Ann Oper Res (2012) 196:591–609
DOI 10.1007/s10479-011-1042-9

# A mixed 0–1 LP for index tracking problem with CVaR risk constraints

**Meihua Wang · Chengxian Xu · Fengmin Xu · Hongang Xue**

Published online: 31 December 2011


**Abstract** Index tracking problems are concerned in this paper. A CVaR risk constraint is introduced into general index tracking model to control the downside risk of tracking portfolios that consist of a subset of component stocks in given index. Resulting problem is a mixed 0–1 and non-differentiable linear programming problem, and can be converted into a mixed 0–1 linear program so that some existing optimization software such as CPLEX can be used to solve the problem. It is shown that adding the CVaR constraint will have no impact on the optimal tracking portfolio when the index has good (return increasing) performance, but can limit the downside risk of the optimal tracking portfolio when index has bad (return decreasing) performance. Numerical tests on Hang Seng index tracking and FTSE 100 index tracking show that the proposed index tracking model is effective in controlling the downside risk of the optimal tracking portfolio.

**Keywords** Index tracking · CVaR constraints · Cardinality constraints · Mixed 0–1 LP

## 1 Introduction

Index tracking problems are popular in the field of passive fund management. Passive fund managers aim to reproduce the performance of a stock market index by investing in a subset

M. Wang
Department of Mathematics of Faculty of Science, Xi'an Jiaotong University, Xi'an, 710049, P.R. China
e-mail: yuyu0504@163.com

C. Xu
SKLMSE Lab. and Department of Mathematics, Xi'an Jiaotong University, Xi'an, 710049, P.R. China
e-mail: mxxu@mail.xjtu.edu.cn

F. Xu (✉)
Department of Mathematics, Xi'an Jiaotong University, Xi'an, 710049, P.R. China
e-mail: fengminxu@mail.xjtu.edu.cn

H. Xue
School of Economics and Finance, Xi'an Jiaotong University, Xi'an, 710061, P.R. China
e-mail: xhg@mail.xjtu.edu.cn

of the stocks included in the index. Such a portfolio is called a *tracking portfolio* (Beasley et al. 2003; Ruiz-Torrubiano and Suárez 2009). Generally speaking, there exist two strategies to reproduce the index: full replication and optimized replication. The former is to purchase each component stock as the same proportion as in the index. The resulting tracking portfolio can track the index well, but incur high transaction frequency and costs in practice. Additionally, it is impossible to purchase all of the component stocks when the index consists of a large number of stocks. The strategy of optimized replication is to minimize tracking errors between the tracking portfolio and the index without purchasing all of the component stocks. This strategy involves much lower transaction frequency and costs in practice. The focus of the paper is the strategy of optimized replication.

The performance of a tracking portfolio is measured by the tracking errors between the tracking portfolio and the index. Different definitions of tracking error have been proposed in literature. Roll takes the variance of differences between the return of the tracking portfolio and the return of the index as the tracking error in Roll (1992), and formulates the index tracking problem by a quadratic programming problem. The same definition is also used in Kwiatkowski (1992), Shapcott (1992), Coleman et al. (2006). However, Beasley et al. point out that the definition is irrational since the variance will be zero when the differences are constants in Beasley et al. (2003). In fact, Roll's definition ignores the deviation between the returns of the tracking portfolio and of the index. Then the definition of tracking error as the mean square difference in the returns of the tracking portfolio and of the index is used in Beasley et al. (2003), Lobo et al. (2007). Clarke et al. define the tracking error as the "absolute difference between the managed portfolio return and the benchmark portfolio return" in Clarke et al. (1994). Sharp emphasizes that linear or absolute deviations between the return of the tracking portfolio and the return of the index are more relevant in Sharpe (1971). In view of this, Markus et al. give four alternative definitions of tracking errors in Markus et al. (1999), and the resulting index tracking problems can be converted into linear programs. In this paper, the mean absolute deviation between the return of the tracking portfolio and the return of the index (denoted in Markus et al. 1999) is used to measure the tracking error. This measure of tracking errors has the advantage of being linearizable. Furthermore, when the constraints are also linear, the resulting index tracking problem can be converted into a linear program so that existing sophisticated algorithms such as simplex algorithms and interior point algorithms can be used (Ross et al. 1986; Ye 1997).

Recently, researchers mainly aim to minimize the tracking error subject to some constraints on the number of stocks in the portfolio, short sale and investment share restriction (Beasley et al. 2003; Ruiz-Torrubiano and Suárez 2009; Canakgoz and Beasley 2009; Krink et al. 2009). In this paper, a CVaR risk constraint on the tracking portfolio is introduced into general index tracking mode to control the downside risk of the optimal tracking portfolio. In recent years, the shortfall risk measures become more popular and practical in risk management area. VaR is an important shortfall risk measure, but it has some undesirable mathematical characteristics such as lack of subadditivity and convexity. When a risk measure satisfies the characteristics of subadditivity, positive homogeneity, monotonicity and translation invariance, it is a coherent risk measure (Artzner et al. 1999). Hence VaR is not coherent except the case when the return rate of the portfolio is normal distributed. Furthermore, VaR is difficult to be optimized when it is calculated from scenarios. Recent researches on the portfolio selection area focus on coherent risk measures, for example the Conditional Value at Risk (CVaR). Rockafellar and Uryasex (2000) propose the application of CVaR in portfolio selection problems. The main advantage of CVaR is that whether CVaR is used in objective function or in a constraint, the resulting model can be converted into a linear program so that the optimal solution can be effectively obtained using existing optimization software. More details about CVaR can be found in Rockafellar and Uryasex (2002). Adding CVaR constraint into index tracking problems can control the downside risk of the optimal tracking portfolio, so as to prevent the investors from large loss. When the return rate of the tracking portfolio maintains increasing, the introduced CVaR constraint is inactive and has no impact on the optimal tracking portfolio. However, when the return rate of the tracking portfolio falls down, the CVaR constraint will limit the downside risk of the optimal tracking portfolio.

When a CVaR constraint is introduced into general index tracking model, the resulting index tracking problem can be converted into a mixed 0–1 linear programming problem. When the number of 0–1 variables is relatively small, the resulting mixed 0–1 linear program can be efficiently solved by the standard optimization software CPLEX. However, it will take long time to solve the problems with high dimension due to the cardinality constraints. Krink et al. propose an evolution and combinatorial search method to determine the subset of the stocks in Krink et al. (2009). Ruiz-Torrubiano et al. design a hybrid optimization approach based on RAR crossover operators to solve the constrained index-tracking problem (a mixed 0–1 quadratic program) in Ruiz-Torrubiano and Suárez (2009). In this paper, a hybrid genetic approach based on the hybrid approach is applied to solve the resulting mixed 0–1 linear program when the number of 0–1 variables is large. Tests of the proposed model are performed on the Hang Seng index tracking and FTSE 100 index tracking. Tests show that adding the CVaR risk constraint can indeed limit the downside risk of the optimal tracking portfolio and do not significantly affect the performance of the index tracking. Furthermore, the hybrid genetic method can solve the resulting mixed 0–1 linear programming problems efficiently and adding the CVaR risk constraint does not substantially increase the solving difficulty.

The rest of the paper is organized as follows. General index tracking problems and related models are described in Sect. 2. CVaR risk constraints on the tracking portfolio are introduced in Sect. 3 and modified index tracking model is presented. Numerical tests on Hang Seng index tracking and FTSE 100 index tracking are reported in Sect. 4. Conclusions and future researches are given in Sect. 5.

Throughout this paper prime ($'$) denotes transposition without special declaration. The notation $s_i$ or $(s)_i$ is used to denote the $i$th component of the vector $s$.

## 2 Index tracking problem

In this section, general index tracking problem will be formulated by a mixed 0–1 linear program, and constraints on cardinality, investment share restriction and so on will be taken into account.

Assumed that an investor hopes to construct a portfolio to track the performance of a stock index which is made up of $N$ component stocks. Historical prices (values) of $N$ component stocks and the index over time period $0, 1, 2, \ldots, T$ will be used. The investor hopes to construct the portfolio at time $T$ and holds it for period $[T+1, T+L]$. Main assumption here is that the past is a guide to the future (Beasley et al. 2003), that is, the tracking portfolio obtained from the data sets over time $[0, T]$ is also suitable for the period $[T+1, T+L]$. However, it should be noticed that the optimal tracking portfolio obtained by the data over time $0, 1, 2, \ldots, T$ may not be optimal after the time when the index is adjusted. The value of $L$ depends upon the frequency with which the component stocks of the index are adjusted. When the component stocks of the index are adjusted, some stocks in the index will be discarded and some new stocks will be added into the index. Nowadays, many stock indices are adjusted in about six months.

## 2.1 Tracking error

Let $I_t$ be the value of the index and $S^t = \{S_1^t, S_2^t, \dots, S_N^t\}'$ be the prices of component stocks at time $t$ ($t \in \{0, 1, 2, \dots, T\}$). Then the return rates of the index and component stocks at time $t$ are calculated by

$$R^t = \frac{I_t - I_{t-1}}{I_{t-1}}; \qquad r_i^t = \frac{S_i^t - S_i^{t-1}}{S_i^{t-1}}; \quad t = 1, 2, \dots, T$$

Let $x = (x_1, x_2, \dots, x_N)'$ be the tracking portfolio, where $x_i$ is the investment weight in the $i$th component stock. It is assumed that the investment vector $x$ keeps no change in whole investment period. An alternative strategy for index tracking is to keep the investment amount of each stock unchanged. It is clear that the later strategy causes lower transaction costs because no reallocation of capital is needed after the investment amount in each asset is initially determined. The strategy to keep $x$ unchanged requires actively managing the portfolio. This may incur high transaction costs, but possible to solve the index tracking problem effectively. There is no final conclusion that which strategy is better (Ruiz-Torrubiano and Suárez 2009). The interested readers can do more in-depth studies.

According to Markus et al. (1999), the mean absolute deviation between the return rate of the tracking portfolio and the return rate of the index

$$TE = \frac{1}{T} \sum_{t=1}^{T} \left| R^t - \sum_{i=1}^{N} r_i^t x_i \right| \tag{1}$$

is used as the measure of the tracking error. The investor hopes to obtain an optimal tracking portfolio by minimizing the tracking error subject to some practical constraints.

## 2.2 Practical constraints

When the optimized replication strategy is applied to generate the optimal tracking portfolio, $K$ ($K < N$) component stocks are generally selected to track the index. 0–1 variables $Z_i$ ($i = 1, 2, \dots, N$) are introduced to indicate the stock selection problem with

$$Z_i = \begin{cases} 1, & \text{the } i\text{th stock is included in the tracking portfolio,} \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

Then, the following constraint ensures that there are precisely $K$ stocks included in the tracking portfolio.

$$\sum_{i=1}^{N} Z_i = K \tag{3}$$

Due to the limitation of investment amount, diversification requirement and actual transaction restriction, the investment proportion on each stock should have a lower bound and an upper bound.

$$L_i Z_i \le x_i \le U_i Z_i, \quad i = 1, 2, \dots, N \tag{4}$$

with $0 < L_i < U_i < 1$, where $L_i$ and $U_i$ are the lower and upper bounds of the investment proportion on stock $i$. (4) shows that if stock $i$ is not selected in the tracking portfolio (i.e., $Z_i = 0$), then $x_i = 0$, and if stock $i$ is selected in the tracking portfolio (i.e., $Z_i = 1$), hence the value of $x_i$ is limited in the interval $[L_i, U_i]$.

Finally, the budget constraint

$$\sum_{i=1}^{N} x_i = 1 \tag{5}$$

is included to ensure total investment proportion equal to one.

### 2.3 General index tracking model

It follows analysis in Sects. 2.1 and 2.2 that the general model for the index tracking problem is given as follows:

$$\begin{aligned}
\min_{x,Z} \quad & TE = \frac{1}{T}\sum_{t=1}^{T}\left|R^t - \sum_{i=1}^{N} r_i^t x_i\right| \\
\text{s.t.} \quad & \sum_{i=1}^{N} Z_i = K \\
& L_i Z_i \le x_i \le U_i Z_i, \quad i = 1, 2, \dots, N \\
& \sum_{i=1}^{N} x_i = 1 \\
& Z_i \in \{0, 1\}, \quad i = 1, 2, \dots, N
\end{aligned} \tag{6}$$

Since the objective function in (6) is piecewise linear and not differentiable, variables $q_t^+ \ge 0, q_t^- \ge 0$ $(t = 1, 2, \dots, T)$ are introduced such that

$$q_t^+ + q_t^- = \left|R^t - \sum_{i=1}^{N} r_i^t x_i\right|$$

$$q_t^+ - q_t^- = R^t - \sum_{i=1}^{N} r_i^t x_i$$

Then problem (6) can be converted into the following mixed 0–1 integer linear programming problem

$$\begin{aligned}
\min_{x,q^+,q^-,Z} \quad & TE = \frac{1}{T}\sum_{t=1}^{T}(q_t^+ + q_t^-) \\
\text{s.t.} \quad & q_t^+ - q_t^- = R^t - \sum_{i=1}^{N} r_i^t x_i, \quad t = 1, 2, \dots, T \\
& \sum_{i=1}^{N} Z_i = K \\
& L_i Z_i \le x_i \le U_i Z_i, \quad i = 1, 2, \dots, N \\
& \sum_{i=1}^{N} x_i = 1 \\
& Z_i \in \{0, 1\}, \quad i = 1, 2, \dots, N \\
& q_t^+ \ge 0, \qquad q_t^- \ge 0, \quad t = 1, 2, \dots, T
\end{aligned} \tag{7}$$

*Remark* In practice, there are transaction costs when the selected stocks are purchased. The transaction cost can be taken into account by modifying model (7). Let $\phi(x)$ be the transaction costs associated with a tracking portfolio $x$. Then the budget constraint can be modified as

$$\sum_{i=1}^{N} x_i + \phi(x) = 1 \tag{8}$$

Generally, the transaction cost function $\phi(x)$ is a separable function, i.e.,

$$\phi(x) = \sum_{i=1}^{N} \phi_i(x_i),$$

where $\phi_i(x_i)$ is the transaction cost function for the $i$th stock. When the function $\phi_i(x_i)$ is linear with respect to $x_i$, the modified index tracking model can also be converted into a mixed 0–1 integer linear program.

## 3 CVaR risk constraints on tracking portfolio

In this section, CVaR risk constraints on the tracking portfolio are included into the index tracking model, so as to control the downside risk of the tracking portfolio. When the benchmark stock index tends to fall down, the optimal tracking portfolio obtained from model (7) will have large downside risk and cause large loss for the investor. It is important for investors to have tracking portfolios with risk controled. Various risk measures are available to control the risk, and it is important to select a proper risk measure to control the downside risk.

CVaR is a coherent risk measure and becomes more popular in practical risk management area (Rockafellar and Uryasex 2000, 2002). Pflug (2000), Ogryczak and Ruszczyński (2002) show that CVaR is stable in the sense of continuity with respect to the confidence level. Furthermore, whether the CVaR is used in the object function or in a constraint, the resulting models can be converted into a linear programming problem.

### 3.1 CVaR risk constraints and corresponding model

Let $l(x,r)$ be the loss function associated with decision vector $x$ (the tracking portfolio) and the random vector $r$ (the return rate vector). Let $p(r)$ be the density function of the return rate vector $r$. Then the probability of $l(x,r)$ not exceeding a threshold $w$ is given by

$$\psi(x,w) = \int_{l(x,r)\le w} p(r)dr$$

**Definition 1** (VaR). The VaR risk of the loss associated with a decision vector $x$ and a specified probability level $\theta$ in (0, 1) is the value

$$\mathrm{VaR}_\theta(x) = \min\{w \in \mathbf{R} : \psi(x,w) \ge \theta\} \tag{9}$$

**Definition 2** (CVaR). The CVaR risk of the loss associated with a decision vector $x$ and a specified probability level $\theta$ in (0,1) is given by

$$\mathrm{CVaR}_\theta(x) = \frac{1}{1-\theta}\int_{l(x,r)\ge \mathrm{VaR}_\theta(x)} l(x,r)p(r)dr \tag{10}$$

In order to control the downside risk of a tracking portfolio $x$, for a given confidence level $\theta$, the CVaR risk of the portfolio will not exceed a loss tolerance $\alpha$ given by the investor, that is,

$$\mathrm{CVaR}_\theta(x) = \frac{1}{1-\theta}\int_{l(x,r)\geq \mathrm{VaR}_\theta(x)} l(x,r)p(r)dr \leq \alpha \tag{11}$$

Since the value of $\mathrm{VaR}_\theta(x)$ is unknown, it is difficult to obtain $\mathrm{CVaR}_\theta(x)$ from the equation. Rockafellar and Uryasex (2000, 2002) propose the following auxiliary function

$$\begin{aligned} F_\theta(x,w) &= w + \frac{1}{1-\theta}\mathbf{E}\{[l(x,r)-w]^+\} \\ &= w + \frac{1}{1-\theta}\int [l(x,r)-w]^+ p(r)dr, \end{aligned} \tag{12}$$

and show that for any given confidence level $\theta$ and loss tolerance $\alpha$, the problem

$$\begin{array}{ll} \min\limits_{(x,w)\in \mathbf{X}\times\mathbf{R}} & h(x) \\ \text{s.t.} & F_\theta(x,w)\leq\alpha \end{array} \tag{13}$$

is equivalent to the problem

$$\begin{array}{ll} \min\limits_{x\in\mathbf{X}} & h(x) \\ \text{s.t.} & \mathrm{CVaR}_\theta(x)\leq\alpha, \end{array} \tag{14}$$

here, $\mathbf{E}\{\xi\}$ denotes the expectation of the random variable $\xi$, $[l(x,r)-w]^+ = \max\{l(x,r)-w, 0\}$, and $h(x)$ is an objective function with $x\in\mathbf{X}$.

When the index tracking problem is considered, the loss function $l(x,r) = -x'r$, and the function $F_\theta(x,w)$ can be further approximated when the values of return vector $r$ are given by a collection of discrete vectors $\{r^1, r^2, \ldots, r^T\}$ where $r^t$ is the return rate of stocks at the period $t$. When the average is used for expectation, an approximation to the function $F_\theta(x,w)$ is given by

$$\tilde{F}_\theta(x,w) \approx w + \frac{1}{(1-\theta)T}\sum_{t=1}^{T}[-x'r^t - w]^+ \tag{15}$$

After introducing the downside risk control with CVaR risk measure, the modified index tracking model is given as follows:

$$\begin{array}{lll} \min\limits_{x,q^+,q^-,Z,w} & TE = \dfrac{1}{T}\displaystyle\sum_{t=1}^{T}(q_t^+ + q_t^-) & \\ \text{s.t.} & w + \dfrac{1}{(1-\theta)T}\displaystyle\sum_{t=1}^{T}[-x'r^t - w]^+ \leq \alpha & \\ & q_t^+ - q_t^- = R^t - \displaystyle\sum_{i=1}^{N} r_i^t x_i, \quad t=1,2,\ldots,T & \\ & \displaystyle\sum_{i=1}^{N} Z_i = K & \\ & L_i Z_i \leq x_i \leq U_i Z_i, \quad i=1,2,\ldots,N & \\ & \displaystyle\sum_{i=1}^{N} x_i = 1 & \\ & Z_i\in\{0,1\}, \quad i=1,2,\ldots,N & \end{array} \tag{16}$$

Since there is $[\cdot]^+$ in the first constraint, the problem is not differentiable. Let

$$p_t = [-x'r^t - w]^+, \quad t = 1, 2, \ldots, T$$

then we have

$$p_t \geq -x'r^t - w, \qquad p_t \geq 0, \quad t = 1, 2, \ldots, T$$

and problem (16) can be converted into:

$$\begin{array}{ll}
\min\limits_{x,q^+,q^-,Z,w,p} & TE = \dfrac{1}{T}\displaystyle\sum_{t=1}^{T}(q_t^+ + q_t^-) \\
\text{s.t.} & w + \dfrac{1}{(1-\theta)T}\displaystyle\sum_{t=1}^{T} p_t \leq \alpha \\
& p_t \geq -x'r^t - w, \quad t = 1, 2, \ldots, T \\
& q_t^+ - q_t^- = R^t - \displaystyle\sum_{i=1}^{N} r_i^t x_i, \quad t = 1, 2, \ldots, T \\
& \displaystyle\sum_{i=1}^{N} Z_i = K \\
& L_i Z_i \leq x_i \leq U_i Z_i, \quad i = 1, 2, \ldots, N \\
& \displaystyle\sum_{i=1}^{N} x_i = 1 \\
& p_t \geq 0, \qquad q_t^+ \geq 0, \qquad q_t^- \geq 0, \quad t = 1, 2, \ldots, T \\
& Z_i \in \{0, 1\}, \quad i = 1, 2, \ldots, N
\end{array} \tag{17}$$

Problem (17) is also a mixed 0–1 linear program with $N + 3T + 1$ continuous variables and $N$ 0–1 variables.

## 3.2 Necessity of adding CVaR constraints

The main contribution of this paper is to add CVaR risk constraints into the index tracking problem to control the downside risk of tracking portfolios. In the following, we will show that adding the CVaR constraint will have no impact when the index has good (return increasing) performance, but can control the downside risk when index has bad (return decreasing) performance.

Consider the index tracking problem

$$\begin{array}{ll}
\min\limits_{(x,w)\in(\mathbf{X}\times\mathbf{R})} & TE \\
\text{s.t.} & \mathrm{CVaR}_\theta(x) \leq \alpha,
\end{array} \tag{18}$$

where $\mathbf{X}$ is a feasible region of the tracking portfolio given by general tracking model, for example, model (7). The confidence level $\theta$ is generally given in $[0.5, 1]$ and the loss tolerance $\alpha > 0$. In the following, we will give some analysis for two cases.

*Case 1: loss function $l(x, r) \leq 0$* For a given confidence level $\theta \in (0, 1)$, loss tolerance $\alpha > 0$ and a feasible portfolio $x$, we have

$$\begin{aligned}
\mathrm{CVaR}_\theta(x) &= \frac{1}{1-\theta}\int_{l(x,r)\geq \mathrm{VaR}_\theta(x)} l(x, r)p(r)dr \\
&\leq 0 \\
&< \alpha.
\end{aligned}$$

This shows that the CVaR constraint is satisfied for any $x \in \mathbf{X}$, and hence is inactive in (18). Thus, we have

$$x^* \in \arg\min_{x\in\mathbf{X}} TE \iff x^* \in \arg\min_{x\in\mathbf{X}} TE, \quad \text{subject to} \quad \mathrm{CVaR}_\theta(x) \le \alpha.$$

Then, the following conclusion can be obtained. When the return keeps positive, the tracking portfolio obtained from the model with CVaR constraints is the same as the one obtained from the model without CVaR constraints, that is, the CVaR constraint has no impact on the optimal tracking portfolio.

*Case 2: loss function* $l(x,r) \ge 0$ For a given confidence level $\theta \in (0,1)$, loss tolerance $\alpha > 0$ and a feasible portfolio $x$, we have

$$\begin{aligned}\mathrm{CVaR}_\theta(x) &= \frac{1}{1-\theta}\int_{l(x,r)\ge \mathrm{VaR}_\theta(x)} l(x,r)p(r)dr\\ &\ge 0\end{aligned}$$

Let $x^*$ be the tracking portfolio obtained from the model without CVaR constraints:

$$x^* \in \arg\min_{x\in\mathbf{X}} TE$$

If $\mathrm{CVaR}_\theta(x^*) < \alpha$, then the CVaR constraint is inactive. When the value of $\alpha$ is not too large, the CVaR constraint will be active and control the downside risk of the portfolio. Most investors are risk-averse and their loss tolerance $\alpha$ are generally not too large. So adding the CVaR constraint into the problem can help risk-averse investors to control the downside risk efficiently.

Above analysis shows that adding the CVaR constraint into the index tracking problem is helpful to investors, especially to risk-averse investors. It has no impact on the optimal tracking portfolio when there is no loss in the market. However, when large loss occurs in the market, the CVaR constraint offers help to investors in controlling the downside risk. Of course, the effect of the CVaR constraint in controlling the downside risk depends upon investor's selection in the values of $\alpha$ and $\theta$, that is, the attitude of the investor to risk.

Problems (7) and (17) are mixed 0–1 linear programs and can be effectively solved by some available optimization software such as CPLEX and Lingo when the number of 0–1 variables $Z_i$ is small. However, CPLEX and Lingo can not solve the problem in short time on a personal computer, when the number of 0–1 variables is large. In this case, a hybrid genetic method based on the method proposed in Ruiz-Torrubiano and Suárez (2009) is applied to solve these problems to obtain the optimal tracking portfolio. A brief description of the hybrid genetic method is given in Appendix. More detailed description and performance about the hybrid genetic method can be found in Ruiz-Torrubiano and Suárez (2009).

## 4 Numerical tests

In this section, the proposed index tracking model with the CVaR risk constraint is tested on practical data sets. Comparison results between the performance of optimal tracking portfolios with and without the CVaR risk constraint are reported. Moreover, results for different values of parameters in the models are also presented. These results show the effectiveness of the proposed model and the necessity of introducing CVaR risk constraint into index tracking problems.

**Table 1** In-sample: "No CVaR constraint" versus "With CVaR constraint (HS)"

| Card. | No CVaR constraints | | | With CVaR risk tolerance $\alpha = 0.1$ | | | With CVaR risk tolerance $\alpha = 0.06$ | | |
|---|---|---|---|---|---|---|---|---|---|
| | TE | CVaR | Time (s) | TE | CVaR | Time (s) | TE | CVaR | Time (s) |
| $K = 5$ | 5.012e-3 | 0.0734 | 3.19 | 5.012e-3 | 0.0734 | 4.78 | 9.047e-3 | 0.0600 | 1.64 |
| $K = 6$ | 4.160e-3 | 0.0742 | 3.97 | 4.160e-3 | 0.0742 | 5.21 | 8.173e-3 | 0.0600 | 1.55 |
| $K = 7$ | 3.736e-3 | 0.0737 | 6.48 | 3.736e-3 | 0.0737 | 8.73 | 7.822e-3 | 0.0600 | 1.65 |
| $K = 8$ | 3.386e-3 | 0.0750 | 7.52 | 3.386e-3 | 0.0750 | 11.36 | 7.331e-3 | 0.0600 | 0.86 |
| $K = 9$ | 3.095e-3 | 0.0745 | 8.64 | 3.095e-3 | 0.0745 | 11.61 | 7.196e-3 | 0.0600 | 1.45 |
| $K = 10$ | 2.807e-3 | 0.0756 | 8.81 | 2.807e-3 | 0.0756 | 11.78 | 6.974e-3 | 0.0600 | 0.65 |

Tests will be performed on the data from the OR-Library (see Beasley et al. 2003) which is a publicly available collection of test data sets for a variety of operations research problems. Two generally used data sets: Hang Seng Index (Hong Kong) and FTSE 100 (UK) are selected for test in this section. Stocks with missing values are dropped, and hence 31 stocks for Hang Seng and 89 stocks for FTSE are used in tests, respectively. Weekly closing prices of these stocks from 1992 to 1997 are selected as test samples. The data sets include 291 history weekly prices and hence 290 weekly return rates for each stock and each index. The data sets of weekly return rates are then divided into two parts. The first 145 return rates are used as in-sample data to obtain the optimal tracking portfolios, and the remaining return rates are used as out-of-sample data to test the performance of the optimal tracking portfolios. Then the experiments are conducted in the following two stages:

*Stage 1 (in-sample calculations)*: Problems (7) and (17) are solved to obtain the optimal tracking portfolios from the in-sample data. The optimization software "CPLEX" is used to solve the problems when $N = 31$ (Hang Seng Index), and the hybrid genetic approach is applied to solve the problems for $N = 89$ (FTSE 100 Index), since the tests are performed on a personal computer with Intel CPU 2.66 GHz and 2 GB memory.

*Stage 2 (out-of-sample calculation)*: The performance of the optimal tracking portfolios obtained in *Stage 1* are tested using the out-of-sample data.

The following parameter values are used in tests: lower bound $L_i = 1\%$ and upper bound $U_i = 50\%$ for investment proportion on each stock, The cardinality $K$ in tests is varied from 5 to 10 for both models and confidence level $\theta = 95\%$ for the model with the CVaR constraint.

## 4.1 Results for Hang Seng index

Tables 1, 2 and Figures 1–6 give the in-sample and out-of-sample results for Hang Seng Index with $\alpha = 0.1$ and $\alpha = 0.06$ (the risk tolerance of investors) in the model (17), where TE denotes the tracking error between the resulting optimal tracking portfolio and the actual index, CVaR denotes the CVaR risk of the resulting portfolio, and Time(s) is the CPU time in seconds.

*Results in in-sample* The CPLEX solves the Hang Seng index tracking problems effectively. It can be observed from Table 1 that the results for the model without the CVaR constraint and the model with the CVaR constraint are the same in the case $\alpha = 0.1$ and difference appears when the value of $\alpha$ reduces to 0.06. It coincides the analysis in Sect. 3.2, and shows that adding CVaR constraint into the model has no impact when the tolerance

**Table 2** Out-of-sample: "No CVaR constraint" versus "With CVaR constraint" (HS)

| Card. | No CVaR constraint | | With CVaR risk tolerance $\alpha = 0.06$ | |
|---|---|---|---|---|
| | TE | CVaR | TE | CVaR |
| $K = 5$ | 6.498e-3 | 0.0587 | 8.953e-3 | 0.0524 |
| $K = 6$ | 5.280e-3 | 0.0584 | 8.612e-3 | 0.0485 |
| $K = 7$ | 4.341e-3 | 0.0567 | 8.246e-3 | 0.0513 |
| $K = 8$ | 4.234e-3 | 0.0560 | 7.695e-3 | 0.0512 |
| $K = 9$ | 3.712e-3 | 0.0547 | 7.718e-3 | 0.0511 |
| $K = 10$ | 3.544e-3 | 0.0553 | 7.506e-3 | 0.0505 |

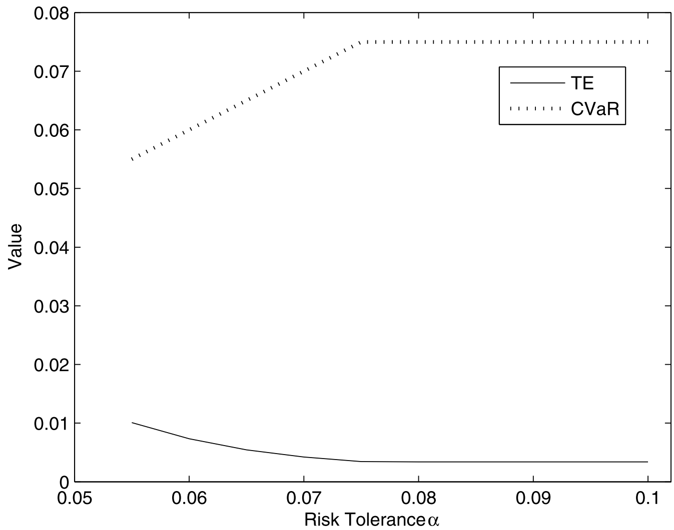


**Fig. 1** In-sample TE and CVaR for various risk tolerance levels $\alpha$ in CVaR constraint (HS)

value is set to relatively large while the CVaR risk of portfolios are small. Observations for the case $K = 8$ are made and it is found that the CVaR constraint in (17) is inactive, when risk tolerance $\alpha = 0.1$, while the CVaR constraint is active, when $\alpha = 0.06$. If an investor holds the optimal tracking portfolio obtained with $\alpha = 0.1$, he will endure CVaR risk of 0.0750. On the contrary, if he holds the optimal tracking portfolio obtained with $\alpha = 0.06$, he only endures CVaR risk of 0.06. In the index tracking problems, CVaR constraints can really reduce the risk of optimal tracking portfolios when risk tolerance is properly set.

Figure 1 gives the curves of tracking errors, CVaR risk of the optimal tracking portfolios when the values of risk tolerance $\alpha$ are changed from 0.055 to 0.1 for the in-sample case with $K = 8$. This figure reveals that the reducing the value of risk tolerance $\alpha$ causes the increase of tracking error and the decrease of CVaR risk. This is an evident consequence of the fact that reducing the value of risk tolerance $\alpha$ diminishes the feasible region of the problem. Figures 2 and 3 give return rates of the index and the optimal tracking portfolio obtained at different time in the in-sample for the cases $\alpha = 0.1$ (the CVaR constraint is inactive) and $\alpha = 0.06$ (the CVaR constraint is active), respectively. These figures show that the optimal tracking portfolio with the CVaR constraint inactive fits the index better, but the

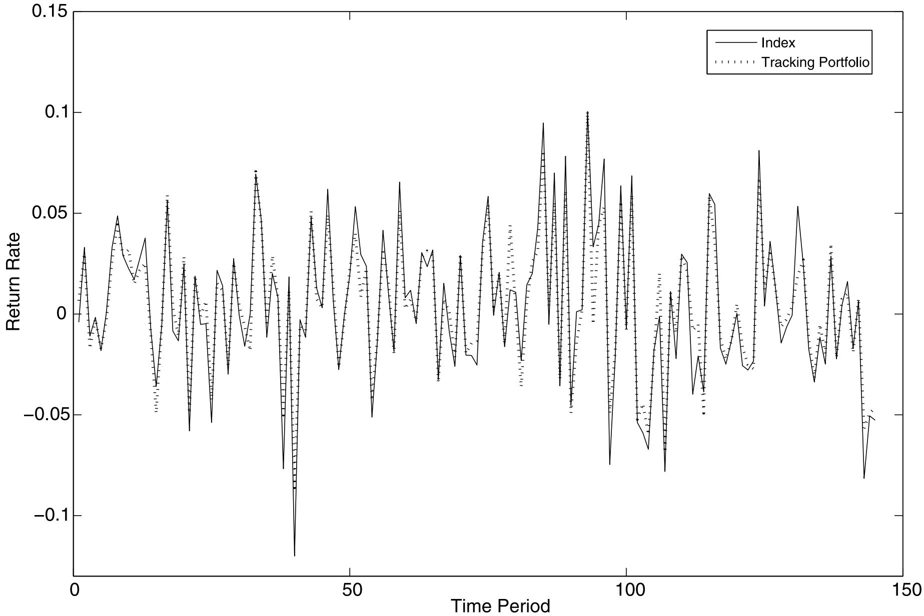


**Fig. 2** In-sample return rates of index and tracking portfolio, CVaR constraint is active ($\alpha = 0.06$) (HS)

corresponding downside risk is larger than the case of the CVaR constraint active. This fact can also be observed from Table 1.

*Results in out-of-sample* The optimal tracking portfolio obtained from in-sample data is used to track the HS index in the out-of-sample. Table 2 presents the tracking errors and CVaR risks of the optimal tracking portfolio obtained from the models with and without the CVaR constraint for different values of the cardinality $K$ and two values of risk tolerance $\alpha$. Since the optimal portfolios are the same obtained from both the model without the CVaR constraint and with the CVaR constraint when $\alpha = 0.1$, the results for both the cases are the same, and then the results with $\alpha = 0.1$ are not given. The results for the optimal portfolios obtained from the model (17) with $\alpha = 0.1$ and $\alpha = 0.06$ are similar to the results of *in-sample*, that is, the tracking errors with $\alpha = 0.1$ is smaller than the tracking errors with $\alpha = 0.06$, while the CVaR risk with $\alpha = 0.1$ is larger than the CVaR risk with $\alpha = 0.06$. The results from both the in-sample and out-of-sample show that the CVaR constraint controls the downside risk of tracking portfolios effectively.

Figure 4 gives curves of tracking errors, CVaR risks of the optimal tracking portfolio with the changes of risk tolerance $\alpha$ and $K = 8$ for the out-of-sample data. Since optimal tracking portfolios in in-sample data will not be always optimal in out-of-sample data, it happens that CVaR risk of tracking portfolio with large risk tolerance is smaller than the one with small risk tolerance in out-of-sample data. Figures 5 and 6 give return rates of the index and the optimal tracking portfolio obtained at different time in the out-of-sample for the cases $\alpha = 0.1$ (the CVaR constraint is inactive) and $\alpha = 0.06$ (the CVaR constraint is active), respectively. The same conclusions as those in Figs. 1, 2 and 3 can be obtained, and the detailed discussion is omitted.

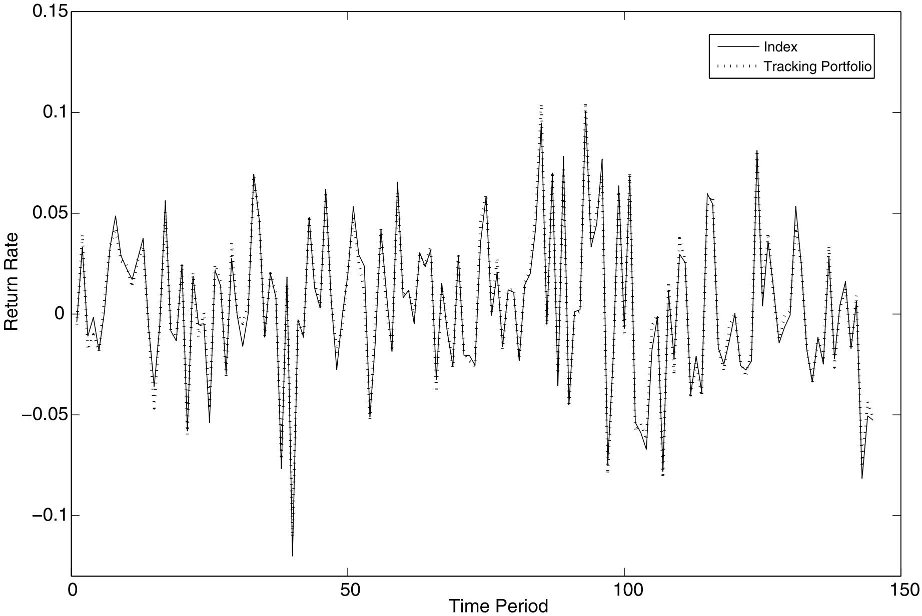


**Fig. 3** In-sample return rates of index and tracking portfolio, CVaR constraint is inactive ($\alpha = 0.1$) (HS)

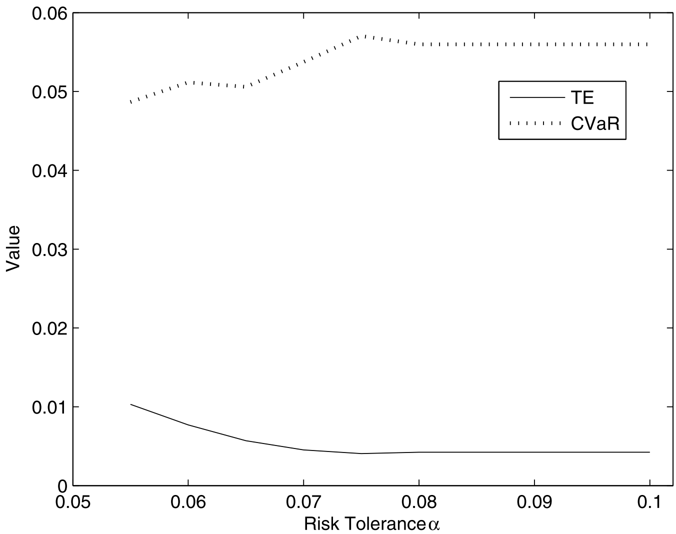


**Fig. 4** Out-of-sample TE, CVaR for various risk tolerance levels $\alpha$ in CVaR constraint (HS)

## 4.2 Results for FTSE index

As described in the last paragraph, problems (7) and (17) for FTSE index tracking are solved using the hybrid genetic method based on the approach in Ruiz-Torrubiano and Suárez (2009). For each problem, results are obtained over 10 executions with different random

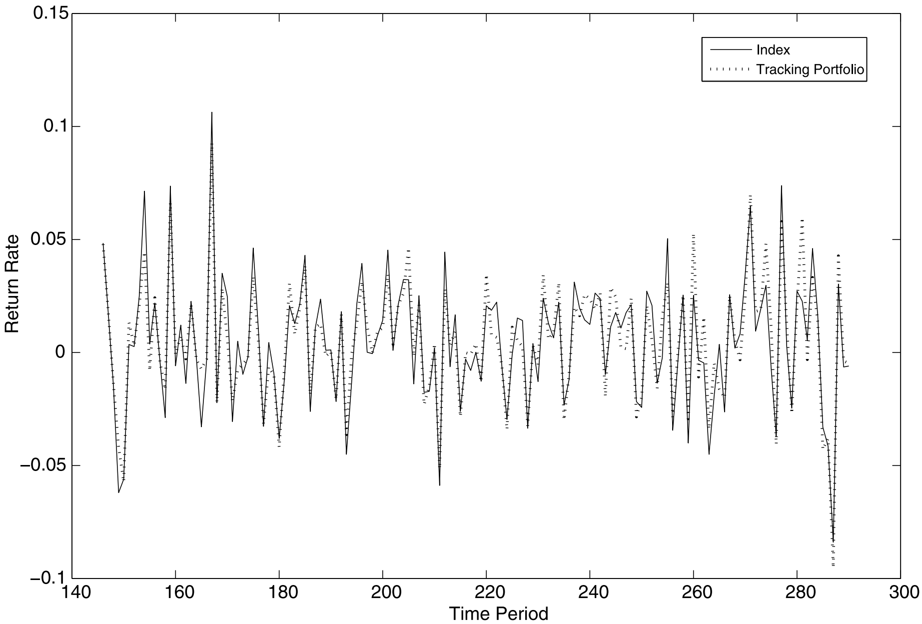


**Fig. 5** Out-of-sample return rates of index and tracking portfolio, CVaR constraint is active ($\alpha = 0.06$) (HS)

**Table 3** In-sample: "No CVaR constraint" versus "With CVaR constraint" (FTSE)

| Card. | No CVaR constraint | | | With CVaR risk tolerance $\alpha = 0.1$ | | | With CVaR risk tolerance $\alpha = 0.03$ | | |
|---|---|---|---|---|---|---|---|---|---|
| | TE | CVaR | Time (s) | TE | CVaR | Time (s) | TE | CVaR | Time (s) |
| $K = 5$ | 6.176e-3 | 0.0375 | 118.27 | 6.176e-3 | 0.0375 | 140.91 | 8.193e-3 | 0.0300 | 244.36 |
| $K = 6$ | 5.380e-3 | 0.0337 | 135.37 | 5.380e-3 | 0.0337 | 168.86 | 7.674e-3 | 0.0300 | 234.19 |
| $K = 7$ | 4.803e-3 | 0.0361 | 184.82 | 4.803e-3 | 0.0361 | 185.76 | 6.962e-3 | 0.0300 | 271.15 |
| $K = 8$ | 4.234e-3 | 0.0354 | 222.76 | 4.234e-3 | 0.0354 | 230.75 | 6.335e-3 | 0.0300 | 292.65 |
| $K = 9$ | 3.863e-3 | 0.0363 | 258.91 | 3.863e-3 | 0.0363 | 266.42 | 6.027e-3 | 0.0300 | 312.86 |
| $K = 10$ | 3.573e-3 | 0.0359 | 259.13 | 3.573e-3 | 0.0359 | 257.84 | 5.905e-3 | 0.0300 | 335.82 |

initial populations, and the best solution is used as the optimal tracking portfolio. The hybrid genetic method is also used to solve problems (7) and (17) for Hang Seng index tracking with the same initial values as used in CPLEX. The same results are obtained by the two methods when parameter values in the models are the same.

Table 3 presents the results for the cases of "No CVaR constraint" and "With CVaR constraint" for in-sample calculations. It can be observed from Table 3 that the performance of the optimal tracking portfolios obtained from the models (7) and (17) for FTSE index are similar to those obtained from models (7) and (17) for Hang Seng index, that is, the results for both the model (7) and the model (17) with $\alpha = 0.1$ are the same, and difference occurs when the value of $\alpha$ in model (17) is set to 0.03. For the case of "With CVaR constraint" when risk tolerance $\alpha = 0.1$, the CVaR constraint is inactive, and the optimal tracking portfolio is identical to the one obtained from model (7). When risk tolerance is $\alpha = 0.03$, the

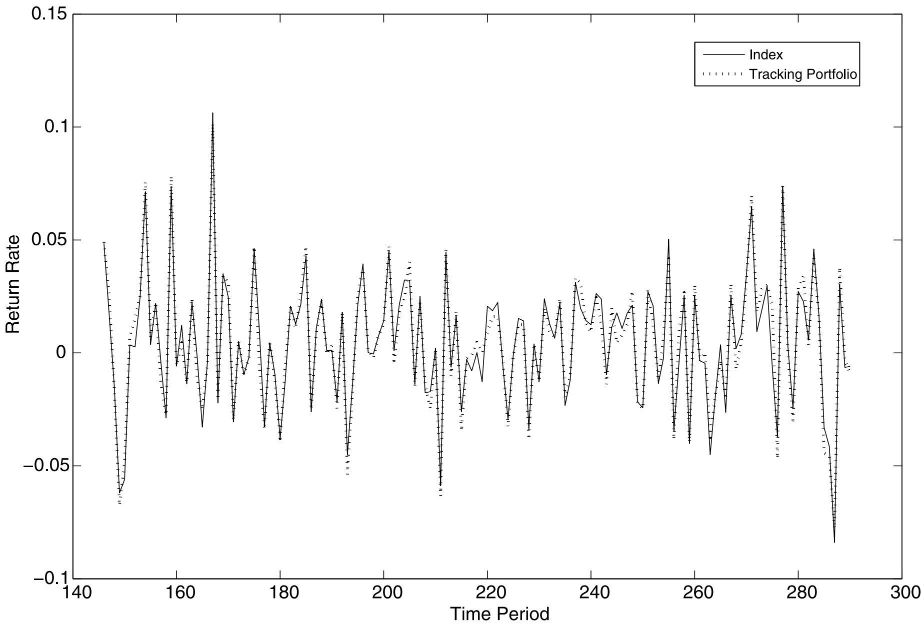


**Fig. 6** Out-of-sample return rates of index and tracking portfolio, CVaR constraint is inactive ($\alpha = 0.1$) (HS)

**Table 4** Out-of-sample: "No CVaR constraint" versus "With CVaR constraints" (FTSE)

| Card. | No CVaR constraints | | With CVaR risk tolerance $\alpha = 0.03$ | |
|---|---|---|---|---|
| | TE | CVaR | TE | CVaR |
| $K = 5$ | 8.678e-3 | 0.0270 | 8.490e-3 | 0.0252 |
| $K = 6$ | 8.248e-3 | 0.0241 | 6.911e-3 | 0.0307 |
| $K = 7$ | 7.123e-3 | 0.0274 | 6.745e-3 | 0.0290 |
| $K = 8$ | 6.955e-3 | 0.0249 | 6.832e-3 | 0.0274 |
| $K = 9$ | 6.950e-3 | 0.0261 | 6.708e-3 | 0.0260 |
| $K = 10$ | 6.632e-3 | 0.0259 | 6.866e-3 | 0.0243 |

CVaR constraint is active. Due to the reduction of the feasible region, the value of tracking error has an increase, but the CVaR risk of the optimal tracking portfolio is reduced.

Table 4 gives the performance of the optimal tracking portfolios obtained from the model (7) and the model (17) with $\alpha = 0.03$. Since the optimal tracking portfolio obtained from the model (17) with $\alpha = 0.1$ is same as the optimal tracking portfolio obtained from model (7), the results of the optimal tracking portfolio obtained from model (17) with $\alpha = 0.1$ are not given. It should be noticed that although the values of tracking errors in the case of "No CVaR constraint" are smaller than the those in the case of "CVaR constraints active" ($\alpha = 0.03$) for in-sample data (Table 3), the values of tracking errors for out-of-sample data are not in the same position. For $K = 5, 9$, both the values of tracking errors and CVaR risks in the CVaR constraint active case ($\alpha = 0.03$) are smaller than those in the "No CVaR constraint" (also the CVaR constraint inactive case). Thus, introducing a CVaR constraint into the index tracking problem is prone to have good performance in out-of-sample data sets for FSTE index. Figure 7 gives the return rates of the index and the optimal tracking

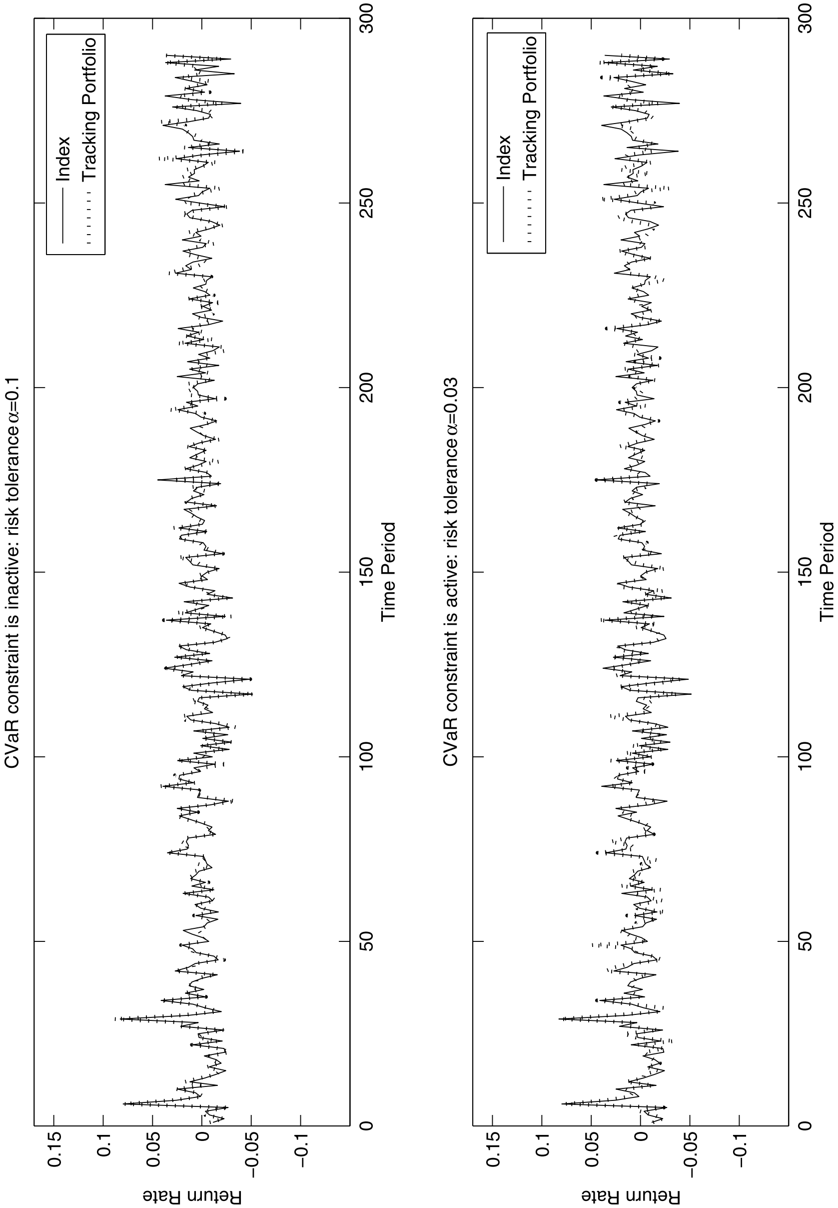


**Fig. 7** Return rates of index and tracking portfolio, CVaR constraint is inactive ($\alpha = 0.1$) and active ($\alpha = 0.03$) (FTSE)

portfolio with $K = 10$ when the CVaR constraints is active ($\alpha = 0.03$) and inactive ($\alpha = 0.1$) over the whole time period.

The index tracking problems can be efficiently solved by either CPLEX when the number of 0–1 variables is small or the hybrid genetic approach. It shows from the tests of both the Hang Seng index tracking and FTSE 100 index tracking that adding the CVaR constraint into the index tracking problems is helpful, and does not increase the difficulty and complexity to solve the resulting index tracking models. When the CVaR constraint is added into the general index tracking problems and the parameter values are properly set, the downside risk of the resulting optimal tracking portfolio can be controlled. Of course, adding the CVaR constraint may cause the increase of tracking errors. Thus, the trade-off between the increase of tracking error and the decrease of the downside risk of the optimal tracking portfolio should be carefully considered in practical application.

## 5 Conclusions and future researches

The index tracking problem is considered in this paper. A CVaR risk constraint is introduced into general index tracking model to limit the downside risk of the resulting optimal tracking portfolio. The aim of the proposed index tracking model is to minimize the tracking error between the tracking portfolio and a specified stock market index with a CVaR risk constraint and other practical market constraints. The tracking portfolio consists of a subset of component stocks in index, and the number of component stocks in the subset is specified by investors. The main contribution of this work is to formulate the index tracking problem with the CVaR constraint in a mixed 0–1 linear programming problem that can be effectively solved by a standard 0–1 LP solver such as CPLEX. Analysis shows that adding a CVaR constraint into the general index tracking model is helpful in the sense that adding the CVaR constraint will have no impact on the optimal tracking portfolio when the index has good (return increasing) performance, but can limit the downside risk of the optimal tracking portfolio when index has bad (return decreasing) performance. The software CPLEX is used to solve the resulting mixed 0–1 linear programming problem when the number of 0–1 variables is small, and a hybrid genetic approach is suggested to solve the resulting mixed 0–1 linear programming problem when the number of 0–1 variables is relatively large. The approach has shown good performance in index tracking and portfolio optimization problems in previous work.

Numerical tests on Hang Seng index and FTSE 100 index show that the proposed model is effective, and the introduced CVaR constraint does limit the downside risk of the resulting optimal tracking portfolios. However, it should be noticed that adding the CVaR constraint will relatively increase the tracking error of the optimal portfolio. Thus the trade-off between reducing the downside risk and increasing the tracking error of the optimal tracking portfolio should be considered by investors when the proposed index tracking model is applied. This can be done by setting proper parameter values in adding the CVaR constraint.

Additionally, theoretical analysis and performance tests show that the optimal tracking portfolio obtained from in-sample data may not be optimal for out-of-sample data. It is meaningful to design an approach to find a tracking portfolio which performs well also in out-of-sample data, since the main concern for the investor is the performance of the tracking portfolio in future. This is under our current investigation.

**Acknowledgements** This research is supported by National Natural Science Foundations, P.R. China (Grant No. 10971162, 11101325 and 71171158).

## Appendix: A hybrid genetic method for index tracking problem

Ruiz-Torrubiano and Suárez (2009) propose a hybrid optimization approach for index tracking problem, where the tracking error is measured by mean square deviation between the returns of the tracking portfolio and of the index. The resulting problem is a mixed 0–1 quadratic programming problem. The approach has been tested on different practical data sets. It can effectively solve index tracking problems. A hybrid genetic method based on the approach is applied to solve the mixed 0–1 linear programming problem generated from the index tracking problem with CVaR constraint. In the following, we summarize the hybrid genetic method:

1. Initialization.
   Generate initial population of $P$ individuals (candidate solutions), where each individual is a portfolio containing $K$ randomly selected stocks. When the stocks in the tracking portfolio is fixed, the problems (7) and (17) with cardinality constraint deleted are linear programs which can be easily solved to obtain an optimal portfolio with $K$ specified component stocks by optimization software such as CPLEX and Matlab. Then the fitness of the optimal portfolio is calculated. The choice of the fitness function is crucial in the design of a genetic algorithm. In the proposed hybrid genetic method, the fitness is set as the negative value of the tracking error (i.e., $-TE$ in problem (7) or (17)).
2. Evaluation.
   Evaluate the fitness of individuals in the population. If the population satisfy stop criterion, stop. Else, goto Step 3;
3. Selection.
   Select individuals for parents from the population;
4. Recombination.
   Recombine parents to produce new generation (children);
5. Mutation.
   Mutate the children with a given mutation probability and obtain a new child, then goto step 2.

The crossover operator (RAR operator) given in Moral-Escudero et al. (2006) is applied in the recombination step. This strategy choice results in a genetic algorithm with high evolutionary pressure which has shown good performance in index tracking and portfolio optimization problems.

## References

Artzner, P., Delbaen, F., Eber, J. M., & Health, D. (1999). Coherent measure of risk. *Mathematical Finance*, *9*, 203–228.

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

Canakgoz, N. A., & Beasley, J. E. (2009). Mixed-integer programming approaches for index tracking and enhanced indexation. *European Journal of Operational Research*, *196*, 384–399.

Clarke, R. C., Krase, S., & Statman, M. (1994). Tracking error, regret and tactical asset allocation. *The Journal of Portfolio Management*, *20*, 16–24.

Coleman, T. F., Henninger, J., & Li, Y. (2006). Minimizing tracking error while restricting the number of assets. *The Journal of Risk*, *8*, 33–56.

Krink, T., Mittnik, S., & Paterlini, S. (2009). Differential evolution and combinatorial search for constrained index-tracking. *Annals of Operational Research*, *172*(1), 153–176.

Kwiatkowski, J. W. (1992). Algorithms for index tracking. *IMA Journal of Mathematics Applied in Business & Industry*, *4*, 279–299.

Lobo, M. S., Fazel, M., & Boyd, S. (2007). Portfolio optimization with linear and fixed transaction costs. *Annals of Operational Research*, *152*(1), 341–365.

Markus, R., Wolter, H. J., & Zimmermann, H. (1999). A linear model for tracking error minimization. *Journal of Banking & Finance*, *23*, 85–103.

Moral-Escudero, R., Ruiz-Torrubiano, R., & Suarez, A. (2006). Selection of optimal investment portfolios with cardinality constraints. In *Proceedings of the IEEE congress on evolutionary computation 2006* (pp. 2382–2388), 16–21 July 2006.

Ogryczak, W., & Ruszczyński, A. (2002). Dual stochastic dominance and related mean-risk models. *SIAM Journal on Optimization*, *13*, 60–78.

Pflug, G. (2000). Some remarks on the value-at-risk and conditional value-at-risk. In S. Uryasev (Ed.), *Probilistic constrained optimization: methodology and application*. Dordrecht: Kluwer Academic.

Rockafellar, R. T., & Uryasex, S. (2000). Optimization of conditional value-at-risk. *The Journal of Risk*, *2*, 21–41.

Rockafellar, T. R., & Uryasex, S. (2002). Conditional value-at-risk for general loss distributions. *Journal of Banking & Finance*, *26*, 1443–1471.

Roll, R. (1992). A mean-variance analysis of tracking error. *Journal of Portfolio Management*, *18*, 13–22.

Ross, C., Terlaky, T., & Vial, J. Ph. (1986). *Theory and algorithm for linear optimization: an interior point approach*. Chichester: Wiley.

Ruiz-Torrubiano, R., & Suárez, A. (2009). A hybrid optimization approach to index tracking. *Annals of Operational Research*, *166*, 57–71.

Shapcott, J. (1992). *Index tracking: genetic algorithms for investment portfolio selection* (Technical report, EPCC-SS92-24). Edinburgh, Parallel Computing Centre.

Sharpe, W. F. (1971). A linear programming approximation for the general portfolio analysis problem. *Journal of Financial and Quantitative Analysis*, *6*, 1263–1275.

Ye, Y.Y. (1997). *Interior point algorithms: theory and analysis*. New York: Wiley.