Optim Eng (2017) 18:443–466
DOI 10.1007/s11081-016-9314-5

# Factor-based robust index tracking

Roy H. Kwon[1] · Dexiang Wu[1]

Received: 30 May 2015 / Revised: 4 February 2016 / Accepted: 19 February 2016 /
Published online: 2 April 2016


**Abstract** We consider a robust optimization approach for the problem of tracking a benchmark portfolio. A strict subset of assets are selected from the benchmark such that the expected return is maximized subject to both risk and tracking error limits. A robust version of the Fama-French 3 factor model is developed whereby uncertatiny sets for the expected return and factor loading matrix are generated. The resulting model is a mixed integer second-order conic problem. Computational results in tracking the S&P 100 out of sample show that the robust model can generate tracking portfolios that have better tracking error and Sharpe ratio than those generated by the nominal model.

**Keywords** Index tracking · Uncertainty · Robust optimization · Factor model

## 1 Introduction

Index tracking is an important passive investment strategy where one seeks a portfolio of securities that emulates a given benchmark portfolio such as the S&P 500. Several studies have concluded that actively managed funds usually cannot outperform broad market indices. Zenios (2006) reported that the average return of 769 all-equity actively managed funds was 2–5 % lower than the S&P 500 index during the period 1983–1989. More recently, SPs (2015) has reported that from the 5 and 10 years before December 31, 2014 more than 88 and 82 % of actively

✉ Roy H. Kwon
rkwon@mie.utoronto.ca

Dexiang Wu
dexter@mie.utoronto.ca

[1] Department of Mechanical and Industrial Engineering, University of Toronto, Ontario M5S 3G8, Canada

managed large cap funds were outperformed by the S&P 500, respectively. Thus, following an index fund can represent a compelling investment strategy. Full replication of a benchmark portfolio is an obvious strategy for tracking where all assets in the benchmark are held in the quantities as specified by the weightings of the benchmark portfolio, but full replication is not practical given the transaction costs this would entail. An alternative strategy is to select a strict subset of assets from the benchmark, however, this results in tracking portfolios that do not match the benchmark as closely as in full replication. A well-known measure of this discrepancy is called tracking error and is defined as the variance of the difference between returns of the tracking portfolio and benchmark. Models that seek to minimize tracking error have emerged as a popular approach for constructing tracking portfolios (Jorion 2003). However, some investors may wish to base a portfolio strategy on tracking a certain benchmark, but may also desire to outperform it. Any such outperformance will come a cost of increased risk and so a trade-off must be made as between how closely a benchmark is to be tracked and the risk associated with additional expected return.

In this paper, we consider an alternative approach where one seeks to maximize the expected return of a tracking portfolio subject to a limit on tracking error. Thus, the formulation we consider allows deviation from portfolios with minimum tracking error enabling a tracking portfolio to possibly outperform a benchmark in terms of expected return. To limit risk in this setting a constraint on portfolio risk is imposed see Cornuejols and Tutuncu (2006). A cardinality constraint is also considered to explicitly allow for the control of the number of assets to be in a tracking portfolio. The resulting nominal model is a mixed integer optimization problem with quadratic constraints. The objective function is linear and so the model can be converted to a mixed integer second-order cone problem (MISOCP).

The nominal tracking model requires a model of risk in order to obtain essential parameters such as expected returns and covariances of the assets. However, numerous research has shown that portfolio optimization, in particular, mean-variance optimization, generates portfolios that are sensitive to estimation error of the required inputs. Tutuncu and Koenig (2004) demonstrated that the efficient frontiers under nominal inputs can be drastically changed within only 5 percentiles for means of monthly log-returns and co-variances of these returns. Chopra and Ziemba (1993) showed that optimal mean-variance portfolios are an order of magnitude more sensitive to estimation errors in the expected returns than errors in co-variances. Our nominal tracking model can be seen as a natural extension of the mean-variance optimization (MVO) portfolio model, and thus optimal tracking portfolios can be sensitive to estimation errors. Therefore, we develop a robust optimization version of our nominal tracking model using a robust factor model approach as in Goldfarb and Iyengar (2003). In robust optimization, parameters that are deemed random or noisy have uncertainty sets created for them. An uncertainty set represents the possible values that a noisy parameter can take on. The optimization principle is then to maximize the worst case benefit (or minimize the maximum possible cost) over all of the various instances of parameters from their uncertainty sets (see Ben-Tal and Nemirovski 1998, 2000). Typically, the number of possible parameter values in an uncertainty set is uncountable. Factor models are an important framework for estimating crucial parameters for portfolio optimization such as expected return and covariance of assets and have been used successfully in the portfolio management industry to not only to obtain parameters for portfolio construction, but test and find factors that explain the cross-section of expected returns of random assets. In the context of index tracking factor models can provide the parameters for accurate risk and reward and tracking error estimation of various tracking portfolios. We employ the well-known Fama-French 3 factor model (Fama and French 1993).

In addition to robust optimization, there are other well-known optimization frameworks to mitigate parameter uncertainty. Earlier robust optimization frameworks as in Mulvey et al. (1995) use scenarios to represent uncertainties in parameter values in a goal programming framework. The aim in this approach to robustness is to generate a sequence of solutions that are progressively less sensitive to data as represented by scenarios. Stochastic programming is a special case of the robust optimization framework in Mulvey et al. (1995) and has been a popular framework to model financial optimization problems where the emphasis is on the construction of here-and-now non-anticipative decisions (see Birge and Louveaux 2011). All of these approaches seek to find solutions immune to uncertainty in parameter values proactively by employing scenarios to represent uncertainty. However, this requires the probabilities of the realizations of the scenarios. In our approach, the robust optimization is data-driven in that the structures that capture randomness i.e. the uncertainty sets can be derived from raw data such as monthly asset returns and so the estimation challenges associated with deriving probability measures is avoided.

We follow as in Goldfarb and Iyengar (2003) in the construction of uncertainty sets for expected returns and factor loadings with the covariance of asset returns assumed to be fixed. Then, the resulting factor-based robust tracking model can also be converted to a MISOCP. Computational results in tracking the S&P 100 out of sample show that the robust model can generate tracking portfolios that have better tracking error and Sharpe ratio than those generated by the nominal model. The main contribution of this paper is development of a robust factor-based enhanced-index tracking model with cardinality constraints.

The rest of the paper is organized as follows: a review of relevant literature in optimization in finance and in particular index tracking is given in Sect. 2. In Sect. 3 we formulate the nominal index tracking models and its robust counterpart. In Sect. 4 we compare the portfolio performance between the models, and we conclude the paper in Sect. 5.

## 2 Literature review

Robust optimization has been considered in many applications to mitigate the effects of parameter uncertainty. A comprehensive survey (over 130 references) of robust optimization is given in Bertsimas et al. (2011). The authors listed several important applications in finance, which include multi-period asset allocation problems as in Ben-Tal et al. (2000) where the authors propose a second-order cone program as a robust counterpart, and Bertsimas and Pachamanova (2008) where under specific norms the problem is cast as a linear program. Goldfarb and Iyengar (2003) consider robust mean-variance optimization formulations based on robust factor models and show that the resulting robust problems can be formulated as SOCPs. Erdogan et al. (2004) incorporate transaction costs into the robust MVO problems and the resulting model remains as an SOCP.

Cardinality restrictions are an important construction in reducing the complexity of portfolio management and is also important in the construction of tracking portfolios. Cardinality constraints in deterministic MVO has been considered by several authors e.g. Chang et al. (2000). Some researchers have considered cardinality constraints in robust MVO models. Sadjadi et al. (2012) apply robust optimization to cardinality constrained mean-variance problem which results in a mixed-integer second-order cone programming and applied genetic algorithms to compute solutions. Gulpinar et al. (2011) also considers robust cardinality constrainted MVO problems and solves the resulting mixed-integer SOCP instances using a commercial solver.

Corielli and Marcellino (2006) presented a two-step index tracking procedure based on factor analysis and illustrated their method through Monte Carlo simulations. However, it is not clear whether practical constraints such as transaction costs can be incorporated in their process. Chen and Kwon (2012) proposed a linearly robust index tracking model that does not require the first moment information between the assets. Karlow and Rossbach (2011) applied a VaR constraint to the tracking error term, and added a regularization term into objective instead of using a cardinality constraint. Chen et al. (2013) used an additional quasi-norm regularization term in the objective function to approximate the portfolio size. However, norm regularization cannot control the portfolio size exactly, and the portfolio may concentrate in one or a few assets. Models that minimize tracking error are considered by Jorion (2003). An enhanced index tracking model with tracking error constraints is considered by Cornuejols and Tutuncu (2006). In this model, the user can set a tracking error limit so that one can generate expected returns that can exceed that of the benchmark, but with possibly increased risk and so the model also incorporates a constraint limiting portfolio risk. Kolbert and Wormald (2010) considers a single factor model for index tracking with a constraint on the norm of tracking error which results in a SOCP, which can be solved by SeDuMi efficiently.

The model we propose in this paper extends the work in Cornuejols and Tutuncu (2006) in that we incorporate cardinality constraints into a robust version based on a robust 3 factor model of Fama and French. We believe this model to be the first factor-based robust enhanced index tracking problem in the literature. Other enhanced index tracking models with cardinality constraints include those by Canakgoz and Beasley (2009) who consider a deterministic mixed integer programming approach. Lejeune and Samatlı-Paç (2013) consider a chance-constrained stochastic integer programming approach for enhanced indexation.

## 3 Robust factor model to index tracking

### 3.1 Nominal index tracking model

In this section we develop the nominal enhanced index tracking model. Let $\mu$ denote the vector of expected returns of assets and $\Sigma$ the covariance matrix, $x$ is vector of portfolio weights, $x_{BM}$ is the vector of weights of a benchmark index. Then the difference in expected returns (or excess returns) between the tracking portfolio and the benchmark is $\mu^T(x - x_{BM})$, and the standard deviation of excess returns (tracking error) is $\sqrt{(x - x_{BM})^T\Sigma(x - x_{BM})}$. Here we adopt the index tracking model from Cornuejols and Tutuncu (2006). In this formulation, a portfolio $x$ is sought that maximizes expected return subject to a limit on portfolio risk and tracking error. The model is given as:

$$\max \quad \mu^T x \tag{1}$$

$$\text{s.t.} \quad \left\|\Sigma^{\frac{1}{2}}x\right\| \le \sigma \quad \text{portfolio risk} \tag{2}$$

$$\left\|\Sigma^{\frac{1}{2}}(x - x_{BM})\right\| \le TE \quad \text{tracking error} \tag{3}$$

$$e^T x = 1 \tag{4}$$

$$x \ge 0 \tag{5}$$

where $e \in R^n$ is a vector all of whose components equal 1, $\|\bullet\|$ denotes the norm value of a vector. $\sigma$ denotes the tracking portfolio risk limit and $TE$ is the tracking error limit. For example, $TE = 5\,\%$ means that a tracking portfolio may not have standard deviation of excess returns of more than 5 %. A cardinality constraint can be added into models (1)–(5) to control the portfolio size exactly:

$$\max \quad \mu^T x \tag{6}$$

$$\text{s.t.} \quad \left\|\Sigma^{\frac{1}{2}}x\right\| \le \sigma \tag{7}$$

MISOCP

$$\left\|\Sigma^{\frac{1}{2}}(x - x_{BM})\right\| \le TE \tag{8}$$

$$e^T x = 1 \tag{9}$$

$$e^T y = q \quad \text{size tracking portfolio} \tag{10}$$

$$lb_i y_i \le x_i \le ub_i y_i, \ \forall i \quad \text{band on the weight} \tag{11}$$

$$x \ge 0, y \in \{0, 1\} \tag{12}$$

where *lb*, *ub* are the lower and upper bounds on the tracking portfolio weights, $q$ is the portfolio size. Models (6)–(12) can eaily be seen to be a mixed-integer second-order cone programming (MISOCP) as the risk and tracking error constraints are quadratic with all other constraints linear and a linear objective function and binary integer restrictions.

Next, we illustrate the nominal model by solving several instances. All instances were solved to optimality by using the mixed integer solver in Gurobi which is based on branch-and-bound to obtain zero gap between lower and upper bounds (Gurobi Optimization 2015). In particular, the effect of the tracking error constraint (8) and the risk control constraint (7) are investigated by repeatedly solving the model with increasing values for the parameter $TE$ under different $\sigma$ value. We used daily returns from June 30, 2005 to December 31, 2007 to generate the parameters $(\mu, \Sigma)$ for the model. We fix $\sigma$ with a large value, e.g. $\sigma = 80 \times \max(\mathrm{diag}(\Sigma))$, then increase $TE$ with a given portfolio size, then we change $\sigma$ to a smaller value, e.g. $\sigma = 0.004 \times \max(\mathrm{diag}(\Sigma))$, and repeat the same computational process by changing $TE$ value. The parameters $(\mu, \Sigma)$ are estimated through linear regression, specifically a three-factor model is applied for our estimation (see Sect. 4). We computed instances over different $q$ sizes that represent low, medium and high density portfolios. For example, we chose the portfolio size from $q \in [5, 65]$ as we found that tracking portfolios will be very close to the index when $q$ over 75, but this will generate higher transaction costs due to holding more assets. We compare the portfolio return, variance and Sharpe ratio with different portfolio tracking sizes $q$ in Figs. 1, 2 and 3.

Figure 1 shows the portfolio return over $TE$ and $q$ under different $\sigma$. Moderate tracking portfolio sizes ($q = 15, 35$) have higher return than larger or smaller sizes ($q = 65$ or 5). When $q = 65$ the effects of diversification become stronger reducing return. From the sub-figures we see that the portfolio return sublinearly increases which means the marginal return decreases with respect to $TE$ value. However, the portfolio returns are generally better than the return of benchmark used i.e. the S&P100 (0.35 ‱ in left sub-figure). As the parameter $\sigma$ decreased to $0.004 \times \max(\mathrm{diag}(\Sigma))$, i.e., from left to right in each sub-figure, the risk control

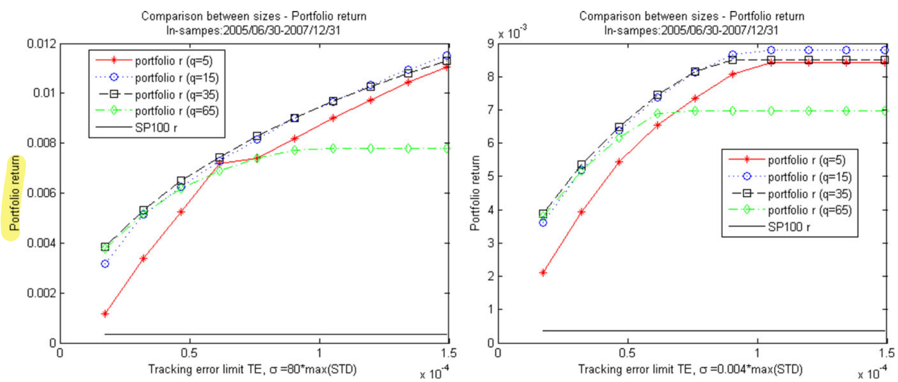


**Fig. 1** Portfolio return versus TE with different q under different $\sigma$ (SP100)

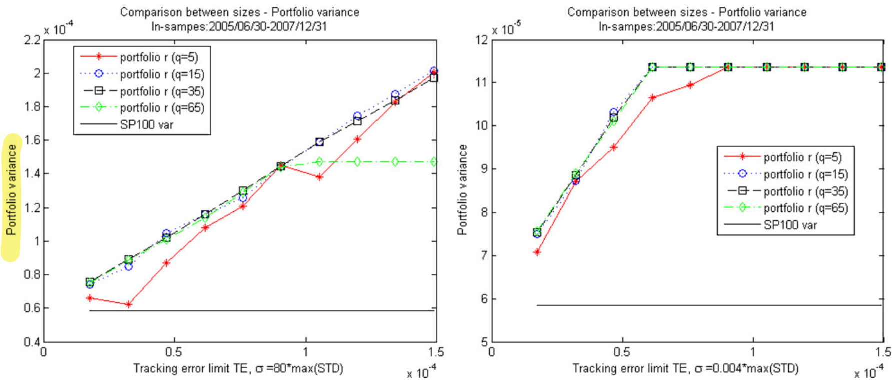


**Fig. 2** Portfolio variance versus TE with different q under different $\sigma$ (SP100)

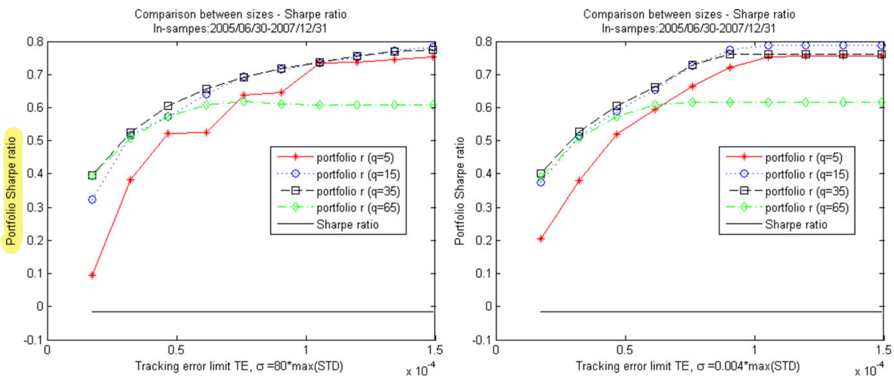


**Fig. 3** Portfolio Sharpe ratio versus TE with different q under different $\sigma$ (SP100)

constraint (7) dominates the tracking error constraint (8), and therefore the portfolio return cannot be improved via changing the *TE* value after $0.8 \times 10\text{e}^{-4}$.

The portfolio variance increases approximately linearly with respect to *TE*, see left side on Fig. 2, for moderate portfolio sizes. The tracking portfolio with size $q = 65$ has lower variance due to the diversification effects of having more assets. The results suggest that if larger *TE* values are allowed, the portfolio return can be improved, however, the portfolio variance may increase quicker than the improvement of return. The variance of S&P100 is lowest out all portfolios most likely due to the diversification effect from having more assets. In particular, the S&P100 variance is 0.06 ‰ and associated standard deviation is 0.24 ‰ in left sub-figure. Again we note that parameter $\sigma$ can also significantly impact the portfolio variance as seen on the right side on Fig. 2. The portfolio variance will be invariant to *TE* after $0.6 \times 10\text{e}^{-4}$ if $\sigma$ is set too small.

We combine the portfolio return on Fig. 1 and variance on Fig. 2 to obtain the portfolio Sharpe ratio on Fig. 3. From Fig. 3, the portfolio Sharpe ratio increases with respect to increasing *TE*, however, the marginal portfolio Sharpe ratio decreases since the marginal variance dominates the marginal return. Thus, by controlling the *TE* one can improve portfolio performance, but increasing *TE* as shown above can lead to increased portfolio volatility in Fig. 2. Thus, the one must be careful about setting *TE* and $\sigma$ too high if one cares about risk. Also, the results suggest that one way to help attain enhanced indexing is to not set $q$ too high. We see that the Sharpe ratio of the benchmark index S&P 100 ($-0.0047$ in left side and $-0.0049$ in right side) is worse than the Sharpe ratio of the tracking portfolios. From both sub-figures in Fig. 3, we see that the portfolio Sharpe ratios can be significantly impacted by both model parameters *TE* and $\sigma$.

In the next section we develop the robust counterpart to models (6)–(12) to errors in parameter estimation.

### 3.2 Robust multi-factor model for index tracking

We follow as in Goldfarb and Iyengar (2003) by employing a robust factor modeling approach. Suppose the return vector $r$ is given by the model:

$$r = \mu + V^T f + \epsilon$$

where $\mu \in R^n$ is the vector of mean returns, $f \sim N(0, F) \in R^m$, is the vector of returns of the factors that drive the market, $V \in R^{m\times n}$ is the matrix of factor loadings of the $n$ assets, and $\epsilon \sim N(0, D)$ is the vector of residual returns where $D = \mathrm{diag}(d)$, $d = [d_i]$, $i = 1, \ldots, n$. In practice we may need $F \succeq 0$. The assumptions for factor model include:

- residual returns $\epsilon_i$ and $\epsilon_j$ are independent, i.e. $\mathrm{cov}\left(\left[\epsilon_i \epsilon_j\right]\right) = 0$ for $i \neq j$;
- residual return $\epsilon_i$ and factor return $f_j$ are independent, i.e. $cov\left(\left[\epsilon_i\, f_j\right]\right) = 0$.

Then $\mathbb{E}(r) = \mu$, $\sigma_{ij} = V_i^T F V_j$, $i \neq j$, $\sigma_{ii} = \sigma_i^2 = V_i^T F V_i + d_i$, $\sigma_i = \sqrt{V_i^T F V_i + d_i}$, or written in matrix form $\Sigma = V^T F V + D$. Given a weight vector $x$, the risk of a portfolio, i.e. $x^T \Sigma x$, can be split as a combination of a systematic risk, i.e. $x^T V^T F V x$, and a individual risk, i.e. $x^T D x$, within a portfolio (Kolbert and Wormald 2010). Then building uncertainty sets around $\Sigma$ is equivalent to build the uncertainty sets for terms $V^T F V$ and $D$ separately. We assume that the market is stable, i.e. $F$ is constant, then generators of uncertainty for parameters $(\mu, \Sigma)$ comes from the generators of uncertainty for the parameters $(\mu, V, D)$. We follow as in Goldfarb and Iyengar (2003) and design the uncertainty sets for parameters $(\mu, V, D)$ separately as follows:

- The uncertainty sets $S_m$ and $S_d$ for parameters $D$ and $\mu$ are defined as intervals:

$$S_d = \left\{D : D = diag(d), d_i \in \left[\underline{d}_i, \overline{d}_i\right], i = 1, \ldots, n\right\} \tag{13}$$

$$S_m = \left\{\mu : \mu = \mu_0 + \xi, \xi_i \in \left[\underline{\gamma}_i, \overline{\gamma}_i\right], i = 1, \ldots, n\right\} \tag{14}$$

– The uncertainty set for parameter $V$ belongs to an ellipsoid:

$$S_v = \left\{ V : V = V_0 + W, \|W_i\|_g \leq \rho_i, i = 1, \ldots, n \right\} \tag{15}$$

where $W_i$ is the $ith$ column of strength matrix $W$ around $V_0$ and $\|W_i\|_g = \sqrt{W_i^T G W_i}$ is an elliptic norm, $G \succeq 0$ denotes the coordinate system that may not be perpendicular. We can always generate a matrix $G \succ 0$ to maintain the strict convexity of the problem.

Then the robust counterpart for objective (6):

la natura butta giù il stesso del rendimento

$$\max_x \min_{\mu \in S_m} \mu^T x = \max_x \min_{|\xi| \leq \gamma} (\mu_0 + \xi)^T x = \max_x \left( \mu_0 + \underline{\gamma} \right)^T x$$

For the constraint (7) that measure the portfolio risk:

$$\left\| \Sigma^{\frac{1}{2}} x \right\|_2^2 \leq \sigma^2 \iff x^T \Sigma x \leq \sigma^2 \iff x^T \left( V^T F V + D \right) x \leq \sigma^2 \iff x^T V^T F V x + x^T D x \leq \sigma^2$$

Then the robust counterpart for above constraint:

$$\max_{V \in S_v, D \in S_d} x^T V^T F V x + x^T D x \leq \sigma^2 \iff \max_{V \in S_v} x^T V^T F V x + \max_{D \in S_d} x^T D x \leq \sigma^2$$

$$\iff \begin{cases} \max\limits_{V \in S_v} x^T V^T F V x \leq v \\ \max\limits_{D \in S_d} x^T D x \leq \delta \\ v + \delta \leq \sigma^2 \end{cases} \iff \begin{cases} \max\limits_{V \in S_v} x^T V^T F V x \leq v \\ \left\| \begin{bmatrix} 2\overline{D}^{1/2} x \\ 1 - \delta \end{bmatrix} \right\| \leq 1 + \delta \\ v + \delta \leq \sigma^2 \end{cases}$$

We use the sum of $v$ and $\delta$ to represent the total risk since the terms $x^T V^T F V x$ and $x^T D x$ are independent. For the robust term $\max\limits_{V \in S_v} x^T V^T F V x \leq v$, Goldfarb and Iyengar ellissoide

Goldfarb and Iyengar (2003) show that it can be converted into a collection of linear and second-order conic constraints through Lemma 1 below.

**Lemma 1** *Let* $r, v > 0$, $y_0, y \in \mathbb{R}^m$ *and* $F, G \in \mathbb{R}^{m \times m}$ *be positive definite matrices. Then the constraint*

$$\max_{\{y : \|y\|_g \leq r\}} \|y_0 + y\|_f^2 \leq v \tag{16}$$

*is equivalent to either of the following:*

(i) *there exist* $\tau, \sigma \geqslant 0$, *and* $t \in \mathbb{R}_+^m$ *that satisfy*

$$v \geqslant \tau + e^T t$$
$$\sigma \leq \frac{1}{\lambda_{\max}(H)}$$
$$r^2 \leq \sigma\tau$$
$$w_i^2 \leq (1 - \sigma\lambda_i)t_i,\ i = 1, \ldots, m$$

*where* $Q\Lambda Q^T$ *is the spectral decomposition of* $H = G^{-1/2}FG^{-1/2}$ , $\Lambda = diag(\lambda_i)$, *and* $w = Q^T H^{1/2} G^{1/2} y_0$;

(ii) there exist $\tau \geqslant 0$, *and* $s \in \mathbb{R}^m_+$ *that satisfy*

$$r^2 \leq \tau\left(v - e^T s\right)$$
$$u_i^2 \leq (1 - \tau\theta_i)s_i,\ i = 1, \ldots, m$$
$$\tau \leq \frac{1}{\lambda_{\max}(K)}$$

*where* $P\Theta P^T$ *is the spectral decomposition of* $K = F^{1/2}G^{-1}F^{1/2}$, $\Theta = diag(\theta_i)$, *and* $u = P^T F^{1/2} y_0$.

Lemma 1 is proved using the $\mathbb{S}$-procedure which has broad application in engineering science Boyd and Vandenberghe (2004). For details of the proof of the Lemma 1 (see Goldfarb and Iyengar 2003). Therefore by using part (ii) of Lemma 1 constraint $\max_{V \in S_v} x^T V^T F V x \leq v$ can be transformed into the following convex constraint set:

$$\max_{V \in S_v} x^T V^T F V x \leq v \iff V \in S_v \max \|Vx\|_f^2 \leq v$$
$$\iff \begin{cases} u = P^T F^{1/2} V_0 x \\ \left\| \begin{bmatrix} 2\rho^T x \\ \tau - v + e^T s \end{bmatrix} \right\| \leq \tau + v - e^T s \\ \left\| \begin{bmatrix} 2u_i \\ v - \tau\theta_i - s_i \end{bmatrix} \right\| \leq v - \tau\theta_i + s_i, \forall i = 1, \ldots, m \\ v - \tau\lambda_{\max}(K) \geq 0 \\ \tau \geq 0 \end{cases} \tag{17}$$

where $K = P\Theta P^T$ is the spectral decomposition of $K = F^{1/2}G^{-1}F^{1/2}$, $\Theta = diag(\theta)$. Note that radius $r = \rho^T|x| = \rho^T x$ in the first norm constraint in (17) because short selling is prohibited, i.e. $x \geq 0$.

For the constraint (8) that measure the tracking error:

$$\left\|\Sigma^{\frac{1}{2}}(x - x_{BM})\right\| \le TE \iff (x - x_{BM})^T \Sigma (x - x_{BM}) \le TE^2 \iff \begin{cases} z^T \Sigma z \le TE^2 \\ z = x - x_{BM} \end{cases} \tag{18}$$

Analogously the robust counterpart of $z^T \Sigma z \le TE^2$ in (18) can be obtained by using Lemma 1. The associated convex constraints are constructed as follows:

$$\max_{V \in S_v, D \in S_d} z^T V^T F V z + z^T D z \le TE^2 \iff \max_{V \in S_v} z^T V^T F V z + \max_{D \in S_d} z^T D z \le TE^2$$

$$\iff \begin{cases} \max\limits_{V \in S_v} z^T V^T F V z \le l \\ \max\limits_{D \in S_d} z^T D z \le \zeta \\ l + \zeta \le TE^2 \end{cases} \iff \begin{cases} \max\limits_{V \in S_v} \|Vz\|_f^2 \le l \\ \left\| \begin{bmatrix} 2\overline{D}^{1/2} z \\ 1 - \zeta \end{bmatrix} \right\| \le 1 + \zeta \\ l + \zeta \le TE^2 \end{cases}$$

$$\iff \begin{cases} w = P^T F^{1/2} V_0 z \\ \left\| \begin{bmatrix} 2\rho^T |z| \\ \tau - l + e^T s \end{bmatrix} \right\| \le \tau + l - e^T s \\ \left\| \begin{bmatrix} 2w_i \\ l - \tau\theta_i - s_i \end{bmatrix} \right\| \le l - \tau\theta_i + s_i, \forall i = 1, \ldots, m \\ l - \tau\lambda_{\max}(K) \ge 0 \\ \tau \ge 0 \\ \left\| \begin{bmatrix} 2\overline{D}^{1/2} z \\ 1 - \zeta \end{bmatrix} \right\| \le 1 + \zeta \\ l + \zeta \le TE^2 \end{cases} \tag{19}$$

The absolute value sign in the radius $r = \rho^T |z| = \sum_{i=1}^n \rho_i |z_i|$ in the first norm constraint in (19) should be removed since variable $z$ could be negative. We replace $|z_i|$ as follows:

$$|z_i| = z_i^+ + z_i^-$$
$$z_i = z_i^+ - z_i^- = x_i - x_i^{BM}$$
$$z_i^+ \ge 0, z_i^- \ge 0$$

Finally the robust counterpart using the factor model for problems (6)–(12) can be formulated as follows:

$$\max \quad \left(\mu_0 + \underline{\gamma}\right)^T x \tag{20}$$

$$\text{s.t.} \quad u = P^T F^{1/2} V_0 x \tag{21}$$

$$w = P^T F^{1/2} V_0 (z^+ - z^-) \tag{22}$$

$$z^+ - z^- = x - x_{BM} \tag{23}$$

$$\left\| \begin{bmatrix} 2\rho^T x \\ \tau - v + e^T s \end{bmatrix} \right\| \leq \tau + v - e^T s \tag{24}$$

$$\left\| \begin{bmatrix} 2u_i \\ v - \tau\theta_i - s_i \end{bmatrix} \right\| \leq v - \tau\theta_i + s_i, \quad \forall i = 1, \ldots, m \tag{25}$$

$$\left\| \begin{bmatrix} 2\rho^T (z^+ + z^-) \\ \tau - l + e^T s \end{bmatrix} \right\| \leq \tau + l - e^T s \tag{26}$$

$$\left\| \begin{bmatrix} 2w_i \\ l - \tau\theta_i - s_i \end{bmatrix} \right\| \leq l - \tau\theta_i + s_i, \quad \forall i = 1, \ldots, m \tag{27}$$

$$\left\| \begin{bmatrix} 2\overline{D}^{1/2} x \\ 1 - \delta \end{bmatrix} \right\| \leq 1 + \delta \tag{28}$$

$$v + \delta \leq \sigma^2 \tag{29}$$

$$\left\| \begin{bmatrix} 2\overline{D}^{1/2} (z^+ - z^-) \\ 1 - \zeta \end{bmatrix} \right\| \leq 1 + \zeta \tag{30}$$

$$l + \zeta \leq TE^2 \tag{31}$$

$$v - \tau\lambda_{\max}(K) \geq 0 \tag{32}$$

$$l - \tau\lambda_{\max}(K) \geq 0 \tag{33}$$

$$e^T x = 1 \tag{34}$$

$$e^T y = q \tag{35}$$

$$lb_i y_i \leq x_i \leq ub_i y_i, \quad \forall i = 1, \ldots, n \tag{36}$$

$$x \geq 0, \tau \geq 0, z^+ \geq 0, z^- \geq 0 \tag{37}$$

$$y \in \{0, 1\} \tag{38}$$

The dimension of different type of variables are $x, z^+, z^- \in \mathbb{R}^n$, $v, \delta, l, \zeta, \tau \in \mathbb{R}$, $u, w, s \in \mathbb{R}^m$, $y \in \mathbb{B}^n$. Therefore, models (20)–(38) keeps the same CCCP structure as norminal tracking model but includes more variables and cone constraints. In practice we apply Fama-French 3 factor model which is an advanced extension of CAMP model to calculate the numerical values of the parameters, details (see Goldfarb and Iyengar 2003; Fama and French 1993).

We then test both nominal models (6)–(12) and the robust counterpart (20)–(38) by commercial solver Gurobi on an AMD Dual-Core laptop with 2GB of RAM. One interesting observation is that Gurobi take much longer running time to model (20 )–

(37) than that to model (6)–(12) in many instances. For example, for the case that $N = 500$, $q = 70$, there still exist 10 % gap between lower and upper bounds after 1000 seconds running time for models (20)–(38) while Gurobi return the optimal solution, i.e. gap equals 0, within 10 s for models (6)–(12), and such instances are common in our testing. The factored robust procedure for index tracking problem simplify the parameter estimation but it may increase the solving time since more conic constraints are included into the model ($2m + 2$ from 2), which make the problem harder and harder. This disadvantage motivated us to apply Lagrangian Relaxation method we designed to speed up the solving process and keep the quality of the solution in next section.

## 4 Computational experiments

We used as the basis of the robust factor model the Fama and French 3 factor model (Fama and French 1993) which can be seen as the extension of Sharpe's one factor CAPM model. We also note that any other notable multi-factor model that can better interpret risks can also used in the robust factor index tracking model. For example, D'ecclesia and Zenios (1994) showed that 98 % of the variability of bonds returns can be explained via multiple risk factors of returns of the Italian bond market and so a robust tracking problem for Italian bond indices can use this multi-factor model. Burmeister et al. (2003) presented a macroeconomic factor model which includes five risk terms in interpreting the historical stock returns. The Fama-French 3 factor model is based on the observation that small capitalization stocks and value stocks (i.e. stocks with high book to price ratio) tend to outperform the market as a whole. In the model, three risk factors reflect the sensitivities of each stock to the market excess return (market factor), the excess of value stocks over growth stocks (book-to-market factor), and the excess of small cap stocks over large cap stocks (size factor). Since the three factor model can explain more of the variability of excess returns than the single factor model, We applied the three factor model as the basis to form the uncertainty sets of expected return and covariance for the factor loadings in (20)–(38) as in Goldfarb and Iyengar (2003). The details of this construction is in Appendix 1. We did not test for stationarity of returns, and assumed that the market was stable i.e. $F$ covariance of factors, in the event of non-stationarity of variance this could affect the estimation of betas (factor loadings). However, our approach was to let this be handled by the robust optimization over different factor loading matrices captured by the uncertainty set $S_v$.

### 4.1 Index tracking using the S&P100 index

In this section, we illustrate the factor-based robust enhanced index tracking model by tracking the S&P100 index. Comparisons of the robust model versus the nominal model illustrate the benefits of robustness. First, in-sample data about the components S&P100 are collected to construct the nominal covariance matrix. We collected the historical price information of all components of S&P100, and calculated the daily return $r_{it} = \frac{P_{i,t} - P_{i,t-1}}{P_{i,t-1}}$, where $P_{i,t}, P_{i,t-1}$ are the adjusted closing prices at time $t$ and $t$-1. Then, daily returns were used to calculate the mean returns of assets and covariance matrix of returns of the assets:

$$\mu_i = \frac{1}{T}\sum_{t=1}^{T} r_{it}, cov_{ij} = \frac{1}{T}\sum_{t=1}^{T}(r_{it} - \mu_i)\left(r_{jt} - \mu_j\right)$$

Daily prices between June 30, 2005 and December 31, 2007 (630 samples) were collected and used for in-samples construction, and daily prices for each end of month between January 1, 2008 and December 31, 2008 were used to build out-of-sample data sets for the nominal and robust models. Some stocks in the S&P100 index can be replaced by some other stocks outside of the index since they may not satisfy the selection criteria of S&P100 in the designed time period, we retrieved the stocks that were moved out in the time periods used above and obtained the associated price information. Usually this replacement was rare and the components of S&P100 were stable, we check the changing history of the composition of the S&P100 and there is no replacement between June 30, 2005 and December 31, 2008, the period we collected data for. For some stocks if there is no adequate data from the Bloomberg work station, we deleted that assets from the index, for example, seven assets (5 % of total market value) were deleted in the period of year 2006–2007 and five assets (2 % of total market value) were deleted in the years of 2007 and 2008 due to lack of data, this reduction did not significantly impact the total market value of S&P100. Table 1 lists the tickers we used for our research grouped them across different sectors.

The Fama-French 3 factor model is used to generate the parameters $\mu_0$, $V_0$, $G$, $\rho_i$, $\gamma_i$, $d_i$ and the associated uncertainty sets for $\mu$ and $V_0$ see Appendix 1 for details on the construction and we set $\omega = 0.95$ which represents the joint confidence level. Figure 4 shows the worst bound for the expected return $\mu$ under the given uncertainty set (14) and the worst bound for covariance $\sigma_i$ under the given uncertainty set (13) and (15) by using in-samples from June 30, 2005 to December 31, 2007. We can see that almost all robust expected returns are below the nominal

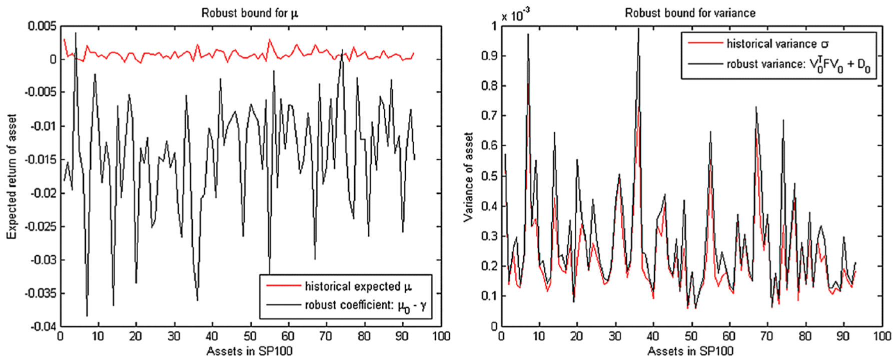


**Fig. 4** Robust bound for expected return and variance (SP100)

expected return from the historical data, and all robust covariances are above the nominal covariance computed from the historical data.

### 4.1.1 *Robust versus nominal portfolio performance*

We then use the computed tracking portfolios to test out-of-sample in a rolling fashion and compare the performance of the portfolios. The four rolling periods are 2008, 2009, 2010, and 2011, respectively. The rolling process is described as follows. We select two years daily data, e.g. year 2006 and 2007, as in sample data to construct the portfolio and then test the next one year performance, e.g. year 2008, without re-balancing. After that we use as in-sample daily data years 2007 and 2008, and test portfolio performance in year 2009, and so on. Both nominal models (6)–(12) and robust counterparts (20)–(38) are solved by Gurobi. For the initial test, we set $lb_i = \frac{1}{n}$ and $ub_i = 0.7$. $\sigma$ equals 8 times of the maximal standard deviation in the assets in SP100, and $TE$ equals 5 times of standard deviation of SP100. The portfolio size was set at $q = 25$.

Figure 5 shows the portfolio return evolution for the rolling out-of-sample periods, there is no rebalancing during the out-of-sample test. The returns from

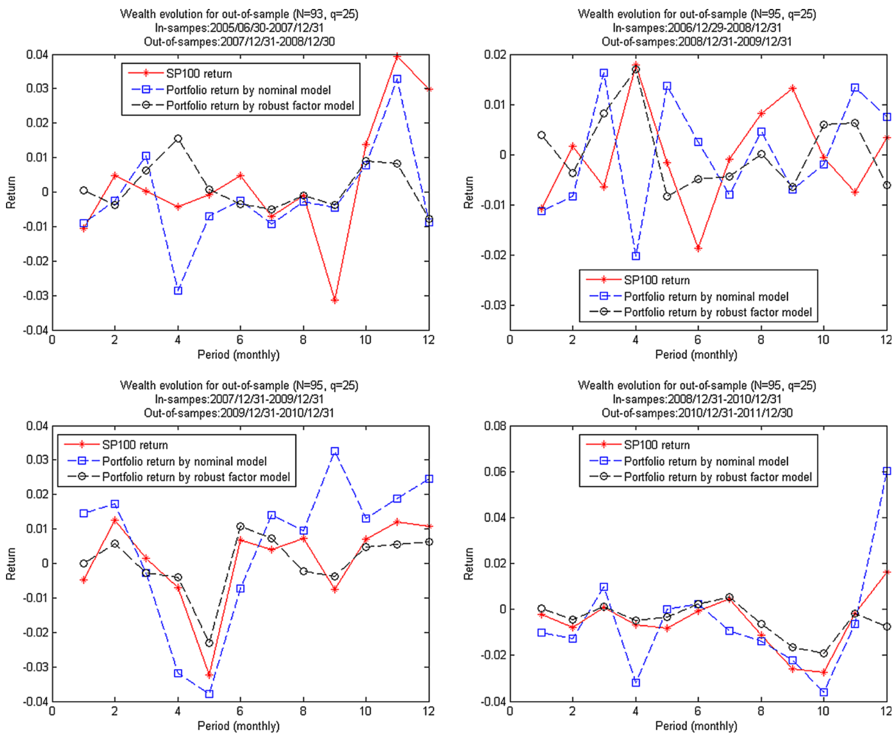


**Fig. 5** Wealth evolutions for rolling out-of-samples

portfolios generated by the robust factor model is reasonably close to the S&P100 index see (5) and are relatively stable without large drops. The portfolio returns generated by the nominal model may be sensitive to perturbations of the coefficient and exhibits wider divergence in returns. For example, for out-of-sample testing in year 2008 see first sub-figure when the market starts to decrease during time periods 2–4, the portfolio generated by the nominal model drops more rapidly than the index but the robust portfolio exhibits good performance and actually dominates the performance of the S&P Index during most of this period of market decline. During time periods 7–9, the portfolios by both models avoided the market plunge and the performance by robust factor model were generally better than that of nominal model. These examples shown that the robust factor model protected against the uncertainty of market movement successfully. During periods 8–11 a market recovery is seen and the returns from the robust portfolios actually lag the returns from the S&P 100 index and nominal portfolio, but then these latter two portfolios drop more steeply in the period from 11 to 12 of decline. This indicates that robustness protects well against large drops but may not accelerate as fast in periods of steady market increases.

Similar protection against downside risk can be seen in the other sub-figures in Fig. 5. For example, when market rapidly increased in years 2009, 2010 and 2011 which represent different parameter structures for the models, the portfolios generated by the factor robust model still displayed relative stable return performance compared to that of the nominal model. It is clear that in the third sub-figure the path of the robust model in periods 3–5 did not drop as fast than that of nominal model and target index.

Next we vary the portfolio size $q$ from 10 to 75 in increments of 5 and solve both nominal and robust models under different portfolio sizes. The mixed integer solver in Gurobi for MISOCP is mainly based on the branch-and-bound algorithm which tries to shrink the gap between the SOCP relaxed lower bound and its feasible upper bound. For the instances of tracking S&P100, we set the running time for Gurobi to 100 seconds and set the relative optimality gap tolerance to 10e-08. In our computations, the hardest instance required 50 seconds to satisfy the gap tolerance, which indicates all instances can be solved to optimality within 50 seconds and that the sizes of these instances were easily handled by Gurobi. The performance metrics include: daily portfolio return, daily portfolio variance, and daily portfolio Sharpe ratio. We compare these performance metrics by using in-sample and out-of sample data. There is no re-balancing of portfolios during a testing period. For example, 630 in-sample daily returns from June 30, 2005 to December 31, 2007 were used to generate data and then tracking portfolios were tested out-of-sample from December 31, 2007 to December 30, 2008 which is a period in which a large market decline was experience.

The size of uncertainty sets are controlled by the joint confidence level $\omega$ in equations (43) and (44) in Appendix 1. In our experience, for very high joint confidence levels, e.g. $\omega = 0.99$, we have high confidence that the solution of robust model protects against parameter uncertainty, but the feasible region of robust model may be restricted and more instances will be infeasible when portfolio size is too small, e.g. $q = 15$. On the other hand, for a low joint confidence level, e.g.

$\omega = 0.55$, more instances with smaller portfolio sizes can admit feasible solutions, but the confidence that the parameters lie in the designed ellipsoid is low. Therefore, we set a reasonable joint confidence level $\omega = 0.95$ in our computation. The parameters $(\mu, \Sigma)$ in nominal model (6)–(12) are approximated by the three factor model where $\mu = \mu_0$, $\Sigma = V_0^T F V_0 + D_0$, and then are used in the robust model as well. We also used the linear regression to approximate the out-of-sample data and then calculated the associated out-of-sample performance. Figures 6, 7, 8, 9 and 10 shows these comparison between two models.

It is clear to see the trend that the portfolio returns by the nominal model decreased as size increased from Fig. 6. All instances were solved to optimality by applying the Gurobi mixed integer solver. The portfolio returns decreased as portfolio size increased because of the increased diversification, that is, the more assets are allocated, the less risky are the portfolios, and thus the smaller portfolio returns. Meanwhile the portfolio return by robust model for both in-sample and out-of-sample seem unchanged too much with respect to portfolio size, however they are generally better than the returns generated by the nominal models for out-of-sample. Figure 6 shown that the robust model can protect against the downside risk in estimation of expected return vector $\mu_0$ due to market uncertainty. We can also see that portfolio return by the robust counterpart in the out-of-sample period (averagely 0.63 ‰) is better than the index return in the same out-of-sample period (−0.10 ‰).

From Fig. 7, we can easily see the diversification process of portfolios generated by nominal model as $q$ gets larger, i.e. as portfolio sizes get larger, the portfolio variance decreased. Portfolio variance for portfolios generated by the robust model for in-sample and out-of-sample are lower than that from the nominal model for corresponding in-sample and out-of-sample periods, which indicates the cardinality constraint had an impact on the conic constraints that represent the portfolio risk in

in questo vincolo subentra la varianza

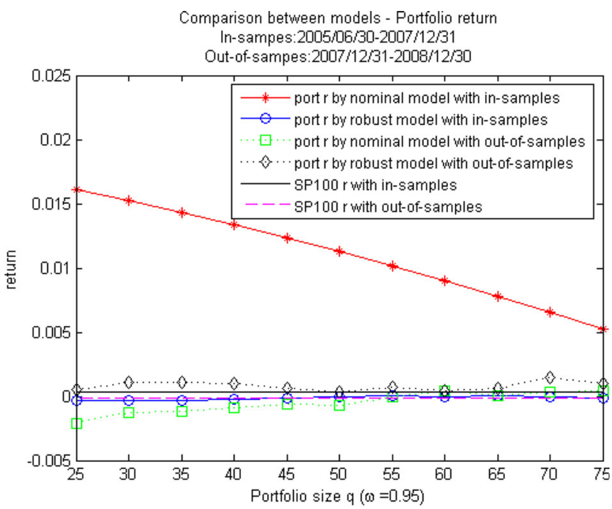


**Fig. 6** Model comparison—portfolio return

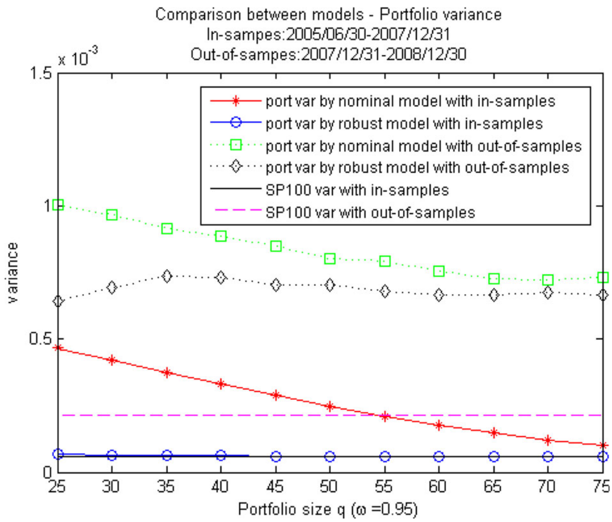


**Fig. 7** Model comparison—portfolio variance

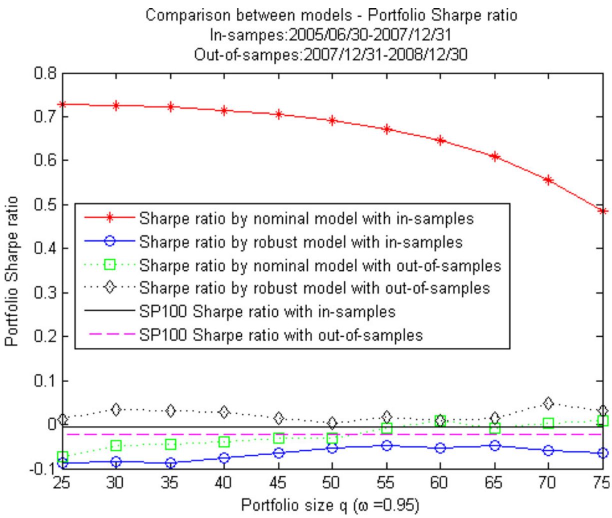


**Fig. 8** Model comparison—portfolio Sharpe ratio

that variance was reduced. The variance of the S&P100 index has the lowest value in the in-sample period, and the value in the out-of-sample period is still lower than robust models due to the diversification effect of of having more assets. The average portfolio variance by the robust model in the out-of-sample period is 0.54 ‰ on average, meanwhile the S&P100 variance equals 0.22 ‰ in the same out-of-sample period.

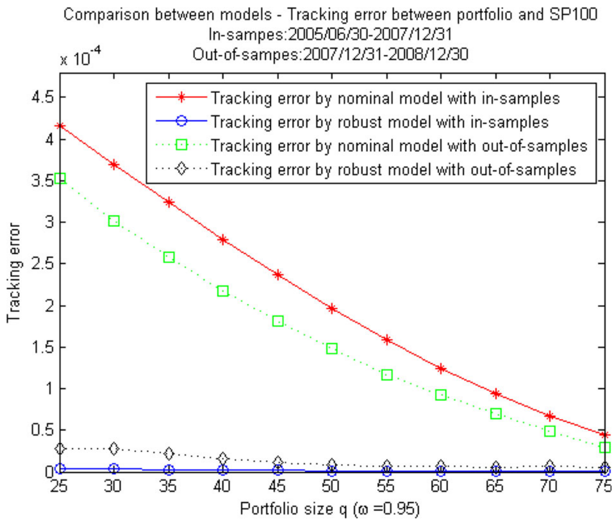


**Fig. 9** Model comparison—tracking error

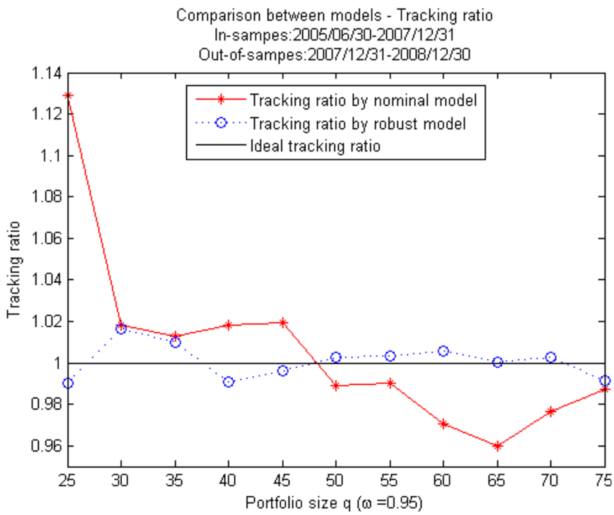


**Fig. 10** Model comparison—tracking ratio

The portfolio Sharpe ratio is defined as $\frac{E(r_{port})-E(r_f)}{\sqrt{\mathrm{var}(r_{port})}}$ where $r_f$ is the return of 10 year U.S Treasury bonds. From Fig. 8, the Sharpe ratio generated by nominal models decreased as the portfolio size increased, which means the portfolio return decreased more quickly than the reduction of portfolio variance across the size. The Sharpe ratio by nominal model for in-sample are better than that by robust factor model for in-sample, this is reasonable since robust counterpart consider the worst

**Table 1** Ticker symbol across sectors (SP100)

| Sector (total number) | Ticker symbol |
|---|---|
| 1 Consumer discretionary (Chen and Kwon 2012) | AMZN, CMCSA, DIS, FOXA, GM, HD, LOW, MCD, NKE, SBUX, TGT, TWX |
| 2 Consumer staples (Boyd and Vandenberghe 2004) | COST, CVS, FB, KO, MDLZ, PEP, WAG, WMT |
| 3 Energy (Canakgoz and Beasley 2009) | APA, APC, COP, CVX, DVN, HAL, NOV, OXY, SLB, XOM |
| 4 Financials (Chopra and Ziemba 1993) | AIG, ALL, AXP, BAC, BK, BRK/B, COF, GS, JPM, MET, PM, SPG, USB, WFC |
| 5 Health care (Chen et al. 2013) | ABBV, ABT, AMGN, BAX, BIIB, BMY, GILD, JNJ, LLY, MDT, MRK, PFE, UNH |
| 6 Industrials (Chen et al. 2013) | CAT, EMR, FDX, GD, GE, HON, LMT, MMM, NSC, RTN, UNP, UPS, UTX |
| 7 Information technology (Corielli and Marcellino 2006) | AAPL, ACN, BA, CSCO, EBAY, EMC, GOOG, HPQ, IBM, INTC, MA, MSFT, ORCL, QCOM, TXN |
| 8 Materials (Bertsimas et al. 2011) | CL, DD, DOW, FCX, MO, MON |
| 9 Telecommunications services (Ben-Tal and Nemirovski 1998) | V, VZ |
| 10 Utilities (Birge and Louveaux 2011) | C, EXC, F, MS, PG, SO, T |

**Table 2** Tracking ratio comparison

| $N = 93$ $q$ | Move index $M_I$ | Move port (nomi. $M_{P1}$) | $\frac{M_I}{M_{P1}}$ | $\left\lvert \frac{M_I}{M_{P1}} - 1 \right\rvert$ | Move port (rob. $M_{P2}$) | $\frac{M_I}{M_{P2}}$ | $\left\lvert \frac{M_I}{M_{P1}} - 1 \right\rvert$ |
|---|---|---|---|---|---|---|---|
| 25 | 0.6575 | 0.5819 | 1.1300 | 0.1300 | 0.6632 | 0.9914 | 0.0086 |
| 30 | 0.6575 | 0.6453 | 1.0190 | 0.0190 | 0.6464 | 1.0172 | 0.0172 |
| 35 | 0.6575 | 0.6486 | 1.0137 | 0.0137 | 0.6507 | 1.0105 | 0.0105 |
| 40 | 0.6575 | 0.6452 | 1.0191 | 0.0191 | 0.6630 | 0.9917 | 0.0083 |
| 45 | 0.6575 | 0.6444 | 1.0205 | 0.0205 | 0.6593 | 0.9973 | 0.0027 |
| 50 | 0.6575 | 0.6642 | 0.9899 | 0.0101 | 0.6551 | 1.0037 | 0.0037 |
| 55 | 0.6575 | 0.6634 | 0.9911 | 0.0089 | 0.6547 | 1.0043 | 0.0043 |
| 60 | 0.6575 | 0.6770 | 0.9713 | 0.0287 | 0.6534 | 1.0064 | 0.0064 |
| 65 | 0.6575 | 0.6843 | 0.9608 | 0.0392 | 0.6568 | 1.0011 | 0.0011 |
| 70 | 0.6575 | 0.6728 | 0.9773 | 0.0227 | 0.6552 | 1.0036 | 0.0036 |
| 75 | 0.6575 | 0.6653 | 0.9883 | 0.0117 | 0.6628 | 0.9920 | 0.0080 |
| Average | 0.6575 | 0.6539 | 1.0074 | 0.0294 | 0.6564 | 1.0018 | 0.0067 |

scenario for parameters. On the other hand, the Sharpe ratio generated by robust factor models for out-of-sample are better than those generated by nominal models out-of-sample, this is crucial since we want to reduce the negative effect of market uncertainty. Therefore, Fig. 8 indicates that the portfolios generated by robust models are more stable than those from the nominal models across different portfolio sizes $q$, this illustrates the benefit of cardinality constraint in the robust factor model. The average portfolio Sharpe ratio of portfolios generated by robust models is 0.0176 in the out-of-sample period and the Sharpe ratio of the S&P100 in the same out-of-sample period is −0.0218.

After solving both the nominal index tracking model and its factored robust counterparts, the tracking errors are calculated by $(x - x_{BM})^T \Sigma (x - x_{BM})$, which represent the variance difference between the portfolio and the target index. As can been seen in Fig. 9, the tracking errors by portfolios from the robust model are generally smaller than those from portfolios generated by the nominal model with respect to size for in-sample and for out-of-sample. This trend can be guaranteed since we generate the worst scenario bound for the parameters and the corresponding tracking error by robust model is also the lower bound for the tracking error by nominal model.

Another way to measure the tracking performance is by the tracking ratio. Similar to the definition of tracking ratio in Cornuejols and Tutuncu (2006), we calculate the tracking ratio through the following formula:

$$R_{0t} = \frac{M_I}{M_P} = \frac{\sum_{i=1}^{n} V_{it} / \sum_{i=1}^{n} V_{i0}}{\sum_{j=1}^{q} x_j V_{jt} / \sum_{j=1}^{q} x_j V_{j0}}$$

where $M_I = \frac{\sum_{i=1}^{n} V_{it}}{\sum_{i=1}^{n} V_{i0}}$ indicates the target index's movement after investment, $M_P = \frac{\sum_{j=1}^{q} x_j V_{jt}}{\sum_{j=1}^{q} x_j V_{j0}}$ denotes the movement of portfolio's market value during the out-of-sample period. The ideal tracking ratio, $R_{0t}$, is 1, a value over 1 means underperformance with respect to the target index, and a value less than 1 indicates excessive return. Figure 10 display the comparison of tracking ratios of portfolios generated from the nominal and robust models.

The straight line indicates that a portfolio perfectly tracks the market index, S&P100. There was no rebalance during the tracking period after investment. From Fig. 10, the tracking ratios by robust model are more closer to 1 than that from the nominal model with respect to size for out-of-sample testing, which indicate the factored robust tracking model has better tracking performance during period from December 31, 2007 to December 30, 2008, a main period in financial crisis.

After obtaining the portfolios by solving the models, we test the movement of the index and portfolios in the out-of-samples period in terms of market value. Table 2 shows more details about the market value movements of index and the portfolios with respect to size. The movement of the target index is constant to size while the movement of portfolios by different models are varying with respect to size. For example, under $q = 25$, $\frac{\sum_{i=1}^{n} V_{it}}{\sum_{i=1}^{n} V_{i0}} = 0.6575$ indicates that the market value of the index at time $t$ is 65.75 % of the market value of the index at time 0, or the index value decreased 34.25 % at the end of the out-of-sample period. Meanwhile, $\frac{\sum_{j=1}^{q} x_j^{no\min al} V_{jt}}{\sum_{j=1}^{q} x_j^{no\min al} V_{j0}} = 0.5819$ denotes the market value of the nominal portfolio dropped 41.81 % in the same out-of-sample period, and the associated tracking ratio

$R_{0t}^{no\min al} = \frac{0.6575}{0.5819} = 1.1300$ denotes the speed of the value shrinkage of the nominal portfolio is faster than that of the index. On the other hand, $\frac{\sum_{j=1}^{q} x_j^{robust} V_{jt}}{\sum_{j=1}^{q} x_j^{robust} V_{j0}} = 0.6632$ denotes the market value of the robust portfolio dropped 33.68 % at the end of the out-of-sample period, which indicates the downward descent in terms of market value is 8.13 % (41.81–33.68 %) less than the descent of the nominal portfolio at the same period, and the associated tracking ratio $R_{0t}^{robust} = \frac{0.6575}{0.6632} = 0.9914$ denotes the decreasing speed of the market value of the robust portfolio is also less than the downside speed of the market value of the index market. The columns with $\left|\frac{M_I}{M_P} - 1\right|$ values indicate how close is a constructed portfolio to the index, and the ideal value is 0. As shown in the Table 2, the portfolios generated by robust model are relative closer to the S&P100 compared with those by the nominal model.

## 5 Conclusions

A factor-based robust enhanced index tracking model was developed in this paper. A robust three factor model of risk of Fama and French was used as the basis of constructing robust counterparts of the nominal tracking model. Computational results using the S&P100 index as a benchmark have shown that the robust counterpart has better tracking performance and Sharpe ratios than portfolios generated by nominal models out-of-sample. Despite the presence of a cardinality constraint the robust counterpart which is a MISOCP is computationally tractable for tracking portfolios with 100 assets and it is easy to incorporate additional convex practical constraints, e.g. transaction costs. However, tracking larger benchmarks e.g. the S&P 500 or Russell 1000 may present serious computational challenges and developing methods for solving robust tracking problems of this size will be the subject of future research.

## Appendix 1: Parameter generation for the robust tracking model

We applied the same procedure described in Goldfarb and Iyengar (2003) to three-factor model for constructing factor-based robust index tracking models. We follow Goldfarb and Iyengar (2003) closely. Suppose the return vector $r$ is given by the linear regression model:

$$r = \mu + V^T f + \epsilon \tag{39}$$

where $\mu \in R^n$ is the vector of mean returns, $f \sim N(0, F) \in R^m$ is the vector of returns of the factors that drive the market, $V \in R^{m \times n}$ is the matrix of factor loadings of the $n$ assets, and $\epsilon \sim N(0, D)$ is the vector of residual returns.

Let $S = [r^1, r^2, \ldots, r^p] \in R^{n \times p}$ be the matrix of asset returns and $B = [f^1, f^2, \ldots, f^p] \in R^{m \times p}$ be the matrix of factor returns, then (39) which can be represented by the following linear model:



m è il numero di fattori di rischio
n è il numero di asset
p è il numero di periodi di tempo(nel caso nostro è il numero di giorni)

$$y_i = Ax_i + \epsilon_i, \forall i = 1, \ldots, n$$

where $y_i = \left[r_i^1, r_i^2, \ldots, r_i^p\right]^T, A = [1, B^T], x_i = [\mu_i, V_{1i}, V_{2i}, \ldots, V_{mi}]^T$ and $\epsilon_i = \left[e_i^1, e_i^2, \ldots, e_i^p\right]^T$.

As we shown in Sect. 4.1, for single factor model, we set $B = [f^1, f^2] = \left[r_M, r_f\right]^T$; for three factor model, $B = [f^1, f^2, f^3, f^4] = \left[r_M, r_f, SMB, HML\right]^T$. The least-squares estimate $\overline{x}_i$ of the true parameter $x_i$ is given by

$$\overline{x}_i = \left(A^T A\right)^{-1} A^T y_i, \forall i = 1, \ldots, n \tag{40}$$

Substituting $y_i = Ax_i + \epsilon_i$ into (40), we get $\overline{x}_i - x_i = \left(A^T A\right)^{-1} A^T \epsilon_i \sim N(0, \Sigma)$ where $\Sigma = \sigma_i^2 \left(A^T A\right)^{-1}$. $\sigma_i^2$ is unknown in practice, so we replace $\sigma_i^2$ by $(m+1)s_i^2$ where $s_i^2$ is the unbiased estimate of $\sigma_i^2$. $\sigma_i^2$ is given by

$$s_i^2 = \frac{\|y_i - A\overline{x}_i\|}{p - m - 1} \tag{41}$$

norma dovrebbe essere al quadrato

and the resulting variable

$$\mathbb{Y} = \frac{1}{(m+1)s_i^2} \left(\overline{x}_i - x_i\right)^T \left(A^T A\right) \left(\overline{x}_i - x_i\right) \tag{42}$$

is a F-distribution with $(m+1)$ degrees of freedom in the numerator and $(p - m - 1)$ degrees of freedom in the denominator Goldfarb and Iyengar (2003).

By setting the joint confidence region $\omega$ for set $(\mu, V)$, Goldfarb and Iyengar (2003) derive the following result for the parameters that can be used in our robust model:

$$\mu_{0,i} = \overline{\mu}_i, \gamma_i = \sqrt{\left(A^T A\right)_{11}^{-1} c_1(\omega) s_i^2}, i = 1, \ldots, n \tag{43}$$

$$V_0 = \overline{V}, G = \left(Q\left(A^T A\right)^{-1} Q^T\right)^{-1}, \rho_i = \sqrt{m c_m(\omega) s_i^2}, i = 1, \ldots, n \tag{44}$$

where $c_J(\omega)$ be the $\omega$-critical value. More prove details read in Goldfarb and Iyengar (2003). Then a worst case bound for the covariance matrix is achieved by 3 factor model, i.e. $cov = V_0^T F V_0 + \overline{D}$, where $\overline{D} = diag\left(s_i^2\right)$. The uncertainty set for $\mu$ in (43) will be used in for robust portfolio returns and $V_0$ in (44) will be used to relative robust covariance.

## References


Ben-Tal A, Nemirovski A (1998) Robust convex optimization. Math Op Res 23(4):769–805

Ben-Tal A, Nemirovski A (2000) Robust solutions of linear programming problems contaminated with uncertain data. Math Program 88(3):411–424

Ben-Tal A, Margalit T, Nemirovski A (2000) Robust modeling of multi-stage portfolio problems. In: Frenk H, Roos K, Terlaky T, Zhang S (eds) High performance optimization, applied optimization, vol 33. Springer, US, pp 303–328

Bertsimas D, Pachamanova D (2008) Robust multiperiod portfolio management in the presence of transaction costs. Comput Oper Res 35(1):3–17

Bertsimas D, Brown DB, Caramanis C (2011) Theory and applications of robust optimization. SIAM Rev 53(3):464–501. doi:10.1137/080734510

Birge JR, Louveaux F (2011) Introduction to stochastic programming. Springer

Boyd S, Vandenberghe L (2004) Convex optimization. Cambridge University Press, New York

Burmeister E, Roll R, Ross SA (2003) Using macroeconomic factors to control portfolio risk. Tech Rep

Canakgoz N, Beasley J (2009) Mixed-integer programming approaches for index tracking and enhanced indexation. Eur J Op Res 196(1):384–399

Chang T, Meade N, Beasley J, Sharaiha Y (2000) Heuristics for cardinality constrained portfolio optimisation. Comput Op Res 27:1271–1302

Chen C, Kwon RH (2012) Robust portfolio selection for index tracking. Comput Op Res 39(4):829–837

Chen C, Li X, Tolman C, Wang S, Ye Y (2013) Sparse portfolio selection via quasi-norm regularization, working paper

Chopra VK, Ziemba WT (1993) The effect of errors in means, variances, and covariances on optimal portfolio choice. J Portf Manag 19(2):6–11

Corielli F, Marcellino M (2006) Factor based index tracking. J Bank Financ 30(8):2215–2233

Cornuejols G, Tutuncu R (2006) Optimization methods in finance. Finance and risk, Cambridge University Press, Mathematics

D'ecclesia RL, Zenios SA (1994) Risk factor analysis and portfolio immunization in the italian bond market. J Fixed Income 4(2):51–58

Erdogan E, Goldfarb D, Iyengar G (2004) Robust portfolio management. Tech Report CORC TR-2004-11, IEOR, Columbia University, New York

Fama E, French K (1993) Common risk factors in the returns on stocks and bonds. J Financ Econ 33(1):3–56

Gulpinar N, Katata K, Pachamanova DA (2011) Robust portfolio allocation under discrete choice constraints. J Asset Manag 12:67–83

Goldfarb D, Iyengar G (2003) Robust portfolio selection problems. Math Op Res 28(1):1–38. doi:10.1287/moor.28.1.1.14260

Gurobi Optimization I (2015) Gurobi optimizer reference manual

Jorion P (2003) Portfolio optimization with tracking error constraints. Financ Anal J 59(5):70–82

Karlow D, Rossbach P (2011) A method for robust index tracking. In: Hu B, Morasch K, Pickl S, Siegle M (eds) Operations research proceedings 2010. Operations research proceedings. Springer, Berlin Heidelberg, pp 9–14

Kolbert F, Wormald L (2010) Robust portfolio optimization using second-order cone programming

Lejeune MA, Samatlı-Paç G (2013) Construction of risk-averse enhanced index funds. INFORMS J Comput 25(4):701–719

Mulvey JM, Vanderbei RJ, Zenios SA (1995) Robust optimization of large-scale systems. Op Res 43(2):264–281. doi:10.1287/opre.43.2.264

Sadjadi SJ, Gharakhani M, Safari E (2012) Robust optimization framework for cardinality constrained portfolio problem. Appl Soft Comput 12(1):91–99

SPs (2015) http://www.us.spindices.com

Tutuncu R, Koenig M (2004) Robust asset allocation. Annals Op Res 132(1–4):157–187

Zenios SA (2006) Practical financial optimization: decision making for financial engineers. Blackwell, Incorporated