for the optimisation for 1000 simulated initial values considering the differential evolution algorithm (DE), genetic algorithm (GA) and initial point algorithm (IP). Objective is the average of the objective function values with std the standard deviation and #Fevals the average of the number of function evaluations.

	DE	GA	IP
objective	0.6255	0.6293	0.2858
std	0.0048	0.0377	0.2966
#Fevals	5	13203	13

Table 2: Descriptive analysis of benchmark and of the 50 assets portfolio returns. TE denotes portfolios where the criterion function considers the tracking error volatility, while VO represents portfolio where the criterion function depends on the portfolio variance. Moreover, 1 and 20 identify the risk aversion coefficient value. The columns report the cumulated returns obtained in the range February 1995 to January 2012, the annualized average monthly return and annualized variance, the minimum and maximum monthly returns, the skewness and kurtosis indices computed on monthly returns, the 5% Value-at-Risk and Expected Shortfall, the Sharpe ratio, the average monthly turnover and the concordance index.

	Cum. Ret.	Ret. Ann.	Vol. Ann.	Min	Max	Skew	Kurt	Var(5%)	ES(5%)	Sharpe	Avg-TO	Conc.Idx
Bench	2.2030	0.1011	0.1538	-0.1601	0.1044	-0.5778	3.8469	0.0650	0.0963	0.1814	-	-
SH	5.3647	0.1404	0.1667	-0.1876	0.1121	-0.8082	4.4216	0.0681	0.1042	0.2288	0.1624	-
TE1	8.5338	0.1783	0.2259	-0.2477	0.1991	-0.5968	4.3472	0.0935	0.1405	0.2111	0.1822	0.5550
TE20	3.2629	0.1169	0.1601	-0.1853	0.1180	-0.6989	4.2987	0.0668	0.0984	0.2002	0.2457	0.5966
VO1	6.6904	0.1611	0.2100	-0.2477	0.1463	-0.7312	4.5592	0.0872	0.1319	0.2067	0.1862	0.6160
VO20	3.6833	0.1186	0.1405	-0.1646	0.1360	-0.8225	5.4076	0.0573	0.0888	0.2314	0.2208	0.5745
Portfolios with a lower bound on Market Value weight in the criterion function (at least 0.1)												
TE1	7.3774	0.1686	0.2194	-0.2383	0.1662	-0.6080	4.1221	0.0911	0.1359	0.2063	0.1723	0.5636
TE20	3.8260	0.1222	0.1524	-0.1777	0.1127	-0.6593	4.2221	0.0627	0.0921	0.2194	0.1972	0.5390
VO1	5.8385	0.1530	0.2054	-0.2383	0.1442	-0.7387	4.3963	0.0856	0.1297	0.2013	0.1773	0.6237
VO20	3.1327	0.1113	0.1361	-0.1598	0.0943	-0.9143	5.1529	0.0558	0.0861	0.2248	0.2013	0.5891

Table 3: Descriptive analysis of performance measure weights in the range February 1995 to January 2012, for the 50 assets portfolio. The minimum is not included since equal to zero for all measures. RA denotes the levels of the risk aversion. TE identifies tracking-error-based screening while VO refers to screening with portfolio variance in the objective function. Row headings refer to the performance measures listed in section 4.1.: MV, market value; ERMAD, return over mean absolut deviation; AR, appraisal ratio; BR, Burke ratio; SR, Sterling ratio; CR, Calmar ratio; RaR, Risk adjusted return; M2, Modigliani-Modigliani index; Sr, Sortino ratio; K3, Kappa 3 measure; VR, return over Value-at-risk ratio; Sh, Sharpe ratio.

	Mean	Max	St.Dev.	Mean	Max	St.Dev.
TE		RA = 1			RA = 20	
MV	0.0041	0.0911	0.0126	0.2450	0.6544	0.1758
ERMAD	0.1198	0.9987	0.2100	0.3353	0.9340	0.3191
AR	0.2904	0.9386	0.1839	0.2623	0.9588	0.2601
BR	0.0091	0.2149	0.0222	0.0335	0.4865	0.0784
SR	0.0030	0.0727	0.0095	0.0614	0.4955	0.1231
CR	0.0024	0.1074	0.0098	0.0035	0.1238	0.0162
RaR	0.0040	0.2078	0.0193	0.0140	0.3064	0.0470
M2	0.5559	0.9342	0.2175	0.0173	0.4235	0.0647
Sr	0.0018	0.1219	0.0105	0.0024	0.2780	0.0200
K3	0.0053	0.2345	0.0258	0.0195	0.6066	0.0994
VR	0.0017	0.1226	0.0100	0.0048	0.2878	0.0299
Sh	0.0026	0.3199	0.0253	0.0009	0.1770	0.0123
VO		RA = 1			RA = 20	
MV	0.0033	0.0668	0.0092	0.0100	0.1988	0.0261
ERMAD	0.1869	0.9987	0.2904	0.1634	0.9666	0.2652
AR	0.3204	0.9817	0.2213	0.3917	0.9999	0.3631
BR	0.0140	0.3627	0.0364	0.1306	0.9108	0.1814
SR	0.0081	0.1227	0.0221	0.0563	0.4404	0.1091
CR	0.0032	0.1064	0.0121	0.0204	0.3385	0.0560
RaR	0.0088	0.2840	0.0357	0.1955	0.9066	0.3111
M2	0.4390	0.9814	0.2285	0.0114	0.1933	0.0373
Sr	0.0012	0.0389	0.0042	0.0188	0.7849	0.1005
K3	0.0073	0.2518	0.0291	0.0007	0.0882	0.0064
VR	0.0075	0.4965	0.0528	0.0010	0.0673	0.0074
Sh	0.0001	0.0064	0.0007	0.0001	0.0180	0.0013

Table 4: Descriptive analysis of performance measure weights in the range February 1995 to January 2012 with MV weight with a lower bound at 10% for the 50 assets portfolio. The minimum is not included since equal to zero for all measures. RA denotes the levels of the risk aversion. TE identifies tracking-error-based screening while VO refers to screening with portfolio variance in the objective function. Row headings refer to the performance measures listed in section 4.1.: MV, market value; ERMAD, return over mean absolut deviation; AR, appraisal ratio; BR, Burke ratio; SR, Sterling ratio; CR, Calmar ratio; RaR, Risk adjusted return; M2, Modigliani-Modigliani index; Sr, Sortino ratio; K3, Kappa 3 measure; VR, return over Value-at-risk ratio; Sh, Sharpe ratio.

	Mean	Max	St.Dev.	Mean	Max	St.Dev.
TE		RA = 1			RA = 20	
MV	0.1035	0.1820	0.0109	0.3236	0.6889	0.1572
ERMAD	0.1065	0.8988	0.1879	0.3025	0.8406	0.2880
AR	0.2636	0.8447	0.1652	0.2358	0.8629	0.2344
BR	0.0082	0.1585	0.0179	0.0301	0.4378	0.0706
SR	0.0025	0.0851	0.0086	0.0520	0.4459	0.1060
CR	0.0021	0.0966	0.0088	0.0029	0.1114	0.0143
RaR	0.0035	0.1870	0.0173	0.0126	0.2758	0.0423
M2	0.5027	0.8408	0.1950	0.0156	0.3812	0.0582
Sr	0.0013	0.1097	0.0081	0.0021	0.2502	0.0180
K3	0.0029	0.1734	0.0132	0.0176	0.5459	0.0895
VR	0.0010	0.0559	0.0046	0.0044	0.2590	0.0269
Sh	0.0023	0.2879	0.0228	0.0008	0.1593	0.0111
VO		RA = 1			RA = 20	
MV	0.1029	0.1601	0.0082	0.1086	0.2789	0.0229
ERMAD	0.1653	0.8988	0.2604	0.1484	0.8699	0.2396
AR	0.2890	0.8835	0.1989	0.3519	0.8999	0.3280
BR	0.0126	0.3265	0.0327	0.1187	0.8289	0.1641
SR	0.0069	0.1104	0.0190	0.0516	0.3964	0.0990
CR	0.0027	0.0958	0.0104	0.0162	0.3046	0.0468
RaR	0.0080	0.2556	0.0321	0.1765	0.8160	0.2813
M2	0.3979	0.8833	0.2074	0.0096	0.1682	0.0316
SR	0.0012	0.0350	0.0043	0.0169	0.7955	0.0925
K3	0.0066	0.2267	0.0262	0.0007	0.0794	0.0060
VR	0.0067	0.4468	0.0475	0.0008	0.0583	0.0055
Sh	0.0001	0.0058	0.0006	0.0001	0.0162	0.0011

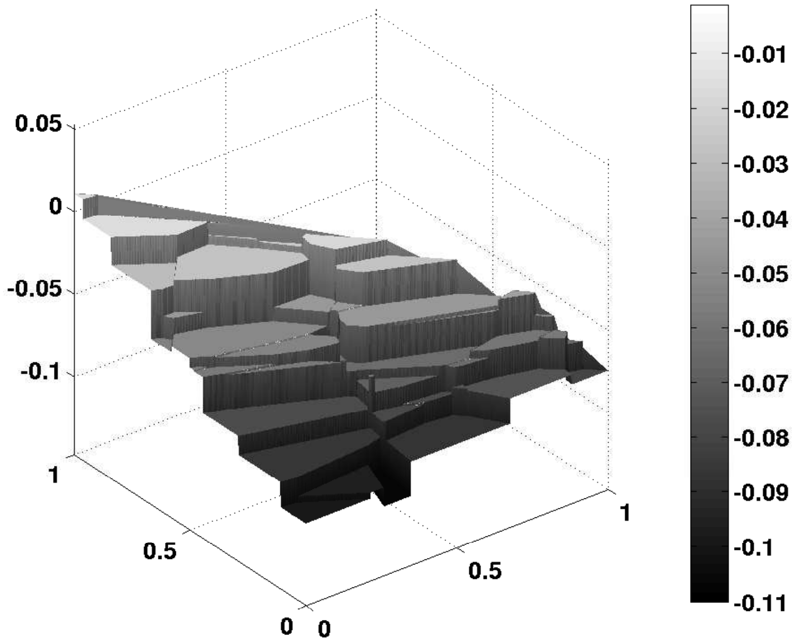


Figure 1: Surface of the objective function for the determination of the composite index weights with three performance measures. The vertical axis refers to the objective function while the other two axis report the weights of two performance measures (the third being obtained through the constraint on combination weights).

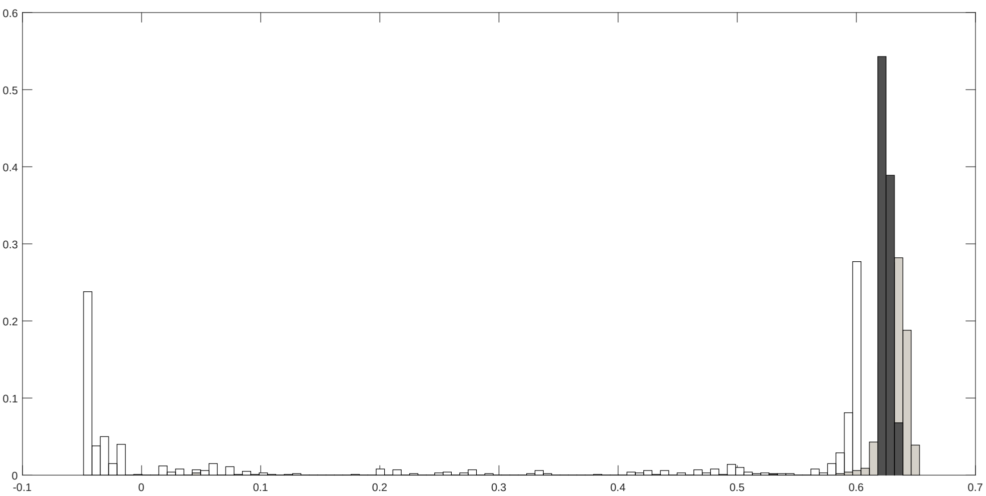
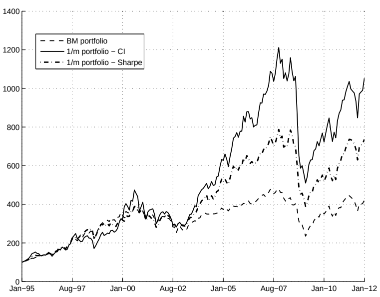
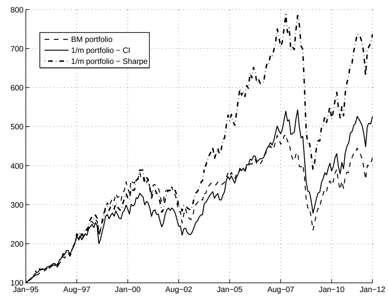


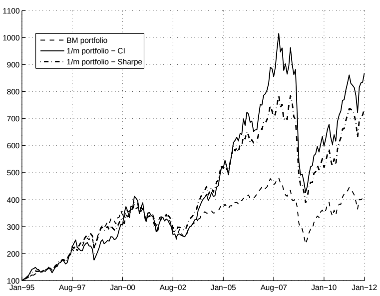
Figure 2: Distributions of the objective function values with 1000 simulated initial values for DE (dark grey), GA (light grey) and IP (white).



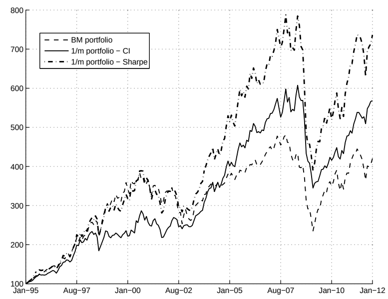
(a) TE1



(b) TE20



(c) VO1



(d) VO20

Figure 3: Cumulated returns of the strategies and of the benchmark in the range 1995-2012

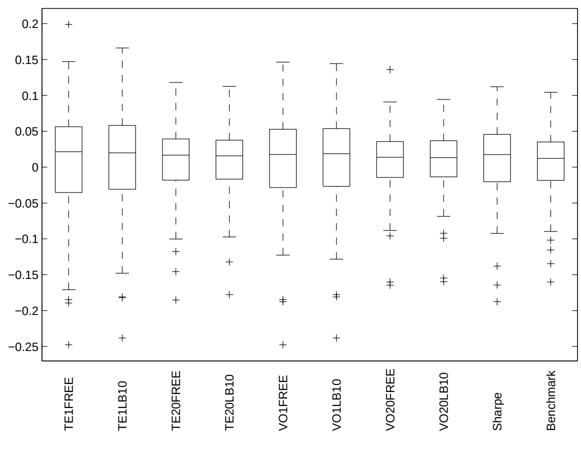
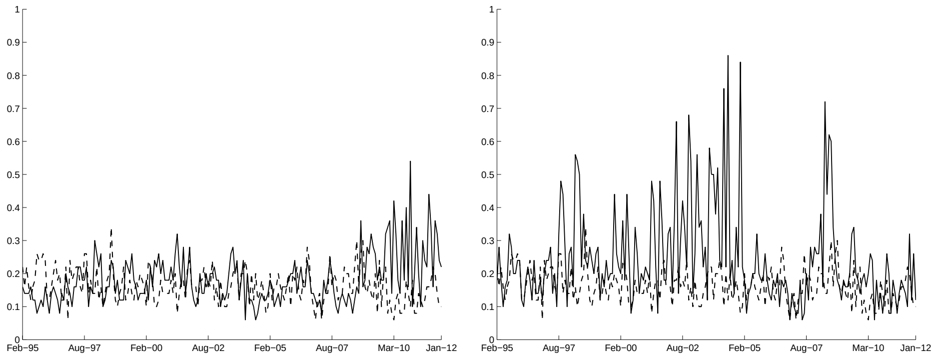
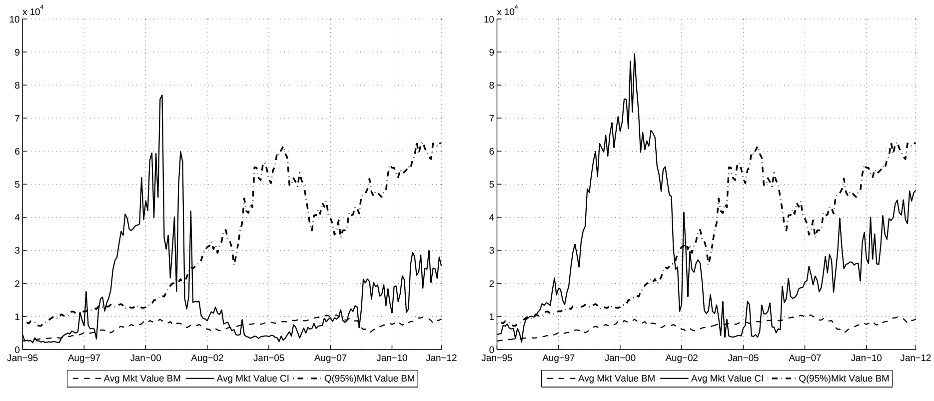


Figure 4: Boxplot for the different strategies with $k = 50$: TE denotes strategies based on the Tracking Error Volatility criterion function, while VO refers to the portfolio variance criterion function; the first number, 1 or 20 is the risk aversion coefficient; finally, FREE denotes strategies without constraints on the Market Value weight, while LB10 indicates the use of a lower bound set at 10% assigned to the Market Value in the construction of the Backward/Forward index.



(a) TE1 (b) TE20

Figure 5: Turnover induced by Sharpe-based screening (dotted line) and by the TE screening with risk aversion coefficient equal to 1 (bold line - left picture) and equal to 20 (bold line - right picture)



(a) Free MV (b) MV=10%

Figure 6: Average market value of the assets selected by the TE screening with risk aversion equal to 1 with unconstrained (bold line - left figure) or constrained market value weight (market value fixed at 10% - bold line - right figure). The figures also report the average market value of the companies included in the benchmark (dashed line), and the 95% quantile of the market value of the companies included in the benchmark (dotted line).