Economics Bulletin

Volume 32, Issue 4

A new stochastic dominance approach to enhanced index tracking problems

Renato Bruni
Università di Roma "Sapienza" Dip. di Ingegneria Informatica, Automatica e Gestionale

Francesco Cesarone
Università degli Studi Roma Tre Dipartimento di Economia

Andrea Scozzari
Università degli Studi "Niccolò Cusano" - Telematica, Roma Facoltà di Economia

Fabio Tardella
Università di Roma "Sapienza" Dip. Metodi e Modelli per l''Economia, il Territorio e la Finanza

Abstract

Enhanced Index Tracking is the problem of selecting a portfolio that should generate excess return with respect to a benchmark index. Here we propose a large-size linear optimization model for Enhanced Index Tracking that selects an optimal portfolio according to a new stochastic dominance criterion and we devise an efficient constraint generation technique to solve such a model. We then compare, on several well-known and publicly available financial data sets, the performances of the portfolios selected by our model to those of the portfolios obtained with other stochastic dominance approaches. The results seem to confirm the practical usefulness of stochastic dominance for portfolio selection.

Citation: Renato Bruni and Francesco Cesarone and Andrea Scozzari and Fabio Tardella, (2012) "A new stochastic dominance approach to enhanced index tracking problems", *Economics Bulletin*, Vol. 32 No. 4 pp. 3460-3470.
Contact: Renato Bruni - bruni@dis.uniroma1.it, Francesco Cesarone - fcesarone@uniroma3.it, Andrea Scozzari - andrea.scozzari@unicusano.it, Fabio Tardella - fabio.tardella@uniroma1.it.
Submitted: October 03, 2012. **Published:** December 24, 2012.

1. Introduction

Portfolio Optimization in Asset Management is the problem of selecting the composition of a portfolio in order to pursue a given objective, generally involving both return maximization and risk minimization. The basic *Index Tracking* (IT) problem, in particular, aims at selecting a portfolio, possibly with a small number of assets, that best tracks the performance of a given *index* or *benchmark*. Extensive reviews in the literature on this problem can be found in Beasley *et al.* (2003) and, more recently, in Canakgoz and Beasley (2008) and in Scozzari *et al.* (2012). A more ambitious task is that of out-performing a given index or benchmark. This problem has been recently addressed with various approaches under the name of *Enhanced Indexation* (EI) or *Enhanced Index Tracking* (see Canakgoz and Beasley 2008, and references therein), and the portfolio selected in this case is sometimes called *Enhanced Indexation Portfolio* (EI portfolio). The return obtained in excess w.r.t. the index is called *excess return*. Note that, in general, no guarantee of always reaching a positive excess return can be given, so the risk of under-performing the index always exists. EI models are usually built and validated using the price data of n assets and of the benchmark index over a time period. In order to simulate practical usage, a part of this time period is considered as the past (and so it is known), and the rest is considered as the future (unknown at the time of portfolio selection). The past (called in-sample) is used for finding the EI portfolio, while the future (called out-of-sample) can only be used for testing the performance of the selected portfolio.

Enhanced Indexation is a relatively recent area of research, and quantitative approaches have been mainly developed in the last decade. Alexander and Dimitriu (2005) propose to extend Index Tracking into Enhanced Indexation by generating, with a cointegration approach, portfolios for tracking two artificial indexes: the index plus a constant and the index minus the same constant. They attempt to generate excess returns by selling the plus tracking portfolio and purchasing the minus tracking portfolio. Konno and Hatagi (2005) compute a portfolio that keeps track of an index-plus-alpha portfolio with minimal transaction costs. The problem is formulated as a concave minimization under linear constraints and is solved with a branch and bound algorithm. Canakgoz and Beasley (2008) consider both Index Tracking and Enhanced Indexation problems, viewing the returns of the tracking (or enhancing) portfolios as depending on benchmark index returns, and perform linear regression. For Enhanced Indexation they propose a two-stage optimization problem using a mixed-integer linear programming formulation. Guastaroba and Speranza (2012) use a heuristic approach (called Kernel Search) for solving mixed-integer linear programming models for IT and EI including also cardinality, buy-in, and transaction costs constraints. They evaluate the efficiency and accuracy of their heuristic by comparing it with a standard exact solver.

In our opinion, there are three main limitations in the majority of existing approaches. First, EI bi-objective models (or their scalarizations) based on minimizing tracking error and maximizing excess return contain a slight contradiction: the first goal penalizes both positive and negative deviations from the index while, on the other hand, one seeks to maximize excess return, i.e. a positive tracking error. This contradiction derives from the use of a symmetric distance measure, which is not suitable for controlling the distance between the returns of the portfolio and those of the benchmark in EI. Thus an asymmetric distance measure should be used. Furthermore, EI is a computationally demanding task (see Roman *et al.* 2011) and several proposed models are too complex for being practically solved to optimality for medium or large size problems. Finally, several authors do not test their models on publicly available datasets, so comparison is often impracticable.

Recent promising approaches, on the other hand, are those based on Stochastic Dominance criteria, which imply optimality with respect to large families of utility functions (see, e.g., Nguema 2005). A practical approach for large Markets has been developed by Roman *et al*. (2011) who apply a Second order Stochastic Dominance (SSD) strategy (see also Fabian *et al.* 2011) to construct a portfolio whose return distribution dominates the one of a benchmark. The proposed model is very large but linear and is solved efficiently with cutting planes techniques.

We present here a new Linear Programming model for selecting an optimal portfolio within the stochastic dominance approach. In particular, in Section 2.1 we propose a new stochastic dominance condition, called *Cumulative Zero-order Stochastic ε-Dominance* (CZSεD), that seems to be one of the strongest stochastic dominance conditions that one could use for portfolio selection in practice. This model uses an asymmetric measure of the tracking error that only minimizes the downside deviations from the benchmark index. In view of its large size, our model is solved to optimality by using a constraint generation approach as described in Section 2.2. This approach allows for efficient solution of the basic model, and reasonable computational complexity when adding further complicating constraints coming from real-world practice such as cardinality constraints. We observe that the number of assets in a portfolio seems to be related to required level of excess return (see Cesarone *et al.* 2012). For this reason, in Section 2.2 we decided to add a constraint on the required excess return both in the model realizing CZSεD and in the one realizing SSD proposed by Roman et al. (2011). Computational results are reported in Section 2.3 for eight major stock markets across the world using data sets publicly available in Beasley (1990). A *rolling window* method is used to evaluate the performance of the selected portfolio over all time periods. Results are very encouraging and show that portfolios selected by our model have a good out-of-sample performance, and exhibit several useful properties. Finally, some conclusions and further research are provided in Section 3.

2. The Enhanced Indexation Problem

Selecting a portfolio over n available assets means deciding how much of each asset i should be purchased, with $i \in A = \{1, \dots, n\}$. Asset prices p_{it} (adjusted for dividends and splits) are observed for $m + 1$ time periods with $t \in \{0,1, \dots, m\}$ and are used to compute over the set $T = \{1, \dots, m\}$ the returns

$$r_{it} = \frac{p_{it} - p_{it-1}}{p_{it-1}} \quad i \in A, t \in T. \qquad (1)$$

For $i \in A$, the fraction of the given capital to be invested in asset i is denoted by x_i and the vector $x = (x_1, \dots, x_n)$ represents the selected portfolio. For $t \in \{0,1, \dots, m\}$, the observed value of the benchmark (e.g., the Market Index) at time *t* is denoted by b_t. The Index returns, are thus given by

$$R_t^I = \frac{b_t - b_{t-1}}{b_{t-1}} \quad i \in A, t \in T. \qquad (2)$$

Adopting a standard approximation, we assume that the portfolio return at time t is

$$R_t(x) = \sum_{i=1}^{n} r_{it}\, x_i \quad \forall t \in T \qquad (3)$$

so that the *excess return*, i.e., the difference between the portfolio return and the index return, is given by

$$\delta_t(x) = R_t(x) - R_t^I, \qquad (4)$$

which can assume any real value. Clearly, $\delta_t(x) < 0$ signals *underperformance* while $\delta_t(x) > 0$ signals *overperformance* at time t.

2.1 Stochastic Dominance

The portfolio returns $R_t(x)$ and benchmark returns R_t^I can be considered the *t*-th realizations of two discrete random variables, namely *Portfolio Return* (*PR*) and *Benchmark Return* (*BR*), respectively. Thus, the excess return $\delta_t(x)$ is the *t*-th realization of the random variable *PR*-*BR*. Two random variables may be compared by means of the Expected Utility Theory (von Neumann and Morgenstein 1944) for which a variable is preferred to another if it presents a larger value of the expected utility than the other. However, this approach depends on a specification of a utility function, which is a fairly subjective matter. Stochastic Dominance (SD) is strictly related to Expected Utility Theory and it can provide a (partial) order in the space of random variables, avoiding the specification of a particular utility function.

Let A and B be two random variables defined in a probability space, and let a_t and b_t be their realizations at time t. Denote by f_A and f_B their probability density functions, and by $F_A(\alpha) = \int_{-\infty}^{\alpha} f_A(\tau)d\tau$ and $F_B(\alpha) = \int_{-\infty}^{\alpha} f_B(\tau)d\tau$ their cumulative distributions. The comparison between A and B can be conducted by using a Stochastic Dominance approach. We recall the following well-known definitions:

- Zero-order (strong) Stochastic Dominance (ZSD): A is preferred to B w.r.t. ZSD iff $a_t \geq b_t \quad \forall t$ and the inequality is strict for at least one t. This means $\mathbb{P}(A \leq B) = 0$.
- First-order Stochastic Dominance (FSD): A is preferred to B w.r.t. FSD iff $F_A(\alpha) \leq F_B(\alpha) \quad \forall \alpha \in (-\infty, \infty)$, and the inequality is strict for at least one α.
- Second-order Stochastic Dominance (SSD): A is preferred to B w.r.t. SSD iff $\int_{-\infty}^{\alpha} F_A(\tau)d\tau \leq \int_{-\infty}^{\alpha} F_B(\tau)d\tau \quad \forall \alpha \in (-\infty, \infty)$, and the inequality is strict for at least one α.

Referring to the Expected Utility Theory, there are relations between the order of the stochastic dominance conditions and the form of the utility functions involved. For example, the FSD condition is connected to the class of non-decreasing utility functions, while the SSD condition relates to non-decreasing and concave functions, which represent an investor risk-averse behavior (see, e.g., Levy 1992).

More generally, any v^{th}-order stochastic dominance can be defined. When increasing the order of a stochastic dominance, the corresponding condition becomes less restrictive, and the v^{th}-order dominance implies the $(v+1)^{th}$-order dominance, while the opposite is not necessarily true (see, e.g., Levy 2006). In view of the above discussion, the ZSD is the strongest condition, and therefore one should aim for a portfolio whose in-sample return is preferred to the benchmark return w.r.t. ZSD. However, this condition cannot be fulfilled in practice otherwise arbitrage opportunities would arise (Meucci 2005). We therefore relax ZSD and propose the following new conditions:

- Zero-order stochastic ε-dominance (ZSεD): A is preferred to B w.r.t. ZSεD iff $a_t + \varepsilon \geq b_t \quad \forall t$, and the inequality is strict for at least one t. This means $\mathbb{P}(A + \varepsilon \leq B) = 0$.
- Cumulative Zero-order stochastic ε-dominance (CZSεD): A is preferred to B w.r.t.

CZSεD iff $\sum_{t\in S} a_t + \varepsilon \geq \sum_{t\in S} b_t \quad \forall S \subseteq T$.

Coming back to the case of EI, PR is preferred to BR w.r.t. ZSεD iff

$$\delta_t(x) \geq -\varepsilon \quad \forall t \in T \tag{5}$$

and the inequality is strict for at least one t. This means that $\delta_t(x)$ can be negative for some of the in-sample time periods (i.e., a loss), but in any case their values cannot be smaller than $-\varepsilon$ (the loss is limited). On the other hand, *PR* is preferred to *BR* w.r.t. CZSεD iff

$$\sum_{t\in S} \delta_t(x) \geq -\varepsilon \quad \forall S \subseteq T \tag{6}$$

This means that $\sum_{t\in S} \delta_t(x)$ can be negative for some subsets of the in-sample time periods (i.e., a cumulative loss), but in any case the value of the above sum cannot be smaller than -ε (the cumulative loss is limited). It is easy to see that CZSεD implies ZSεD, so it appears to be the strongest condition that a real-world portfolio could satisfy in practice under a no-arbitrage assumption.

2.2 The Optimization Model

Among all portfolios that are preferred to the Market Index with respect to the CZSεD criterion, we are interested in the one(s) having the smallest absolute value for ε. This can be obtained by solving an optimization problem in term of the above introduced decision variables x_i. Seeking the smallest absolute value for ε clearly amounts to maximizing $-\varepsilon$, while the above stochastic dominance conditions can be imposed as constraints that we here call *limiting constraints*. As usual, we also require the *budget constraint* ($\sum_{i\in A} x_i = 1$), the *no-short-selling* condition $x \in \mathbb{R}^n_+$, and we allow for the possibility of a set $\mathcal{C}$ of other further linear constraints. We thus obtain the following Linear Programming problem

$$\begin{cases} \max \ -\varepsilon \\ \sum_{t\in S} \delta_t(x) \geq -\varepsilon \qquad \forall S \subseteq T \\ \sum_{i\in A} x_i = 1 \\ x \in \mathcal{C} \\ x \in \mathbb{R}^n_+, \ \varepsilon \in \mathbb{R} \end{cases} \tag{7}$$

Note that the number of limiting constraints is exponential: one for every subset *S* of *T*. Thus, in order to practically solve this very large model, we use a *constraint generation* framework (see also Bertsimas and Tsitsiklis 1997) as follows. First we solve the Linear Program

$$
\begin{cases}
\max -\varepsilon \\
\delta_1(x) \geq -\varepsilon \\
\quad\vdots \\
\delta_m(x) \geq -\varepsilon \\
\sum_{i\in A} x_i = 1 \\
x \in \mathcal{C} \\
x \in \mathbb{R}^n_+, \;\; \varepsilon \in \mathbb{R}
\end{cases}
\tag{8}
$$

over the in-sample set T, obtaining an optimal solution x^* and values $\delta_t(x^*)$ for each $t\in T$. Then, if we can find a set of time periods B such that

$$
\sum_{t\in B\subseteq T} \delta_t\,(x^*) < -\varepsilon \tag{9}
$$

this means that there is a limiting constraint that is violated. Therefore, according to the constraint generation framework, we search for such a violated limiting constraint by means of a *separation procedure*. When such a set B is obtained, we generate the following constraint and add it to the previously solved model:

$$
\sum_{t\in B\subseteq S} \delta_t\,(x^*) \geq -\varepsilon \tag{10}
$$

We then solve the updated model, obtaining a new solution, and we iterate the procedure. After a number of generated constraints, a solution $x^{*\prime}$ is found such that a set B corresponding to a violated limiting constraint does not exists. This means that $x^{*\prime}$ is the portfolio that is preferred to the Market Index w.r.t. CZSεD.

The separation procedure can be realized by means of a so-called *oracle* that, for a given x^*, either returns a constraint violated by x^*or guarantees that no constraint is violated by x^*. The oracle is implemented here by solving the following Integer Programming problem, where γ is a numeric tolerance:

$$
\begin{cases}
\min \sum_{t\in T} y_t\, \delta_t(x^*) \\
\sum_{t\in T} y_t\, \delta_t(x^*) \leq -(\varepsilon^* + \gamma) \\
y \in \{0,1\}^m
\end{cases}
\tag{11}
$$

To try to improve the stochastic dominance models, we decided to add a constraint on the minimum required excess return both in the model realizing CZSεD and in the one realizing SSD proposed by Roman *et al*. (2011). To this end, we force the total in-sample return to be at least a fraction $k\in[0,1]$ (called *return level*) of the maximum obtainable total in-sample return R^{max} of a portfolio

$$\sum_{t \in T} R_t(x) \geq kR^{max}. \qquad (12)$$

This is mainly intended to cut away low gain solutions, but we found that it also has the effect of limiting the number of assets in the solution without explicitly introducing a cardinality constraint.

2.3 Computational Results

The described approaches have been preliminary tested using a rolling time window scheme, with 200 in-sample periods and 12 out-of-sample periods, and rebalancing every 12 periods (= 3 months). We use publicly available datasets (Beasley 1990) from Beasley's OR-Library (http://people.brunel.ac.uk/~mastjjb/jeb/orlib/indtrackinfo.html) frequently used in studies on portfolio management. They include weakly price data from March 1992 to September 1997 for the Hang Seng (Hong Kong), DAX100 (Germany), FTSE100 (UK), S&P100 (USA), Nikkei 225 (Japan), S&P500 (USA), Russell 2000 (USA) and Russell 3000 (USA) capital market indices, with 31, 85, 89, 98, 225, 457, 1318 and 2151 assets, respectively. Such prices have previously been adjusted for dividends and splits. The return rates for these eight markets have been computed as relative variations of the quotation prices ($(P_t - P_{t-1})/P_{t-1}$). The historical realizations consist in 290 rates of return.

In order to evaluate performance, for each dataset we compare our EI portfolios (CZSεD portfolios), to the EI portfolios obtained by the original model introduced in Roman *et al.* (2011) (original SSD portfolios), and to the original SSD portfolios modified by the introduction of the minimum excess return constraint (modified SSD portfolios). We therefore report, in the following 3 figures, the portfolio values on the whole sequence of the out-of-sample data for:

- the CZSεD portfolios with return level *k* = 0.8 ('CZSεD_0.8' in red),
- the original SSD portfolios ('SSD_FMR11' in green),
- the modified SSD portfolios with return level *k* = 0.8 ('SSD_0.8' in blue),
- the Market Index considered as benchmark ('Benchmark' in black).

Figure 1: Comparison between CZSεD_0.8, SSD_FMR11, SSD_0.8 and Benchmark on DAX100 dataset (German stock market)

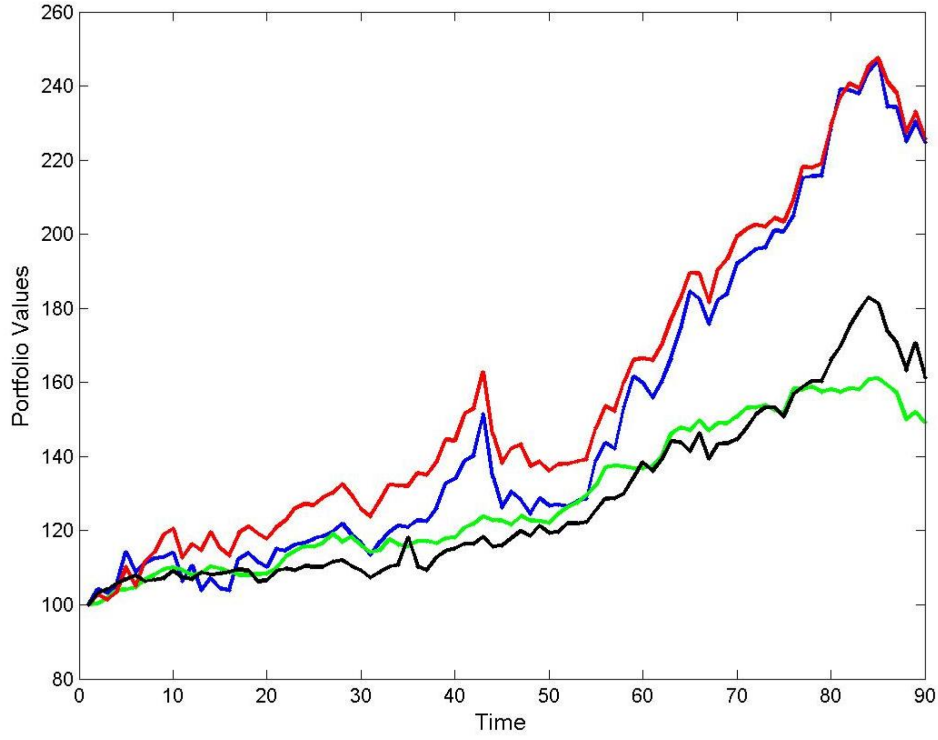


Figure 2: Comparison between CZSεD_0.8, SSD_FMR11, SSD_0.8 and Benchmark on FTSE100 dataset (UK stock market)

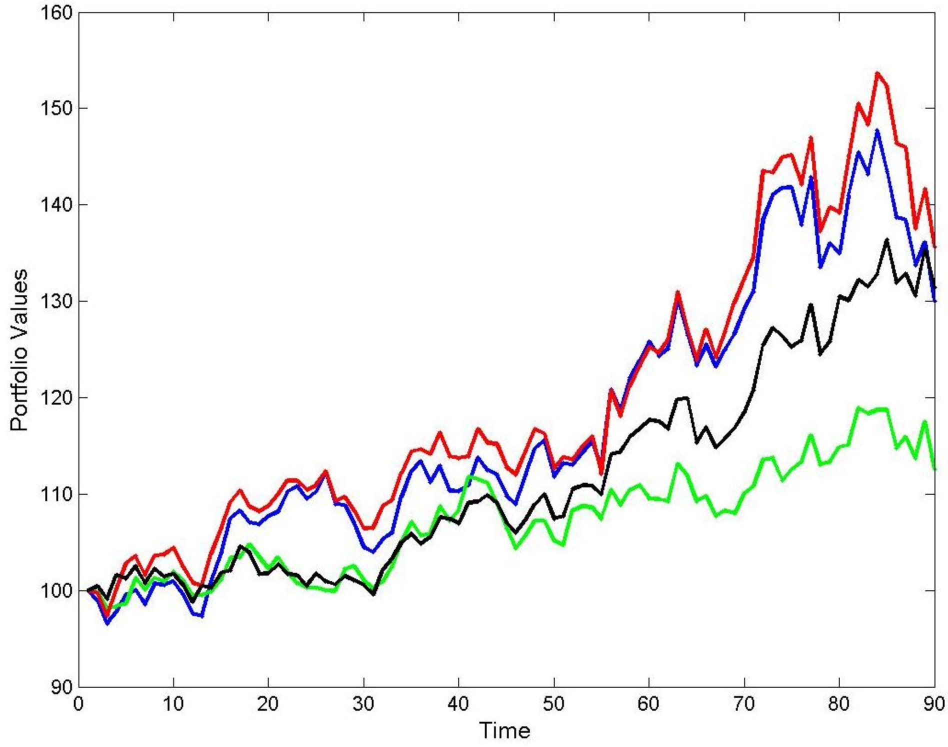
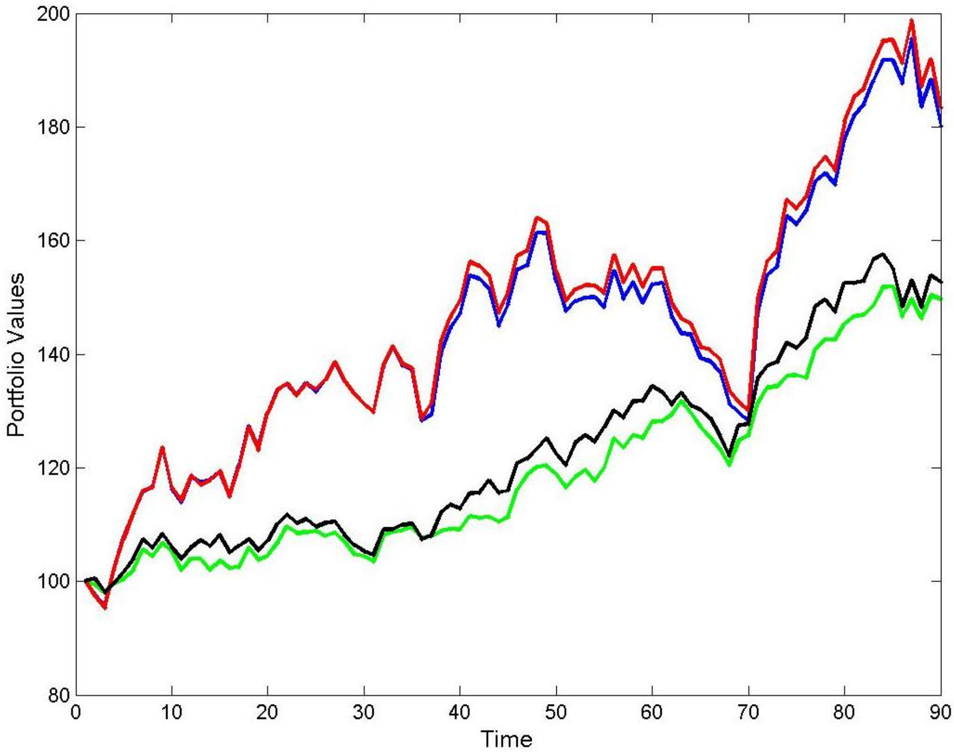


Figure 3: Comparison between CZSεD_0.8, SSD_FMR11, SSD_0.8 and Benchmark on S&P100 dataset (USA stock market)



We observe that the proposed CZSεD portfolios with return level k = 0.8 (in red) produce the best out-of-sample values for the considered datasets, closely followed by the modified SSD portfolios with return level k = 0.8 (in blue).

We moreover analyze, in the following 2 tables, the overall out-of-sample performance of the same approaches on all the 8 datasets. In particular, Table 1 reports the average difference between the portfolio values and the Market Index. For each dataset (row), the best result is in bold face. The CZSεD portfolios with return level k = 0.8 produce the best result in half of the cases, while each of the other two portfolios considered produces the best result in one fourth of the cases.

Table 1: Average differences between out-of-sample portfolio values and the Benchmark (the best result in each row is in bold face)

	SSD_FMR11 – Benchmark	**CZSεD_0.8 – Benchmark**	**SSD_0.8 – Benchmark**
HS31	-2.19	**5.95**	2.17
DAX100	0.84	**29.21**	22.02
FTSE100	-4.61	**7.92**	5.50
SP100	-3.32	**23.03**	21.50
NIKKEI	-5.00	-6.39	**0.14**
SP500	**25.30**	8.85	8.95
RUSSEL2000	34.61	16.89	**39.19**
RUSSEL3000	**47.44**	16.21	37.38

Table 2 reports an analysis based on the Sharpe index, defined as the average value of $\delta_t(x)$ divided by its standard deviation, which is undefined when negative. The proposed CZSεD portfolios with return level k = 0.8 give again the best result in half of the cases, followed, this time, by the original SSD portfolios and then by the modified SSD portfolios with return level k = 0.8.

Table 2: Out-of-sample Sharpe index (the best result in each row is in bold face)

	SSD_FMR11	CZSεD_0.8	SSD_0.8
HS31	0.0108	**0.0578**	0.0270
DAX100	-	**0.1470**	0.0534
FTSE100	-	**0.1283**	0.0820
SP100	-	**0.1323**	0.1188
NIKKEI	0.0245	-	**0.1330**
SP500	**0.2392**	0.0469	0.0783
RUSSEL2000	**0.2161**	-	0.1437
RUSSEL3000	**0.2689**	0.0665	0.1960

3. Conclusions and Further Research

Stochastic dominance approaches to the Enhanced Index Tracking problem seem to be very attractive both from a theoretical and from a practical viewpoint. However, they lead to large-size models that need to be solved with adequate techniques like the constraint generation procedure developed in this work. In this paper we proposed a model based on new approximate stochastic dominance relations and we compared it with a state-of-the-art stochastic dominance model for Enhanced Indexation on publicly available datasets finding encouraging results. The reported results refer to the case of equity investments, but, in principle, the proposed approach can be applied to other types of assets such as derivatives or bonds, possibly considering their prices directly.

From a theoretical viewpoint, in an extended version of this paper we plan to study in greater depth the relation between the ε-stochastic dominance conditions introduced here and the Expected Utility Theory. Moreover, the model computationally analyzed in this paper, adopting classical short sale constraints, could be modified for better capturing further real market conditions (see, e.g., Bottazzi *et al*. 2012). From a practical viewpoint, the tuning of the required return (risk) level for each different in-sample window will be investigated.

References

Alexander, C. and A. Dimitriu (2005) "Indexing, cointegration and equity market regimes", *International Journal of Finance and Economics* **10**, 213-231.

Beasley, J.E. (1990) "OR-Library: Distributing test problems by electronic mail", *Journal of the Operational Research Society* **41**, 1069-1072.

Beasley, J.E., N. Meade, T.-J. Chang (2003) "An evolutionary heuristic for the index tracking problem", *European Journal of Operational Research* **148**, 621-643.

Bertsimas, D. and J.N. Tsitsiklis (1997) *Introduction to Linear Optimization,* Athena Scientific: Belmont, Massachusetts.

Bottazzi, J.-M., J. Luque, M.R. Páscoa (2012) "Securities market theory: Possession, repo and rehypothecation", *Journal of Economic Theory* **147**, 477–500.

Canakgoz, N.A. and J.E. Beasley (2008) "Mixed-integer programming approaches for index tracking and enhanced indexation", *European Journal of Operational Research* **196**, 384-399.

Cesarone, F., A. Scozzari, F. Tardella (2012) "A new method for mean-variance portfolio optimization with cardinality constraints", *Annals of Operations Research* DOI 10.1007/s10479-012-1165-7.

Fabian, C.I., G. Mitra, D. Roman, V. Zverovich (2011) "An enhanced model for portfolio choice with SSD criteria: a constructive approach", *Quantitative Finance* **11**, 1525-1534.

Guastaroba, G. and M.G. Speranza (2012) "Robust investment strategies with discrete asset choice constraints using DC programming", *European Journal of Operational Research* **217**, 54-68.

Konno, H. and T. Hatagi (2005) "Index-plus-alpha tracking under concave transaction cost", *Journal of Industrial and Management Optimization* **1**, 87-98.

Levy, H. (1992) "Stochastic dominance and expected utility: Survey and analysis", *Management Science* **38**, 555-593.

Levy, H. (2006) *Stochastic Dominance: Investment Decision Making under Uncertainty*, 2nd ed. Springer, New York.

Meucci, A. (2005) *Risk and Asset Allocation*, Springer: Heidelberg, Germany.

Nguema, J.F. (2005) "Stochastic dominance on optimal portfolio with one risk-less and two risky assets", *Economics Bullettin* **7**, 1-7.

Roman, D., G. Mitra, V. Zviarovich (2011) "Enhanced indexation based on second-order stochastic dominance", Social Science Research Network (SSRN): http://ssrn.com/abstract=1776966.

Scozzari, A., F. Tardella, S. Paterlini, T. Krink (2012), "Exact and Heuristic Approaches for the Index Tracking Problem with UCITS Constraints", *Annals of Operations Research* DOI 10.1007/s10479-012-1207-1.

Von Neumann, J., and O. Morgenstern (1944) *Theory of Games and Economic Behavior*, 2nd edition, 1947; 3th edition, 1953. Princeton University Press: Princeton, New Jersey.