Accepted Manuscript

Index tracking model, downside risk and non-parametric kernel estimation

Jinbo Huang, Yong Li, Haixiang Yao

PII: S0165-1889(18)30128-3
DOI: 10.1016/j.jedc.2018.04.008
Reference: DYNCON 3592

To appear in: *Journal of Economic Dynamics & Control*

Received date: 15 July 2017
Revised date: 12 April 2018
Accepted date: 28 April 2018

Please cite this article as: Jinbo Huang, Yong Li, Haixiang Yao, Index tracking model, downside risk and non-parametric kernel estimation, *Journal of Economic Dynamics & Control* (2018), doi: 10.1016/j.jedc.2018.04.008

Index tracking model, downside risk and non-parametric kernel estimation ☆

Jinbo Huang[a], Yong Li[b,c], Haixiang Yao[d,*]

[a]School of Finance, Guangdong University of Finance and Economics, Guangzhou 510320, China
[b]School of Finance, University of International Business and Economics, Beijing 100029, China
[c]UQ Business School, the University of Queensland, St Lucia, QLD 4072, Australia
[d]School of Finance, Guangdong University of Foreign Studies, Guangzhou 510006, China

☆This research is partially supported by the National Natural Science Foundation of China (grant numbers 71471045, 71603058); the Innovative Research Group Project of National Natural Science Foundation of China (grant number 71721001); the Social Science Research Foundation of the National Ministry of Education of China (grant numbers 15YJAZH051, 16YJC790033); the Natural Science Foundation of Guangdong Province of China (grant numbers 2016A030313656, 2017A030313399, 2017A030313397); the Innovative School Project in Higher Education of Guangdong Province of China (grant number GWTP-GC-2017-03); the Innovation Team Project of Guangdong Colleges and Universities (grant number 2016WCXTD012); the Philosophy and Social Science Programming Foundation of Guangdong Province of China (grant number GD15XYJ03); and the AFAANZ Grant 2016: "A new non-parametric model for measuring portfolio downside risk". We thank Yue Dong at the University of Queensland for research assistance.

*Corresponding author. Tel.:+86 2037105360.

Email addresses: j.b.huang@gdufe.edu.cn (Jinbo Huang), y.li@business.uq.edu.au (Yong Li), yaohaixiang@gdufs.edu.cn (Haixiang Yao)

Index tracking model, downside risk and non-parametric kernel estimation

Abstract: In this paper, we propose an index tracking model with the conditional value-at-risk (CVaR) constraint based on a non-parametric kernel (NPK) estimation framework. In theory, we demonstrate that the index tracking model with the CVaR constraint is a convex optimization problem. We then derive NPK estimators for tracking errors and CVaR, and thereby construct the NPK index tracking model. Monte Carlo simulations show that the NPK method outperforms the linear programming (LP) method in terms of estimation accuracy. In addition, the NPK method can enhance computational efficiency when the sample size is large. Empirical tests show that the NPK method can effectively control downside risk and obtain higher excess returns, in both bearish and bullish market environments.

Keywords: Non-parametric kernel estimation; Index tracking model; Conditional value-at-risk.
JEL classification: G11; G10

1. Introduction

Over the past 10 years, financial markets have witnessed rapid developments in indexation funds. Indexation funds deliver returns for a benchmark index with low turnover, diversified portfolios, and low expenses. Appel et al. (2016) report that the proportion of total market capitalization of indexation funds quadrupled from 2% to more than 8%. Among indexation products, enhanced index tracking funds that operate according to trade-offs between tracking errors and excess returns have developed more quickly than index replication funds (Filippi et al., 2016). These index-linked trades have become especially prevalent in the asset management industry, as investors tend to require benchmarking as a mechanism to evaluate portfolio performance. A good example is that sharp increases in defined contribution pension plans require fund managers to beat the benchmarks but effectively control tracking errors. This requirement significantly changes fund managers' decisions (Christoffersen and Simutin, 2017). For index-linked fund managers, the key challenge is to efficiently track the benchmark index while also capturing higher potential excess returns. In this paper, we use a non-parametric kernel (NPK) estimation method to study index tracking models with a conditional value-at-risk (CVaR) constraint. We show that our proposed NPK model can effectively address computational difficulty when the sample size is large. Both numerical and empirical tests demonstrate that the NPK method outperforms the classic linear programming (LP) method in downside risk control and in obtaining excess returns, in both bullish and bearish market environments.

Our study contributes to the growing number of index investing strategies. In fact, effective index tracking strategies benefit not only passive investors, such as index funds and pension funds, but also active portfolio managers (Beasley et al., 2003; Alexander and Baptista, 2010; Christoffersen and Simutin, 2017). To ensure that our works can be beneficial to the greatest number of investors, we study three classes of index tracking models. The first is the index replication model (IRM), the aim of which is to strictly control tracking errors without considering excess returns (Beasley et al., 2003; Haugen and Baker, 1990; Larsen Jr and Resnick, 1998; Hodges, 1976; Roll, 1992; Franks, 1992; Rohweder, 1998; Wang, 1999). The second class uses index returns as a benchmark for measuring a portfolio's excess returns. We refer to this model as the active investment model (AIM), which seeks to maximize portfolio returns relative to the benchmark index. The third is the enhanced indexation model (EIM). Canakgoz and Beasley (2009) review the differences between EIM and IRM. EIM chases excess returns when minimizing tracking errors (Roman et al., 2013; Filippi et al., 2016). There is one strand of literature that treats EIM as a multi-objective decision model. For example, Filippi et al. (2016) and Wu et al. (2007) convert EIM to a bi-objective model, Anagnostopoulos and Mamanis (2010) and Hirschberger et al. (2013) convert it to a tri-objective model. However, the issue here is that the optimal solution set to a multi-objective optimization typically has very high or even infinite dimensional cardinality. This makes obtaining the complete solution set computationally problematic (Filippi et al., 2016). Although in this study we have three targets - excess return, tracking error, and downside risk - we use a balance parameter to connect excess return and tracking error and use downside risk as a constraint condition. Hence, our model is a mono-objective optimization.[1] By adjusting the balanced parameter, EIM can be degenerated to AIM and IRM.

The first task of the index tracking models is to determine a way of measuring tracking error. Roll (1992) uses the variance of differences between tracking portfolio returns and benchmark index returns to measure tracking error. He adopts the

[1]For other mono-objective optimization models, see for example, Roman et al. (2013), who adopt second-order stochastic dominant theory to construct EIM and use a cutting plane approach to solve this model; Valle et al. (2014), who propose a three-stage solution approach to select absolute return portfolios and extend this approach to EIM; and Guastaroba et al. (2016), who build EIM based on the Omega ratio and convert their models to a mixed-integer linear programming problem.

mean-variance framework (Markowitz, 1952) to show that the mean-TEV (tracking error volatility) efficient portfolio is mean-variance inefficient. In addition, Coleman et al. (2006) and Alexander and Baptista (2008, 2010) use the same variance definition. However, Beasley et al. (2003) state that variance is irrational and challenges this definition by demonstrating that when the difference in returns between a tracking portfolio and a benchmark index is constant, the variance is zero. Therefore, studies tend to adopt linear or absolute deviations to measure tracking errors (Sharpe, 1971; Clarke et al., 1994; Rudolf et al., 1999). In this paper, we use a higher-order origin moment of absolute difference between tracking portfolio returns and index returns as our measure of tracking error, which has three advantages. First, our definition is a convexity function of decision variables, which is helpful for optimization. Second, our definition is more general and when we select a different order, our measure can be directly related to the works of Sharpe (1971), Rudolf et al. (1999), Beasley et al. (2003) and Clarke et al. (1994), all of which define tracking error by an absolute difference that is a special case of our definition. Third, whereas these studies use sample data to define tracking error, we treat tracking error as a random variable and adopt its higher-order moment to define it. By using this definition, we obtain an expression of tracking error when the distribution is known; otherwise, we use the sample data to estimate tracking error.

Next, our index tracking model focuses on controlling market jump risk (Wang et al., 2012),and imposes transaction costs and investment proportion as important constraints (Beasley et al., 2003; Canakgoz and Beasley, 2009). We intend to prevent the tracking portfolio from jumping in conjunction with severe market recession. To control downside risk, Alexander and Baptista (2008) and Palomba and Riccetti (2012) impose a value-at-risk (VaR) measure on Roll's (1992) model. These studies assume normal distributions, an assumption that is not consistent with the reality of financial markets. In addition, the VaR measure does not satisfy the sub-additive condition; therefore, VaR is not a coherent risk measure (Artzner et al., 1999). With superior mathematical properties, CVaR was developed to overcome some of the difficulties of VaR (Pflug, 2000). CVaR is a coherent risk measure (Acerbi and Tasche, 2002), satisfying the axioms of translation invariance, subadditivity, positive homogeneity, and monotonicity, all of which address the non-coherent issues of VaR that Acerbi and Tasche (2002) argue might not stimulate investment diversification. In addition, CVaR indicates the tail conditional expectation and is more sensitive to changes in the tail distribution than VaR. As a comparison, VaR is only a quantile measure; it ignores all losses larger than its value. CVaR was recently adopted by the Basel III to serve as a key measure of risk

(Basel Committee on Banking Supervision, 2013) . Wang et al. (2012) use the CVaR as a constraint to control the downside risk of the optimal tracking portfolio. They use a mixed 0-1 linear programming method to show that the CVaR constraint can effectively protect a tracking portfolio from market jump risk. However, the shortcoming of this model is that the number of constraints and decision variables increases dramatically when the asset and sample size increase. This limits applications of this model to large-scale asset allocations.

In this paper, we control the tracking portfolio's downside risk by adding the CVaR constraint to the index tracking model. In theory, we demonstrate that the index tracking model with the CVaR constraint is a convex optimization. As opposed to Wang et al. (2012), our contribution is to propose an NPK framework to solve the index tracking model under general distributions. Specifically, we obtain NPK estimators of CVaR and tracking error, respectively. Next, we embed these estimators into our index tracking models with the CVaR constraint. More importantly, we emphasize that the number of decision variables and constraints do not rely on sample size. This significantly mitigates the computational difficulties that arise with large sample sizes. Our NPK model can therefore be applied to solve various problems in asset management, especially when dealing with large samples and unknown distributions.

We carry out Monte Carlo simulations to examine the performance of the NPK method when sample size and portfolio size increase. Simulation results show that the NPK method outperforms the LP method in terms of estimation accuracy and that this disparity is statistically significant. With respect to computational time, our findings are mixed. The sample size tests show that the computing time of the LP method increases significantly when sample size increases. However, the computing time of the NPK method remains quite stable. The LP method requires considerably more time than NPK when the sample size is large. In the portfolio size tests, for IRM and EIM, we show that the NPK method performs better with a moderate portfolio size. However, for AIM, the LP method requires less time than NPK. In simulated market environments, out-of-sample simulations show that the NPK method outperforms LP in controlling downside risk and obtaining excess return.

Finally, we highlight that our model is suited for optimized index tracking rather than full replication. In contrast to full replication, optimized index tracking adopts only part of the constituents in constructing a portfolio to track the benchmark index while minimizing tracking errors (Yao et al., 2006). This means that we need to select a subset of constituents to construct an optimized tracking portfolio. A popular method is to use cardinality constraints to pick up stocks from constituents (Canakgoz and Beasley, 2009). However, cardinality constraints require the introduction of binary variables to implement stock selections in addition to optimization; this significantly increases the computational complexity of the models. To address this problem, researchers have developed a series of computational methods, such as the heuristic frameworks (Beasley et al., 2003; Guastaroba and Speranza, 2012; Filippi et al., 2016), mixed-integer linear programming (Canakgoz and Beasley, 2009; Guastaroba et al., 2016), a hybrid genetic approach (Wang et al., 2012), and a cutting plane approach (Roman et al., 2013). Canakgoz and Beasley (2009) review these computational methods in detail and conclude that an individual method can only address certain problems in terms of index tracking. However, we argue that, as pointed out by Roman et al. (2013), prior literature concentrates more on the methods of solving models while ignoring the essence of index tracking.

In this study, we implement *ex ante* unbiased Beta criteria (Ling et al., 2014) to select a subset of constituents to optimize the tracking portfolio. This method avoids making use of numerous auxiliary variables and significantly reduces the complexity of solving the model. Empirically, for the American and British stock markets, we use the S&P 500 and FTSE 100 constituents to track benchmark indices in both bearish and bullish environments to test our model. We document that our proposed NPK method performs better in controlling downside risk than the LP method, in both market environments. The advantage of NPK in controlling downside risk results in deviations from the benchmark index but higher excess returns.

This paper is organized as follows. Section 2 introduces the NPK estimator for CVaR. In Section 3, we build the index tracking model with the CVaR constraint and derive the NPK estimation framework. Sections 4 and 5 carry out simulations and empirical tests, respectively, and Section 6 concludes the paper.

2. Non-parametric conditional value-at-risk

2.1. Definition of CVaR

Let $\Pr(\cdot)$ be the probability measure. Suppose that the cumulative distribution function of an asset's or a portfolio's return X is $P(x)$, i.e., $P(x) = \Pr(X \leqslant x)$. For a confidence level $1 - \alpha$, the value-at-risk $VaR(X, \alpha)$ of this asset or portfolio is defined by (Jorion, 1997):

$$VaR(X, \alpha) := -\inf\{x \in R : P(x) \geqslant \alpha\}. \tag{1}$$

When the distribution function $P(x)$ is continuous, $VaR(X,\alpha)$ is a negative value of the lower α quantile. Based on VaR, the CVaR of this asset or portfolio is (Rockafellar and Uryasev, 2000, 2002):

$$CVaR(X,\alpha) := -\mathrm{E}[X|X \leqslant -VaR(X,\alpha)]. \tag{2}$$

In the following, we specifically give analytical exressions of CVaR under normal, t and asymmetric Laplace (AL) distributions, respectively

If $X \sim N(\mu,\sigma^2)$, the analytical expression of CVaR is (Alexander and Baptista, 2004):

$$CVaR_N(X,\alpha) = k_1(\alpha)\sigma - \mu, \tag{3}$$

where $k_1(\alpha) = \varphi(z_\alpha)/\alpha$, z_α is α quantile of standard normal distribution and $\varphi(\cdot)$ is density function of standard normal distribution.

If X follows $t(\mu,\sigma^2,m)$, μ is a location parameter and σ is a scale parameter, m is degree of freedom, then (Andreev and Kanto, 2004):

$$CVaR_t(X,\alpha) = k_2(\alpha)\sigma - \mu, \tag{4}$$

where $k_2(\alpha) = \frac{1}{\alpha}\frac{m}{m-1}\frac{\Gamma\left(\frac{m+1}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\sqrt{m\pi}}\left(1+\frac{t_{1-\alpha}^2}{m}\right)^{\frac{1-m}{2}}$, $t_{1-\alpha}$ is $1-\alpha$ quantile of t-distribution with degree of freedom m , $\Gamma(\cdot)$ is a Gamma function.

In general, if X follows an elliptical distribution with a location parameter μ , a scale parameter σ , it follows that (Landsman and Valdez, 2003)

$$CVaR_E(X,\alpha) = k_\aleph(\alpha)\sigma - \mu, \tag{5}$$

where $k_\aleph(\alpha)$ is a function of α relying on the specific distribution form $\aleph$ (e.g., normal, t, logistic and exponential power distributions) of the elliptical distribution. Therefore, under elliptical distributions, CVaR is a linear function of μ and σ.

In fact, however, it is difficult to select a $k_\aleph(\alpha)$ from numerous ellipsoidal distributions to adapt complex financial markets. Additionally, ellipsoidal distributions can only reflect heavy tails but cannot depict asymmetric characteristics of financial data. Hence, the AL distribution $AL(\mu,\sigma^2)$ is used to reflect both heavy tails and asymmetric characteristics, where μ is an asymmetry parameter, and σ is a scale parameter. The following Lemma 1 shows an analytical expression of CVaR under the AL distribution.

Lemma 1 *(Zhao et al., 2015): If random variable $X \sim AL(\mu,\sigma^2)$, then the CVaR* *of X is*

$$CVaR_{AL}(X,\alpha) = -\frac{\sigma^2}{\mu + \sqrt{\mu^2+2\sigma^2}} \ln\alpha \left(2 + \frac{\mu^2 + \mu\sqrt{\mu^2+2\sigma^2}}{\sigma^2}\right) + \frac{\sigma^2}{\mu+\sqrt{\mu^2+2\sigma^2}}. \quad (6)$$

We see that $CVaR(X,\alpha)$ is not a linear function of μ and σ under the AL distribution.

2.2. NPK estimator of CVaR

As shown in Section 2.1, we can obtain analytical expressions of CVaR under some specific distribution settings. However, in realistic financial markets we have little knowledge about assets' distributions, and *ex ante* assumptions of distributions may cause model specification errors. Non-parametric methods driven by historical data can give estimations of a distribution function without any assumption. In this Section, we derive a NPK estimator of CVaR.

Suppose $\{x_t\}_{t=1}^T$ is sample data of asset return. Sample mean and standard deviation are $\bar{x} = \frac{1}{T}\sum\limits_{t=1}^{T} x_t$ and $\hat{\sigma}(x) = \left(\frac{1}{T-1}\sum\limits_{t=1}^{T}(x_t-\bar{x})^2\right)^{1/2}$. In order to estimate the CVaR, we first give an equivalent definition of CVaR (Rockafellar and Uryasev, 2000, 2002).

$$CVaR(X,\alpha) = \min_v F_\alpha(X,v),$$

where $F_\alpha(X,v) = v + \frac{1}{\alpha}\mathrm{E}[(-X-v)^+]$, and $(x)^+ = \max(x,0)$.

Next, we adopt the NPK method to estimate $F_\alpha(X,v)$. Using sample data, we can obtain a kernel estimator of density function $p(x)$ of X as follows (Li and Racine, 2007).

$$\hat{p}(x) = \frac{1}{Tb}\sum_{t=1}^{T} g\left(\frac{x-x_t}{b}\right), \quad (7)$$

where $g(\cdot)$ is a smooth kernel function, $b = b(T)$ is a smoothing parameter, called the bandwidth, which depends on the sample size T. Li and Racine (2007) show that the kernel estimator $\hat{p}(x)$ is a consistent estimator of $p(x)$ when kernel function $g(\cdot)$ and bandwidth b satisfy

(i) $g(\cdot)$ is nonnegative and bounded, $\int_{-\infty}^{+\infty} g(u)du = 1$, $g(-u) = g(u)$, $\int_{-\infty}^{+\infty} u^2 g(u)du > 0$;

(ii) $b(T) \to 0$ and $Tb(T) \to \infty$ as $T \to \infty$.

Guided by Yao et al. (2013) and Yao et al. (2015), we choose the Gaussian kernel function[1], $g(u) = (\sqrt{2\pi})^{-1}\exp(-u^2/2)$, and adopt the rule of thumb[2] to choose bandwidth b , i.e., $b = 1.06 \times T^{-1/5} \times \hat{\sigma}(x) = b_0\hat{\sigma}(x)$. Then, utilizing Eq (13) in Yao et al. (2013), we can obtain the kernel estimator of $F_\alpha(X, v)$ as

$$\hat{F}_\alpha(X, v) = v - \frac{1}{T\alpha}\sum_{t=1}^{T}\left((x_t + v)G\left(\frac{-v - x_t}{b}\right) + bH\left(\frac{-v - x_t}{b}\right)\right),$$

where $G(z) = \int_{-\infty}^{z} g(y)dy$, and $H(z) = \int_{-\infty}^{z} yg(y)\,dy$ .Then we have the NPK estimator of CVaR

$$CVaR_{npk}(X, \alpha) = \min_{v}\left(v - \frac{1}{T\alpha}\sum_{t=1}^{T}\left((x_t + v)G\left(\frac{-v - x_t}{b_0\hat{\sigma}(x)}\right) + b_0\hat{\sigma}(x)H\left(\frac{-v - x_t}{b_0\hat{\sigma}(x)}\right)\right)\right). \quad (8)$$

3. Index tracking model with CVaR

In this section, we propose an index tracking model with the CVaR constraint. Moreover, we prove that this model is a convex optimization model. In addition, we derive data-driven NPK and LP index tracking models with CVaR.

3.1. Index tracking model

Suppose that we use n ($n < N$) stocks, which are a subset of the N constituent stocks, to track the index return r_I . Denoted by $\mathbf{r} = (r_1, r_2, ..., r_n)'$, the returns of n stocks, and by $\mathbf{a} = (a_1, a_2, ..., a_n)'$, the weights of these n stocks in the tracking portfolio.

An index tracking problem aims to seek optimal investment strategy to replicate index returns and obtain potential excess returns. Consistent with Beasley et al. (2003) and Filippi et al. (2016), we define the objective function OF of an index tracking model as a trade-off between tracking error TE and excess return ΔR as follows.

$$OF = \lambda TE - (1-\lambda)\Delta R = \lambda\left(\mathrm{E}\left[|\mathbf{a}'\mathbf{r} - r_I|^\gamma\right]\right)^{1/\gamma} - (1-\lambda)\mathrm{E}\left[\mathbf{a}'\mathbf{r} - r_I\right],$$

[1]In this paper, we focus only on the Gaussian kernel function for two reasons. First, the Gaussian kernel function has well-known analytical properties, by which we can derive analytical expression for our proposed objective function and CVaR. Second, Li and Racine (2007) show that non-parametric estimation is insensitive to the choice of kernel function. The Gaussian function can provide a robust estimator for the density and distribution functions of a univariable.

[2]We acknowledge that for example, least square cross validation (LSCV) is another popular method selecting the bandwidth. LSCV-based LPM model is a dual-optimization model and it is beyond the scope of this paper.

where $\gamma > 0$ is the order. In particular, when $\gamma = 1$, the tracking error is the mean absolute deviation; when $\gamma = 2$, the tracking error is the root mean square error. $\lambda \in [0,1]$ reflects an investor's attitude. If this investor is conservative and concentrates on tracking errors, then $\lambda = 1$. By contrast, if this investor is aggressive and intends to obtain excess returns, then $\lambda = 0$. Generally, when $0 < \lambda < 1$, the aim is to obtain excess returns as well as track index trend.

Suppose the initial positions are $\mathbf{a}_0 = (a_{1,0}, a_{2,0}, ..., a_{n,0})'$. Then we need to adjust positions $\mathbf{\Delta a} = (\Delta a_1, \Delta a_2, ..., \Delta a_n)'$ to obtain optimal portfolios, where $\Delta a_i = a_i - a_{i,0}$. Our model takes transaction costs into account because they significantly affect a portfolio's performance (Brown and Smith, 2011). In this paper, we consider a proportional transaction costs function $TC_i = \delta_i^b \Delta a_i^+ + \delta_i^s \Delta a_i^-$, where $\delta_i^b \geqslant 0$ and $\delta_i^s \geqslant 0$ are the proportional costs for buying and selling asset i , respectively, and $\Delta a_i^+ = \max(a_i - a_{i,0}, 0)$, $\Delta a_i^- = \max(a_{i,0} - a_i, 0)$. Hence, the total cost is denoted by

$$TC = \sum_{i=1}^{n} TC_i = \sum_{i=1}^{n} \left(\delta_i^b \Delta a_i^+ + \delta_i^s \Delta a_i^-\right).$$

When $\delta_i^b = \delta_i^s = 0$, our models degenerate to the special case without transaction costs. We also assume an investor would impose constraints on investment positions and require minimum positions l_i and maximum positions u_i on asset i, $l_i \leqslant a_i \leqslant u_i,\ i = 1, 2..., n$. In particular, $l_i = -\infty, u_i = \infty$ allows for short sale, and $l_i = 0, u_i = 1$ indicates that short sale is prohibited. In this paper, we assume the initial wealth is standardized as 1, and an investor would like to control the transaction cost of asset i under c_i and total cost under c . In summary, we state the index tracking model as follows:

$$(\mathrm{P}_0)\quad \begin{cases} \min\limits_{\mathbf{a}\in\Re^n}\ OF = \lambda \left(\mathrm{E}\left[|\mathbf{a}'\mathbf{r} - r_I|^{\gamma}\right]\right)^{1/\gamma} - (1-\lambda)\mathrm{E}\left[\mathbf{a}'\mathbf{r} - r_I\right], \\ s.t.\ \ \delta_i^b \Delta a_i^+ + \delta_i^s \Delta a_i^- \leqslant c_i, \quad i = 1, 2..., n, \\ \qquad \sum\limits_{i=1}^{n} (\delta_i^b \Delta a_i^+ + \delta_i^s \Delta a_i^-) \leqslant c, \\ \qquad \Delta a_i^+ = \max(a_i - a_{i,0}, 0), \quad i = 1, 2..., n, \\ \qquad \Delta a_i^- = \max(a_{i,0} - a_i, 0), \quad i = 1, 2..., n, \\ \qquad \sum\limits_{i=1}^{n} a_i + \sum\limits_{i=1}^{n} (\delta_i^b \Delta a_i^+ + \delta_i^s \Delta a_i^-) = 1, \\ \qquad l_i \leqslant a_i \leqslant u_i, \quad i = 1, 2..., n. \end{cases}$$

In the model above, when $\lambda = 1$, problem (P_0) degenerates to IRM; when $\lambda = 0$, it degenerates to AIM; when $0 < \lambda < 1$, it represents EIM.

In order to tackle problem (P$_0$) easily, we take Δa_i^+ and Δa_i^- $(i = 1, 2..., n\ )$ as the additional decision variables subject to the linear conditions

$$\Delta a_i^+ \geqslant a_i - a_{i,0}, \Delta a_i^- \geqslant a_{i,0} - a_i,\ \Delta a_i^+ \geqslant 0, \Delta a_i^- \geqslant 0, \quad i = 1, 2..., n.$$

We define the set Ω below.

$$\Omega = \left\{ \mathbf{a}, \mathbf{\Delta a}^+, \mathbf{\Delta a}^- \in \Re^n \middle| \begin{array}{l} a_i - \Delta a_i^+ \leqslant a_{i,0}, \quad i = 1, 2..., n, \\ -a_i - \Delta a_i^- \leqslant -a_{i,0}, \quad i = 1, 2..., n, \\ \delta_i^b \Delta a_i^+ + \delta_i^s \Delta a_i^- \leqslant c_i, \quad i = 1, 2..., n, \\ \sum\limits_{i=1}^{n} (\delta_i^b \Delta a_i^+ + \delta_i^s \Delta a_i^-) \leqslant c, \\ \sum\limits_{i=1}^{n} a_i + \sum\limits_{i=1}^{n} (\delta_i^b \Delta a_i^+ + \delta_i^s \Delta a_i^-) = 1, \\ l_i \leqslant a_i \leqslant u_i, \quad i = 1, 2..., n, \\ \Delta a_i^+ \geqslant 0, \Delta a_i^- \geqslant 0, \quad i = 1, 2..., n, \end{array} \right\},$$

where $\mathbf{\Delta a}^+ = (\Delta a_1^+, \Delta a_2^+, \cdots, \Delta a_n^+)'$, $\mathbf{\Delta a}^- = (\Delta a_1^-, \Delta a_2^-, \cdots, \Delta a_n^-)'$.

Then problem (P$_0$) is equivalent to the following optimization with only linear constraints:

$$(\mathrm{P}_1) \qquad \min_{\mathbf{a}, \mathbf{\Delta a}^+, \mathbf{\Delta a}^- \in \Omega} OF = \lambda \left(\mathrm{E} \left[|\mathbf{a'r} - r_I|^\gamma \right] \right)^{1/\gamma} - (1 - \lambda) \mathrm{E} \left[\mathbf{a'r} - r_I \right].$$

3.2. The index tracking model with the CVaR constraint

We emphasize that model (P$_1$) ignores extreme risk control. This means a portfolio may suffer from market jump risk because it closely replicates a benchmark index. Hence, it is crucial to control downside risk measured by CVaR, which is a popularly applied convex risk measure. If $X = \mathbf{a'r}$ is a tracking portfolio's return whose distribution function is $P(\mathbf{a}, x) = \Pr(\mathbf{a'r} \leqslant x)$, then according to Eqs.(1)-(2), this tracking portfolio VaR is $VaR(\mathbf{a}, \alpha) := -\inf\{x \in R : P(\mathbf{a}, x) \geqslant \alpha\}$, and CVaR is $CVaR(\mathbf{a}, \alpha) := -\mathrm{E}[\mathbf{a'r} | \mathbf{a'r} \leqslant -VaR(\mathbf{a}, \alpha)]$.

Suppose an investor's maximum risk tolerance is ρ , then we propose our index tracking model with the CVaR constraint to be

$$(\mathrm{P}_{CVaR}) \quad \left\{ \begin{array}{l} \min\limits_{\mathbf{a}, \mathbf{\Delta a}^+, \mathbf{\Delta a}^- \in \Omega} OF = \lambda \left(\mathrm{E} \left[|\mathbf{a'r} - r_I|^\gamma \right] \right)^{1/\gamma} - (1 - \lambda) \mathrm{E} \left[\mathbf{a'r} - r_I \right], \\ s.t.\ \ CVaR(\mathbf{a}, \alpha) \leqslant \rho. \end{array} \right.$$

Next, we prove that (P$_{CVaR}$) is a convex optimization model. The following two lemmas serve for Theorem 1.

Lemma 2: *For $\gamma \geqslant 1$ and random variables $X_1,\ X_2,\ \cdots,\ X_m$ satisfying $(E\,|X_j|^{\gamma})^{\frac{1}{\gamma}} < \infty$, $j = 1, 2, \cdots, m$, then we have*

$$\left(E\left|\sum_{j=1}^{m} X_j\right|^{\gamma}\right)^{\frac{1}{\gamma}} \leqslant \sum_{j=1}^{m} (E\,|X_j|^{\gamma})^{\frac{1}{\gamma}}.$$

Proof: see Lin and Bai (2010).

Lemma 3: *For any distribution satisfying regularity condition, $CVaR(\mathbf{a}, \alpha)$ is a convex function of portfolio position $\mathbf{a}$.*

The proof follows immediately from Corollary 11 in Rockafellar and Uryasev (2002).

Theorem 1: *Suppose $\gamma \geqslant 1$, for any distribution satisfying regularity condition, problem (P_{CVaR}) is a convex optimization model.*

Proof: All other constraints in problem (P_{CVaR}) are linear except CVaR, and by Lemma 3, $CVaR(\mathbf{a}, \alpha)$ is a convex function of portfolio position $\mathbf{a}$. Therefore, the feasible set of problem (P_{CVaR}) is a convex set, so we need only to prove that the objective function of problem (P_{CVaR}) is a convex function of decision variable $\mathbf{a}$. The objective function of problem (P_{CVaR}) is $\lambda\,(\mathrm{E}\,[|\mathbf{a}'\mathbf{r} - r_I|^{\gamma}])^{1/\gamma} - (1-\lambda)\mathrm{E}\,[\mathbf{a}'\mathbf{r} - r_I]$. Note that the second part $(1-\lambda)\mathrm{E}\,[\mathbf{a}'\mathbf{r} - r_I]$ is a linear function of $\mathbf{a}$, so we need only to prove that the first part $\lambda\,(\mathrm{E}\,[|\mathbf{a}'\mathbf{r} - r_I|^{\gamma}])^{1/\gamma}$ is a convex function of decision variable $\mathbf{a}$. In the following, we prove that $f(\mathbf{a}) = (\mathrm{E}\,[|\mathbf{a}'\mathbf{r} - r_I|^{\gamma}])^{1/\gamma}$ is a convex function of $\mathbf{a}$.

For any two decision variables $\mathbf{a}_1$ and $\mathbf{a}_2$, and any number κ satisfying $\kappa \in [0, 1]$. Then by Lemma 2, we have

$$\begin{aligned}
&f(\kappa\mathbf{a}_1 + (1-\kappa)\mathbf{a}_2) = (\mathrm{E}\,[|(\kappa\mathbf{a}'_1 + (1-\kappa)\mathbf{a}'_2)\,\mathbf{r} - r_I|^{\gamma}])^{1/\gamma}\\
&= (\mathrm{E}\,[|\kappa(\mathbf{a}'_1\mathbf{r} - r_I) + (1-\kappa)(\mathbf{a}'_1\mathbf{r} - r_I)|^{\gamma}])^{1/\gamma}\\
&\leqslant (\mathrm{E}\,[|\kappa(\mathbf{a}'_1\mathbf{r} - r_I)|^{\gamma}])^{1/\gamma} + (\mathrm{E}\,[|(1-\kappa)(\mathbf{a}'_1\mathbf{r} - r_I)|^{\gamma}])^{1/\gamma}\\
&= \kappa\,(\mathrm{E}\,[|\mathbf{a}'_1\mathbf{r} - r_I|^{\gamma}])^{1/\gamma} + (1-\kappa)\,(\mathrm{E}\,[|\mathbf{a}'_1\mathbf{r} - r_I|^{\gamma}])^{1/\gamma}\\
&= \kappa f(\mathbf{a}_1) + (1-\kappa) f(\mathbf{a}_2),
\end{aligned}$$

which means that $\lambda\,(\mathrm{E}\,[|\mathbf{a}'\mathbf{r} - r_I|^{\gamma}])^{1/\gamma} = \lambda f(\mathbf{a})$ is a convex function of $\mathbf{a}$. This completes the proof of theorem 1.

3.3. The NPK index tracking model with the CVaR constraint

Only under specific distributions we can obtain objective function OF and CVaR in problem (P_{CVaR}). However, in reality, we have limited information about distribution functions. Therefore, we have to use sample data to estimate objective function OF and CVaR in problem (P_{CVaR}) to build data-driven optimization models.

First, we use sample data to estimate CVaR. Suppose $\mathbf{r}_t = (r_{1,t}, r_{2,t}, ..., r_{n,t})'$ is the sample return of risky assets for $t = 1, 2, ..., T$, then $x_t = \mathbf{a}'\mathbf{r}_t$ is the sample return of the tracking portfolio, whose sample standard deviation is $\hat{\sigma}(x) = \sqrt{\mathbf{a}'\hat{\Sigma}\mathbf{a}}$, where $\hat{\Sigma} = \frac{1}{T-1}\sum_{t=1}^{T}(\mathbf{r}_t - \bar{\mathbf{r}})(\mathbf{r}_t - \bar{\mathbf{r}})'$and $\bar{\mathbf{r}} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{r}_t$. Using sample data and Eq. (8), we have the tracking portfolio's NPK estimator of CVaR as

$$\begin{aligned} CVaR_{npk}(\mathbf{a},\alpha) &= \min_{v} \hat{F}_{\alpha}(\mathbf{a},v) \\ &= \min_{v}\left(v - \frac{1}{T\alpha}\sum_{t=1}^{T}\left((\mathbf{a}'\mathbf{r}_t + v)G\left(\frac{-v-\mathbf{a}'\mathbf{r}_t}{b_0\sqrt{\mathbf{a}'\hat{\Sigma}\mathbf{a}}}\right) + b_0\sqrt{\mathbf{a}'\hat{\Sigma}\mathbf{a}}H\left(\frac{-v-\mathbf{a}'\mathbf{r}_t}{b_0\sqrt{\mathbf{a}'\hat{\Sigma}\mathbf{a}}}\right)\right)\right), \end{aligned} \tag{9}$$

Second, we use sample data to derive the NPK estimator of OF. To obtain the NPK estimator of OF , we derive Theorem 2 and Theorem 3 as follows:

Theorem 2: *For any random variable X , the NPK estimator of* $\mathrm{E}\left[|X|^{\gamma}\right]$ *is*

$$\hat{\mathrm{E}}\left[|X|^{\gamma}\right] = \begin{cases} \frac{1}{T}\sum_{t=1}^{T}\left(\sum_{i=0}^{\gamma} C_{\gamma}^{i} x_t^{\gamma-i} b^i\left(-F_i(-\frac{x_t}{b}) + (-1)^i F_i(\frac{x_t}{b})\right)\right), \ \gamma = 1,3,5,\cdots, \\ \frac{1}{T}\sum_{t=1}^{T}\left(\sum_{i=0}^{\gamma} C_{\gamma}^{i} x_t^{\gamma-i} b^i M_i\right), \ \gamma = 0,2,4,\cdots, \end{cases}$$

where $b = b_0\hat{\sigma}(x)$,$F_i(z) = \int_{-\infty}^{z} y^i g(y)\,dy$, $M_i = \int_{-\infty}^{+\infty} y^i g(y)\,dy = \lim_{z\to+\infty} F_i(z)$.

Proof: see Appendix A.

Note that $g(z) = \varphi(z) \triangleq \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2}z^2}$ is the Gauss kernel function, which follows that

$$\begin{cases} F_0(z) = \int_{-\infty}^{z} g(y)dy = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z}\mathrm{e}^{-\frac{1}{2}y^2}dy = \Phi(z), \\ F_1(z) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z} y\mathrm{e}^{-\frac{1}{2}y^2}dy = -\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2}z^2} = -\varphi(z), \end{cases} \tag{10}$$

where $\Phi(z)$ is the standard normal distribution function. In order to derive the expression of $F_i(z)$ with any nonnegative integer i , we give the iterative formula for $F_i(z)$.

Theorem 3: *For any integer $i \geqslant 2$, we have*

$$F_i(z) = -z^{i-1}\varphi(z) + (i-1)F_{i-2}(z).$$

Proof: see Appendix B.

With Theorem 3 and Eq. (10), we obtain the expression of $F_i(z)$ for $i = 2, 3, 4, \cdots$. Because $M_i = \lim_{z\to+\infty} F_i(z)$, according to Eq. (10), we have $M_0 = \lim_{z\to+\infty} F_0(z) = \Phi(+\infty) = 1$, and $M_1 = \lim_{z\to+\infty} F_1(z) = -\varphi(+\infty) = 0$.

By Theorem 3, for any integer $i \geqslant 2$, we have

$$\lim_{z\to+\infty} F_i(z) = -\lim_{z\to+\infty} z^{i-1}\varphi(z) + (i-1)\lim_{z\to+\infty} F_{i-2}(z) = (i-1)\lim_{z\to+\infty} F_{i-2}(z)$$

where, $\lim_{z\to+\infty} z^{i-1}\varphi(z) = \lim_{z\to+\infty} \frac{z^{i-1}}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}z^2} = 0$. When i is an even number,

$$M_i = \lim_{z\to+\infty} F_i(z) = (i-1)\times(i-3)\times(i-5)\times \ldots \times 3\times 1\times \lim_{z\to+\infty} F_0(z) = (i-1)!!$$

When i is an odd number,

$$M_i = \lim_{z\to+\infty} F_i(z) = (i-1)\times(i-3)\times(i-5)\times \ldots \times 2\times \lim_{z\to+\infty} F_1(z) = 0$$

Let $X = \mathbf{a}'\mathbf{r} - r_I$, and $x_t = \mathbf{a}'\mathbf{r}_t - r_{I,t}$ $(t = 1, 2, ..., T)$ are sample data of $\mathbf{a}'\mathbf{r} - r_I$, $r_{I,t}$ is sample returns of r_I. By Theorem 2, we get the NPK estimator of TE:

$$\widehat{TE}^{npk} = \left(\hat{\mathrm{E}}\left[|\mathbf{a}'\mathbf{r} - r_I|^{\gamma}\right]\right)^{1/\gamma} = \\ \begin{cases} \left(\frac{1}{T}\sum\limits_{t=1}^{T}\left(\sum\limits_{i=0}^{\gamma} C_{\gamma}^{i}\left(\mathbf{a}'\mathbf{r}_t - r_{I,t}\right)^{\gamma-i} b^i \left(-F_i(-\frac{\mathbf{a}'\mathbf{r}_t - r_{I,t}}{b}) + (-1)^i F_i(\frac{\mathbf{a}'\mathbf{r}_t - r_{I,t}}{b})\right)\right)\right)^{1/\gamma}, & \gamma = 2m+1, \\ \left(\frac{1}{T}\sum\limits_{t=1}^{T}\left(\sum\limits_{i=0}^{\gamma} C_{\gamma}^{i}\left(\mathbf{a}'\mathbf{r}_t - r_{I,t}\right)^{\gamma-i} b^i M_i\right)\right)^{1/\gamma}, & \gamma = 2m, \end{cases} \tag{11}$$

where $m = 0, 1, 2, \cdots$, $b = b_0\hat{\sigma}(x) = b_0\sqrt{\mathbf{a}'\hat{\Sigma}\mathbf{a} - 2\mathbf{a}'\hat{g} + \hat{\sigma}_I^2}$, $\hat{\sigma}_I^2$ is the sample variance of r_I , $\hat{g} = \frac{1}{T-1}\sum\limits_{t=1}^{T}(\mathbf{r}_t - \bar{\mathbf{r}})(r_{I,t} - \bar{r}_I)$ is the covariance vector of $\mathbf{r}$ with r_I and $\bar{r}_I = \frac{1}{T}\sum\limits_{t=1}^{T} r_{I,t}$.

Now, we have the NPK estimator of OF:

$$OF^{npk} = \lambda\widehat{TE}^{npk} - (1-\lambda)\frac{1}{T}\sum_{t=1}^{T}(\mathbf{a}'\mathbf{r}_t - r_{I,t}).$$

Lemma 4 *(Rockafellar and Uryasev, 2000, 2002): The following two optimization problems are equivalent:*

$$\min_{\mathbf{a}\in X} f(\mathbf{a}), \quad s.t.\ CVaR(\mathbf{a},\alpha) \leqslant \rho, \quad \Leftrightarrow \quad \min_{(\mathbf{a},v)\in X\times R} f(\mathbf{a}), \quad s.t.\ F_{\alpha}(\mathbf{a},v) \leqslant \rho.$$

Utilizing Lemma 4, we substitute the NPK estimator of OF (OF^{npk}) and the NPK estimator of CVaR - Eq. (9) - into the index tracking problem (P_{CVaR}). We propose our NPK index tracking model with CVaR to be

$$(\mathrm{P}_{CVaR_{npk}}) \quad \begin{cases} \min\limits_{\mathbf{a},\boldsymbol{\Delta}\mathbf{a}^+,\boldsymbol{\Delta}\mathbf{a}^-\in\Omega,\ v\in\Re} & OF^{npk} = \lambda\widehat{TE}^{npk} - (1-\lambda)\frac{1}{T}\sum\limits_{t=1}^{T}(\mathbf{a}'\mathbf{r}_t - r_{I,t}), \\ s.t. \quad v - \frac{1}{T\alpha}\sum\limits_{t=1}^{T}\left((\mathbf{a}'\mathbf{r}_t + v)G\left(\frac{-v-\mathbf{a}'\mathbf{r}_t}{b_0\sqrt{\mathbf{a}'\hat{\Sigma}\mathbf{a}}}\right) + b_0\sqrt{\mathbf{a}'\hat{\Sigma}\mathbf{a}}H\left(\frac{-v-\mathbf{a}'\mathbf{r}_t}{b_0\sqrt{\mathbf{a}'\hat{\Sigma}\mathbf{a}}}\right)\right) \leqslant \rho, \end{cases}$$

where $\widehat{TE}^{npk}$ is given by Eq. (11) for any positive number γ .

When the $\gamma = 1$, according to Wang et al. (2012), the problem (P_{CVaR}) can be transformed into a linear programming (LP) index tracking optimization problem:

$$(\mathrm{P}_{CVaR_{lp}}) \quad \begin{cases} \min\limits_{\substack{\mathbf{a},\boldsymbol{\Delta}\mathbf{a}^+,\boldsymbol{\Delta}\mathbf{a}^-\in\Omega, \\ \eta,\mathbf{z}^+,\mathbf{z}^-\in\Re^T,v\in\Re}} & OF^{lp} = \frac{\lambda}{T}\sum\limits_{t=1}^{T}(z_t^+ + z_t^-) - \frac{1-\lambda}{T}\sum\limits_{t=1}^{T}(\mathbf{a}'\mathbf{r}_t - r_{I,t}), \\ s.t. \quad v + \frac{1}{T\alpha}\sum\limits_{t=1}^{T}\eta_t \leqslant \rho, \\ \qquad -\mathbf{a}'\mathbf{r}_t - v - \eta_t \leqslant 0, \quad t = 1,2,...,T, \\ \qquad z_t^+ - z_t^- - \mathbf{a}'\mathbf{r}_t = -r_{I,t}, \quad t = 1,2,...,T, \\ \qquad \eta_t \geqslant 0, z_t^+ \geqslant 0, z_t^- \geqslant 0, \quad t = 1,2,...,T, \end{cases}$$

where $\eta = (\eta_1, \eta_2, \cdots, \eta_T)', \mathbf{z}^+ = (z_1^+, z_2^+, \cdots, z_T^+)', \mathbf{z}^- = (z_1^-, z_2^-, \cdots, z_T^-)'$ are ancillary variables, subject to the conditions (Rudolf et al., 1999; Krokhmal et al., 2002; Wang et al., 2012): $z_t^+ + z_t^- = |\mathbf{a}'\mathbf{r}_t - r_{I,t}|, z_t^+ - z_t^- = \mathbf{a}'\mathbf{r}_t - r_{I,t}, \eta_t = (-\mathbf{a}'\mathbf{r}_t - v)^+$.

In the following simulation and empirical sections, we concentrate on comparing the performance of our proposed NPK model ($\mathrm{P}_{CVaR_{npk}}$) with the performance of the LP model ($\mathrm{P}_{CVaR_{lp}}$). Notice that after imposing CVaR as a constraint, the number of decision variables increases by $3T+1$ and the number of constraints increases by $2T+1$ in the ($\mathrm{P}_{CVaR_{lp}}$) model. By comparison, in the ($\mathrm{P}_{CVaR_{npk}}$) model, the number of decision variables and constraints only increases by one, respectively. Hence, when the sample size is large, our NPK index tracking model an effectiely mitigate the computational difficulty compared with the LP model. In this paper, we use the *cplexlp* routine to solve model $P_{CVaR_{lp}}$ and use $fmincon$ routine to solve model $P_{CVaR_{npk}}$. In addition, we carry out tests using a Macbook Pro with 2.6GHz dual-core Intel i5 processor and 8GB memory. The software suites used are MATLAB 2013 and CPLEX V12.6.1.

4. Simulations

In Section 4, we carry out Monte Carlo simulations to compare the performance of model $\left(\mathrm{P}_{CVaR_{npk}}\right)$ with model $\left(\mathrm{P}_{CVaR_{lp}}\right)$. In Section 4.1, we first derive analytical expressions of model (P_{CVaR}) under normal, t and AL distributions. Utilizing these true expressions as a benchmark, we compare the accuracy of model $\left(\mathrm{P}_{CVaR_{npk}}\right)$ with the accuracy of model $\left(\mathrm{P}_{CVaR_{lp}}\right)$ in Section 4.2. Finally, in Section 4.3, we simulate market states to examine the out-of-sample performance of models $\left(\mathrm{P}_{CVaR_{npk}}\right)$ and $\left(\mathrm{P}_{CVaR_{lp}}\right)$.

4.1. Analytical expressions of model (P_{CVaR})

First, we illustrate the analytical expressions of CVaR under normal, t and AL distributions as follows:

If asset returns $\mathbf{r} \sim N(\mu, \Sigma)$, then the tracking portfolio's return $\mathbf{a}'\mathbf{r} \sim N(\mathbf{a}'\mu, \mathbf{a}'\Sigma\mathbf{a})$, according to Eq. (3), we have $CVaR_N(\mathbf{a}, \alpha) = k_1(\alpha)\sqrt{\mathbf{a}'\Sigma\mathbf{a}} - \mathbf{a}'\mu$.

If asset returns $\mathbf{r}$ follow the n dimensional $t(\mu, \Sigma, m)$, then the tracking portfolio's return $\mathbf{a}'\mathbf{r}$ follows the one-dimensional $t(\mathbf{a}'\mu, \mathbf{a}'\Sigma\mathbf{a}, m)$. Therefore, according to Eq. (4), we have $CVaR_t(\mathbf{a}, \alpha) = k_2(\alpha)\sqrt{\mathbf{a}'\Sigma\mathbf{a}} - \mathbf{a}'\mu$.

In order to obtain the analytical expression of CVaR under AL distribution, we introduce the following Lemma 5:

Lemma 5 *(Kotz et al., 2012): Suppose that random vector* $\mathbf{r} = (r_1, r_2, ..., r_n)' \sim AL(\mu, \Sigma)$ *and* $\mathbf{a} = (a_1, a_2, ..., a_n)'$ *is an* $n \times 1$ *real vector. Then the random variable* $\mathbf{a}'\mathbf{r} \sim AL(\mathbf{a}'\mu, \mathbf{a}'\Sigma\mathbf{a})$ *.*

According to Lemma 5 and Eq. (6), if asset returns $\mathbf{r} \sim AL(\mu, \Sigma)$, then the CVaR of a tracking portfolio is

$$\begin{aligned} CVaR_{AL}(\mathbf{a}, \alpha) = & -\frac{\mathbf{a}'\Sigma\mathbf{a}}{\mathbf{a}'\mu + \sqrt{(\mathbf{a}'\mu)^2 + 2\mathbf{a}'\Sigma\mathbf{a}}} \ln\alpha \left(2 + \frac{(\mathbf{a}'\mu)^2 + \mathbf{a}'\mu\sqrt{(\mathbf{a}'\mu)^2 + 2\mathbf{a}'\Sigma\mathbf{a}}}{\mathbf{a}'\Sigma\mathbf{a}}\right) \\ & + \frac{\mathbf{a}'\Sigma\mathbf{a}}{\mathbf{a}'\mu + \sqrt{(\mathbf{a}'\mu)^2 + 2\mathbf{a}'\Sigma\mathbf{a}}}. \end{aligned} \tag{12}$$

Next, we derive the analytical expressions of OF under these three distributions. We emphasize that we can give analytical expressions of OF in any order $\gamma \geqslant 1$ for model (P_{CVaR}) . However, we only give analytical expressions at $\gamma = 1$ in order to compare with model $\left(\mathrm{P}_{CVaR_{lp}}\right)$ in the following simulation subsections[1]:

[1]Another motivation is that when $\gamma = 1$ the tracking error is measured by absolute deviation (Sharpe, 1971; Konno and Yamazaki, 1991), which has been widely applied to reward fund managers' performance (Clarke et al., 1994; Rudolf et al., 1999).

If asset returns and benchmark index returns $\tilde{\mathbf{r}} = (\mathbf{r}', r_I)' \sim N(\tilde{\mu}, \tilde{\Sigma})$, where $\tilde{\mu} = \begin{pmatrix} \mu \\ \mu_I \end{pmatrix}$ and $\tilde{\Sigma} = \begin{pmatrix} \Sigma & g \\ g' & \sigma_I^2 \end{pmatrix}$, then we have $\mathbf{r} \sim N(\mu, \Sigma)$, $r_I \sim N(\mu_I, \sigma_I^2)$ and $\mathbf{a}'\mathbf{r} - r_I \sim N(\mu_0, \sigma_0^2)$, where $\mu_0 = \mathbf{a}'\mu - \mu_I$ and $\sigma_0^2 = \mathbf{a}'\Sigma\mathbf{a} - 2\mathbf{a}'g + \sigma_I^2$. Thus the excess return is $\mathrm{E}[\mathbf{a}'\mathbf{r} - r_I] = \mu_0$, and TE is

$$\begin{aligned}
\mathrm{E}[|\mathbf{a}'\mathbf{r} - r_I|] &= \mathrm{E}\left|\tfrac{\mathbf{a}'\mathbf{r} - r_I - \mu_0}{\sigma_0} + \tfrac{\mu_0}{\sigma_0}\right| \sigma_0 \\
&= \mathrm{E}\left|Y + \tfrac{\mu_0}{\sigma_0}\right| \sigma_0 = \int_{-\infty}^{\infty} \left|y + \tfrac{\mu_0}{\sigma_0}\right| \sigma_0 \varphi(y) dy \\
&= -\int_{-\infty}^{-\frac{\mu_0}{\sigma_0}} \left(y + \tfrac{\mu_0}{\sigma_0}\right) \sigma_0 \varphi(y) dy + \int_{-\frac{\mu_0}{\sigma_0}}^{\infty} \left(y + \tfrac{\mu_0}{\sigma_0}\right) \sigma_0 \varphi(y) dy \\
&= 2\sigma_0 \varphi\left(-\tfrac{\mu_0}{\sigma_0}\right) - 2\mu_0 \Phi\left(-\tfrac{\mu_0}{\sigma_0}\right) + \mu_0,
\end{aligned} \tag{13}$$

where $Y = \frac{\mathbf{a}'\mathbf{r} - r_I - \mu_0}{\sigma_0} \sim N(0,1)$, $\varphi(\cdot)$ and $\Phi(\cdot)$ are density and cumulative distribution functions of standard normal distribution, respectively.

Hence, the analytical expression of model (P_{CVaR}) under normal distribution can be formulated as

$$(\mathrm{P}_{CVaR_N}) \quad \begin{cases} \min\limits_{\mathbf{a}, \mathbf{\Delta a}^+, \mathbf{\Delta a}^- \in \Omega} OF^N = \lambda \left[2\sigma_0 \varphi\left(-\tfrac{\mu_0}{\sigma_0}\right) - 2\mu_0 \Phi\left(-\tfrac{\mu_0}{\sigma_0}\right) + \mu_0\right] - (1-\lambda)\mu_0, \\ s.t. \;\; \mu_0 = \mathbf{a}'\mu - \mu_I, \;\; \sigma_0 = \sqrt{\mathbf{a}'\Sigma\mathbf{a} - 2\mathbf{a}'g + \sigma_I^2}, \\ \quad k_1(\alpha)\sqrt{\mathbf{a}'\Sigma\mathbf{a}} - \mathbf{a}'\mu \leqslant \rho. \end{cases}$$

If the joint distribution of the asset returns and benchmark return, $\tilde{\mathbf{r}} = (\mathbf{r}', r_I)' \sim t(\tilde{\mu}, \tilde{\Sigma}, m)$, where $\tilde{\mu}$ and $\tilde{\Sigma}$ are the same as normal distribution above, then $\mathbf{r} \sim t(\mu, \Sigma, m)$, $r_I \sim t(\mu_I, \sigma_I^2, m)$ and $\mathbf{a}'\mathbf{r} - r_I \sim t(\mu_0, \sigma_0^2, m)$, where $\mu_0 = \mathbf{a}'\mu - \mu_I$ and $\sigma_0^2 = \mathbf{a}'\Sigma\mathbf{a} - 2\mathbf{a}'g + \sigma_I^2$. Thus excess return is $\mathrm{E}[\mathbf{a}'\mathbf{r} - r_I] = \mu_0$ and TE is

$$\mathrm{E}[|\mathbf{a}'\mathbf{r} - r_I|] = \frac{2\sigma_0}{m-1}\left(m + \left(\frac{\mu_0}{\sigma_0}\right)^2\right) f_t\left(-\frac{\mu_0}{\sigma_0}\right) - 2\mu_0 F_t\left(-\frac{\mu_0}{\sigma_0}\right) + \mu_0, \tag{14}$$

where $f_t(\cdot)$ and $F_t(\cdot)$ are the density and cumulative distribution function of one-dimensional t-distribution.

Hence, the analytical expression of model (P_{CVaR}) under t-distribution can be formulated as

$$(\mathrm{P}_{CVaR_t}) \quad \begin{cases} \min\limits_{\mathbf{a}, \mathbf{\Delta a}^+, \mathbf{\Delta a}^- \in \Omega} OF^t = \lambda \left[\tfrac{2\sigma_0}{m-1}\left(m + \left(\tfrac{\mu_0}{\sigma_0}\right)^2\right) f_t\left(-\tfrac{\mu_0}{\sigma_0}\right) - 2\mu_0 F_t\left(-\tfrac{\mu_0}{\sigma_0}\right) + \mu_0\right] - (1-\lambda)\mu_0, \\ s.t. \;\; \mu_0 = \mathbf{a}'\mu - \mu_I, \;\; \sigma_0 = \sqrt{\mathbf{a}'\Sigma\mathbf{a} - 2\mathbf{a}'g + \sigma_I^2}, \\ \quad k_2(\alpha)\sqrt{\mathbf{a}'\Sigma\mathbf{a}} - \mathbf{a}'\mu \leqslant \rho. \end{cases}$$

Finally, by Lemma 5, if $\tilde{\mathbf{r}} = (\mathbf{r}', r_I)' \sim AL(\tilde{\mu}, \tilde{\Sigma})$, where $\tilde{\mu}$ and $\tilde{\Sigma}$ are the same as normal distribution above, then $\mathbf{r} \sim AL(\mu, \Sigma)$, $r_I \sim AL(\mu_I, \sigma_I^2)$, and $\mathbf{a}'\mathbf{r} - r_I \sim AL(\mu_0, \sigma_0^2)$, where $\mu_0 = \mathbf{a}'\mu - \mu_I$ and $\sigma_0^2 = \mathbf{a}'\Sigma\mathbf{a} - 2\mathbf{a}'g + \sigma_I^2$. Thus excess return is $\mathrm{E}\left[\mathbf{a}'\mathbf{r} - r_I\right] = \mu_0$ and according to Kotz et al. (2012), we have TE

$$\mathrm{E}\left[|\mathbf{a}'\mathbf{r} - r_I|\right] = \frac{\sigma_0}{\sqrt{2}k_0}\frac{1+k_0^4}{1+k_0^2}, \tag{15}$$

where $k_0 = \frac{\sqrt{2}\sigma_0}{\mu_0 + \sqrt{\mu_0^2 + 2\sigma_0^2}}$.

Hence, the analytical expression of model (P_{CVaR}) under AL distribution can be formulated as

$$(\mathrm{P}_{CVaR_{AL}}) \quad \begin{cases} \min\limits_{\mathbf{a}, \boldsymbol{\Delta}\mathbf{a}^+, \boldsymbol{\Delta}\mathbf{a}^- \in \Omega} \quad OF^{AL} = \lambda\left[\frac{\sigma_0}{\sqrt{2}k_0}\frac{1+k_0^4}{1+k_0^2}\right] - (1-\lambda)\mu_0, \\ s.t. \;\; k_0 = \frac{\sqrt{2}\sigma_0}{\mu_0 + \sqrt{\mu_0^2 + 2\sigma_0^2}}, \; \mu_0 = \mathbf{a}'\mu - \mu_I, \; \sigma_0 = \sqrt{\mathbf{a}'\Sigma\mathbf{a} - 2\mathbf{a}'g + \sigma_I^2}, \\ \quad CVaR_{AL}(\mathbf{a}, \alpha) \leqslant \rho. \end{cases}$$

We can use the optimizer tool $fmincon$ in MATLAB to obtain the true index tracking strategies and the true objective function value OF^N, OF^t and OF^{AL} from models (P_{CVaR_N}), (P_{CVaR_t}) and ($\mathrm{P}_{CVaR_{AL}}$) respectively.[1]

4.2. Accuracy simulation

In Section 4.2, we compare the estimation accuracy of the NPK model $\left(\mathrm{P}_{CVaR_{npk}}\right)$ and the LP model$\left(\mathrm{P}_{CVaR_{lp}}\right)$ using true OF values obtained in Section 4.1 under normal, t and AL distributions. We not only carry out sample size examinations by setting T from 250 to 5000, but also portfolio size examinations by setting n from 20 to 400. We assume n assets and one index together follow $n+1$-dimensional $N(\tilde{\mu}, \tilde{\Sigma})$, $t(\tilde{\mu}, \tilde{\Sigma}, m)$ and $AL(\tilde{\mu}, \tilde{\Sigma})$, where $m = 5$. We generate sample data sets from $N(\tilde{\mu}, \tilde{\Sigma})$, $t(\tilde{\mu}, \tilde{\Sigma}, m)$ and $AL(\tilde{\mu}, \tilde{\Sigma})$ respectively, according to the methods in Appendix C. $\tilde{\mu}$ and $\tilde{\Sigma}$ are from uniform distribution.[2]

The number of repetitive sampling N in sample (portfolio) size tests is 1000 (500). We denote the objective function value in the jth simulation by $\widehat{OF}_j^{lp}$ and $\widehat{OF}_j^{npk}$ -

[1] According to Theorem 1, the three models P_{CVaR_N}, P_{CVaR_t}, $P_{CVaR_{AL}}$ are convex optimization problems. Therefore, the $fmincon$ routine can obtain globally optimal solutions. In addition, the $exitflag$ reported by $fmincon$ equals 1, which means that first-order optimality conditions are satisfied and that the solutions are locally optimal. Theorem 1 verifies that these locally optimal solutions are globally optimal.

[2] We extract elements from uniform distribution to have a matrix A. Furthermore, we define $\tilde{\Sigma} = AA^T$ to make sure $\tilde{\Sigma}$ is a positive definite matrix. These data are available upon requests.

obtained by solving LP model $\left(\mathrm{P}_{CVaR_{lp}}\right)$ and NPK model $\left(\mathrm{P}_{CVaR_{npk}}\right)$, respectively - and the true objective function value by OF - obtained by solving the true models (P_{CVaR_N}) , (P_{CVaR_t}) , (P_{CVaR_AL}) under normal, t and AL distributions, respectively. We define statistical indicators of estimation errors as follows:

$$mse^{lp} = \frac{1}{N}\sum_{j=1}^{N}\left(\widehat{OF}_j^{lp} - OF\right)^2, \quad mse^{npk} = \frac{1}{N}\sum_{j=1}^{N}\left(\widehat{OF}_j^{npk} - OF\right)^2.$$

where mse^{lp} and mse^{npk} represent the mean square error of LP model and NPK model, respectively. In order to compare the NPK model's $\widehat{OF}^{npk}$ with the LP model's $\widehat{OF}^{lp}$, we define

$$\Delta\% = \frac{mse^{lp} - mse^{npk}}{mse^{lp}} \times 100\%, \quad freq = \frac{1}{N}\sum_{j=1}^{N}\mathrm{I}\left(\left|\widehat{OF}_j^{lp} - OF\right| > \left|\widehat{OF}_j^{npk} - OF\right|\right),$$

where $I(\cdot)$ is an indicator function.

Clearly, if $freq > 0.5$, then among N samples, the frequency of the NPK method outperforming the LP method is greater than 50%. In addition, $\Delta\%$ measures the extent to which the NPK method outperforms LP in terms of estimation accuracy. Moreover, by constructing a statistic z^*, we intend to demonstrate that the advantage of NPK method in terms of estimation accuracy is statistically significant. As the $\widehat{OF}_j^{lp}$ and $\widehat{OF}_j^{npk}$ are random variables, we denote $Prob(|\widehat{OF}_j^{lp} - OF| > |\widehat{OF}_j^{npk} - OF|) = p$, $j = 1, 2, 3, ..., N$ and define $X_j = \mathrm{I}(|\widehat{OF}_j^{lp} - OF| > |\widehat{OF}_j^{npk} - OF|)$, then $Prob(X_j = 1) = p$ and $Prob(X_j = 0) = 1 - p$. Therefore, X_j follows a Bernoulli distribution and $X = \sum_{j=1}^{N} X_j$ follows a Binomial distribution $B(N, p)$. According to the central limit theorem, we know that $z^* = \frac{X - Np}{\sqrt{Np(1-p)}} \to N(0, 1)$. Hence, we can use z^* to test the null hypothesis $H_0 : p = 0.5$ and the alternative hypothesis $H_1 : p > 0.5$. Under H_0, if $z^* > 1.64(2.33)$, we reject the null hypothesis and conclude that NPK outperforms LP at the 5% (1%) significance level. To compare computational efficiencies, we also report the average time AT^{lp} and AT^{npk} in N simulations (unit: second).

We simulate our models under three cases.[1]

(I) IRM ($\lambda = 1$), which implies that we target to minimize tracking errors, i.e., the objective function $OF = TE$.

[1]For the sake of space, we report detailed results with respect to estimation accuracy and computing time in the online appendix.

The parameter settings are as follows: we assume an investor steps into the market with 1 initial capital, and the initial positions $a_{i,0} = 0$, $i = 1, 2..., n$. The transaction cost in each stock is less than 0.01 and the total transaction cost is less than 0.1, i.e., $c_i = 0.01$, $c = 0.1$. The maximum risk tolerance is 3, i.e., $\rho = 3$. We select $\alpha = 0.01$ and we do not allow short selling, so the lower and upper proportion on each stocks are $l_i = 0$ and $u_i = 1$. The proportional costs for both buying and selling are equal to 0.01, i.e., $\delta_i^b = \delta_i^s = 0.01$, $i = 1, 2, ...n$. In the sample size tests, the portfolio size $n = 10$; in the portfolio size tests, the sample size $T = 1500$.

Table 1 presents the simulation results for IRM. The left half of this table reports the results when the sample size increases (sample size test) and the right half of this table reports the results when the portfolio size increases (portfolio size test). Panel A, B, and C report results under normal, t, and AL distributions, respectively. It is evident that the mse^{npk} values are less than the mse^{lp} values. Specifically, as indicated by $\Delta\%$, in sample size tests, the NPK method enhances estimation accuracy by 9.50% to 24.01% under normal distribution, 52.31% to 82.14% under t distribution, and 47.23% to 72.07% under the AL distribution when compared with the LP method. In portfolio size tests, the NPK method enhances estimation accuracy by 12.36% to 15.18%, 36.59% to 64.05%, and 34.12% to 46.89% under normal, t, and AL distributions, respectively. These numbers mean that the NPK method has more accurate estimation than LP. Next, in the sample size test, it is evident that the $freq$s are at least 70% and z^* values are greater than 12; in portfolio size test, the $freq$s equal 1 and z^* values are around 22.36. These findings mean that NPK outperforms LP in terms of estimation accuracy not only robustly, but also statistically significantly at the 1% level. Last, regarding computing time, in the sample size test, we find that the computing time of LP considerably climbs when T increases; however, the computing time of NPK performs stably when T increases. The LP method requires more time to complete optimization than NPK when T is large. In the portfolio size test, we document that when the portfolio size n is small, the time consumption of the NPK method is less than that of LP. However, when the portfolio size n expands, the computing time of the NPK method quickly increases and exceeds the computing time of LP. We attribute this finding to the facts that NPK is a nonlinear optimization model and that the large portfolio size increases its computing time.

INSERT TABLE 1 HERE

(II) AIM ($\lambda = 0$), which implies that we target to maximize the excess return, i.e., the objective function $OF = -\Delta R$. Minimizing OF is equivalent to maximizing ΔR .

Different from the simulation in Part (I), we set $\alpha = 0.05$ to compare the model performance under different loss probability. We allow short sale, which means that $l_i = -\infty$, $u_i = \infty$. Other parameters are the same as those in Part (I).

Table 2 presents the simulation results regarding AIM. The table structure is the same as Table 1. With respect to the estimation accuracy, it is evident that the mse^{npk} values are less than the mse^{lp} values. Specifically, in sample size tests, the NPK method enhances estimation accuracy by 8.84% to 39.61% under normal distribution, 10.95% to 34.04% under t distribution, and 8.59% to 39.61% under AL distribution when compared with the LP method. In portfolio size tests, the NPK method enhances estimation accuracy by 0.4% to 1.96%, 0.7% to 11.01%, and 0.2% to 0.6% under normal, t, and AL distributions, respectively. These numbers mean that the NPK method has better accuracy estimation than LP. Next, in the sample size test, it is evident that the $freq$s are at least 53% and z^* values are greater than 2.09; in the portfolio size test, the $freq$s are at least 56% and z^* values are greater than 2.95. These findings mean that NPK outperforms LP in terms of estimation accuracy not only robustly but also statistically significantly at the 5% level. Finally, regarding computing time, in the sample size test, we find that the computing time of LP increases significantly and is considerably greater than that of the NPK method. Similarly, in the portfolio size test, we find that when n is small, the NPK method also requires less time than the LP method; however, when n increases, the LP method outperforms NPK. The reasons might be two-fold. First, the objective function of AIM is linear and therefore, unlike EIM and IRM, the LP method does not need to introduce auxiliary variables $\mathbf{z}^+, \mathbf{z}^-$. Second, NPK AIM has a nonlinear constraint while the constraints on LP AIM are all linear.

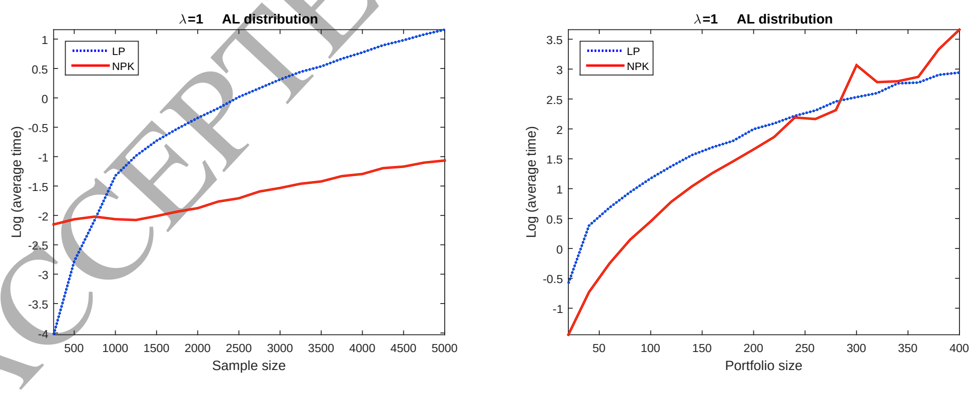
INSERT TABLE 2 HERE

(III) EIM ($\lambda = 0.5$), which implies a mixed strategy to balance the active investment and the index replication, i.e., the objective function $OF = \frac{1}{2}TE - \frac{1}{2}\Delta R$.

Table 3 presents the simulation results for EIM. The table structure is the same as Tables 1 and 2. With respect to the estimation accuracy, it is evident that the mse^{npk} values are lower than the mse^{lp} values. Specifically, in sample size tests, the NPK method enhances estimation accuracy by 13.04% to 31.32%, 33.16% to 51.17%, and 23.78% to 47.48% under normal, t, and AL distributions, respectively, when compared with the LP method. In portfolio size tests, the NPK method enhances estimation accuracy by 10.39% to 20.19%, 17.49% to 47.62%, and 20.82% to 36.22% under normal, t, and AL distributions, respectively. These numbers mean that the NPK method has better estimation accuracy than LP. Next, both in the sample size and portfolio size tests, it is evident that the $freq$s are at least 95% and z^* values are greater than 22. These findings mean that the NPK method outperforms LP in terms of estimation accuracy not only robustly but also statistically significantly at the 1% level. Finally, regarding computing time, in the sample size test, we find that the computing time of LP increases significantly and is considerably greater than that of the NPK method. However, in the portfolio size test, we find that computing time of both LP and NPK increase, but NPK still requires less time than LP.

INSERT TABLE 3 HERE

To sum up, we provide evidence that the NPK method outperforms LP in terms of estimation accuracy. This finding is robust for AIM, IRM and EIM under three distributions and is robust across both sample size and portfolio size tests. With respect to computing time, our findings are mixed. To give a direct impression, as shown in Tables 1 through 3, Figure 1 depicts the computing time curves for the three models under the AL distribution.[1] It is evident that the computing time of the LP method increases significantly with an increase in sample size T; however, the computing time of the NPK method performs quite stably. The LP method spends considerably more time than the NPK method when T is large. In the portfolio size test, regarding IRM, we show that, when n is small, the NPK method outperforms LP in terms of time saving; otherwise, the LP method outperforms the NPK method. Regarding AIM, LP requires less time than the NPK method. Regarding EIM, the computing time of both methods increase with an increase in n, but the NPK method outperforms LP in terms of time consumption overall.



[1]Figures about normal and t distributions are similar with Figure 1. All the figures are reported in the online appendix.

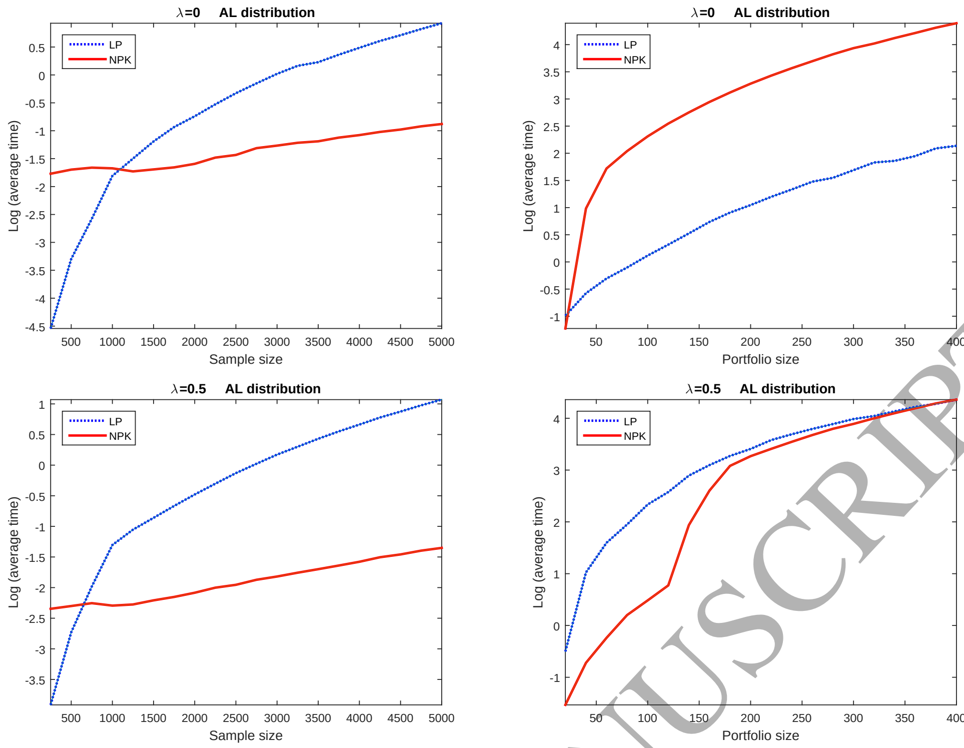


Figure 1. **Computing time when sample size or portfolio size changes: AL distribution.** This figure depicts computing time of IRM ($\lambda = 1$), AIM ($\lambda = 0$) and EIM ($\lambda = 0.5$) under AL distribution. The left column depicts the computing time when the sample size changes. The right column depicts computing time when the portfolio size changes. The vertical axis is the log of the average time (unit: second) over 1000 simulations for sample size tests and 500 simulations for portfolio size tests.

4.3. The NPK model performance in simulated market conditions

In Section 4.3, we implement out-of-sample analyses to compare the performance of the NPK model $\left(\mathrm{P}_{CVaR_{npk}}\right)$ with that of the LP model $\left(\mathrm{P}_{CVaR_{lp}}\right)$ in simulated bullish and bearish markets.[1]

Suppose there are $n = 100$ stocks in the market. We set 30 stocks to follow joint normal distribution $N(\mu_1, \Sigma_1)$, another 30 stocks to follow $t(\mu_2, \Sigma_2, 5)$ and the remaining 40 stocks to follow $AL(\mu_3, \Sigma_3)$. The parameters $\mu_1, \Sigma_1, \mu_2, \Sigma_2, \mu_3, \Sigma_3$ come from random numbers generated by uniform distribution.We generate $T = 1000$ sample returns for these 100 stocks and then construct an index using these 100 stocks. We use 100 random numbers from [0,1] uniform distribution divided by their sum as weights. We assume no knowledge about these weights and track index performance.

[1] We also carry out in-sample analyses and the results are reported in the online appendix.

We split our $T = 1000$ sample returns into two parts. The first 500 returns are training samples and the other 500 returns are used for out-of-sample test. We adopt static index tracking strategies obtained from training samples to carry out out-of-sample examinations. We set the initial wealth as 1 and the other parameters are: $\alpha = 0.01, a_{i,0} = 0, c_i = 0.01, c = 0.1, l_i = 0, u_i = 1, \delta_i^b = \delta_i^s = 0.01, i = 1, 2, ...n.$

We consider three cases: AIM ($\lambda = 0$), EIM ($\lambda = 0.5$) and IRM ($\lambda = 1$).[1] risk tolerance levels ρ are 1 and 2. Moreover, in order to test our models under different market environments, we control the index trend in a random number-generation process. To be specific, we create two cases: index upwards represents the bullish market, and index downwards represents the bearish market. Figure 2 illustrates out-of-sample performances when the market falls. Figure 3 shows out-of-sample performances when the market rises. From Figures 2-3, we have the following findings:

We can obtain maximum excess returns when $\lambda = 0$ (AIM) and minimum excess returns when $\lambda = 1$ (IRM), which is consistent with our expectation in the model settings. This is because $\lambda = 0$ (AIM) implies the objective function is to maximize the excess return, $\lambda = 1$ (IRM) implies the objective is to minimize the tracking error, and $\lambda = 0.5$ (EIM) is a mixed strategy to trade off index replication and excess returns.

When $\lambda = 0$ (AIM), investors can obtain excess returns in both the NPK and LP models, however, the LP model performs greater volatility. When $\lambda = 0.5$ (EIM), both LP and NPK methods can track index trend, however, we find that the NPK method obtains higher excess return. When $\lambda = 1$ (IRM), both methods can replicate index performance. In particular, when $\rho = 2$ three lines almost overlap, which implies the NPK and LP methods almost replicate the weights of the index constituents. It is clear that the tracking error when $\rho = 2$ is less than when $\rho = 1$. This is because, with the decrease in ρ , the feasible set shrinks, which in turn enlarges tracking errors. This finding means the CVaR constraint is efficient.

[1] We also test the models' performance when sample size increases to 2000 and portfolio size increases to 400. In addition, to show that our results apply for more general cases, we also test EIMs when $\lambda = 0.25$ and $\lambda = 0.75$. The results are robust and consistent. All the results are reported in the online appendix.

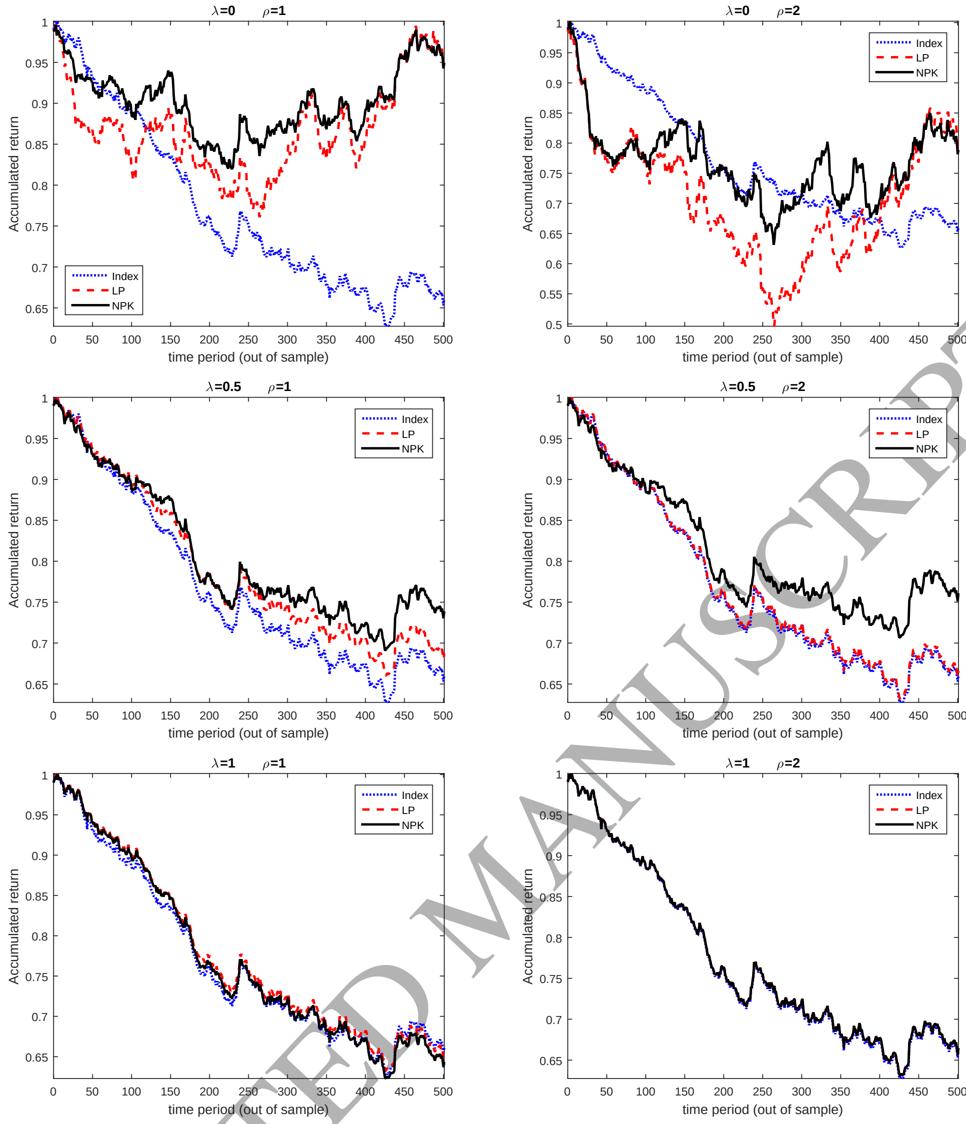


Figure 2. Accumulated returns of index tracking models when the market decreases: out of sample analysis. In this figure, we apply the (static) optimized tracking portfolio obtained from training sample to show the accumulated performance in the out-of-sample period. We show the out-of-sample accumulated returns of AIM ($\lambda = 0$), EIM ($\lambda = 0.5$) and IRM ($\lambda = 1$) when the market falls.

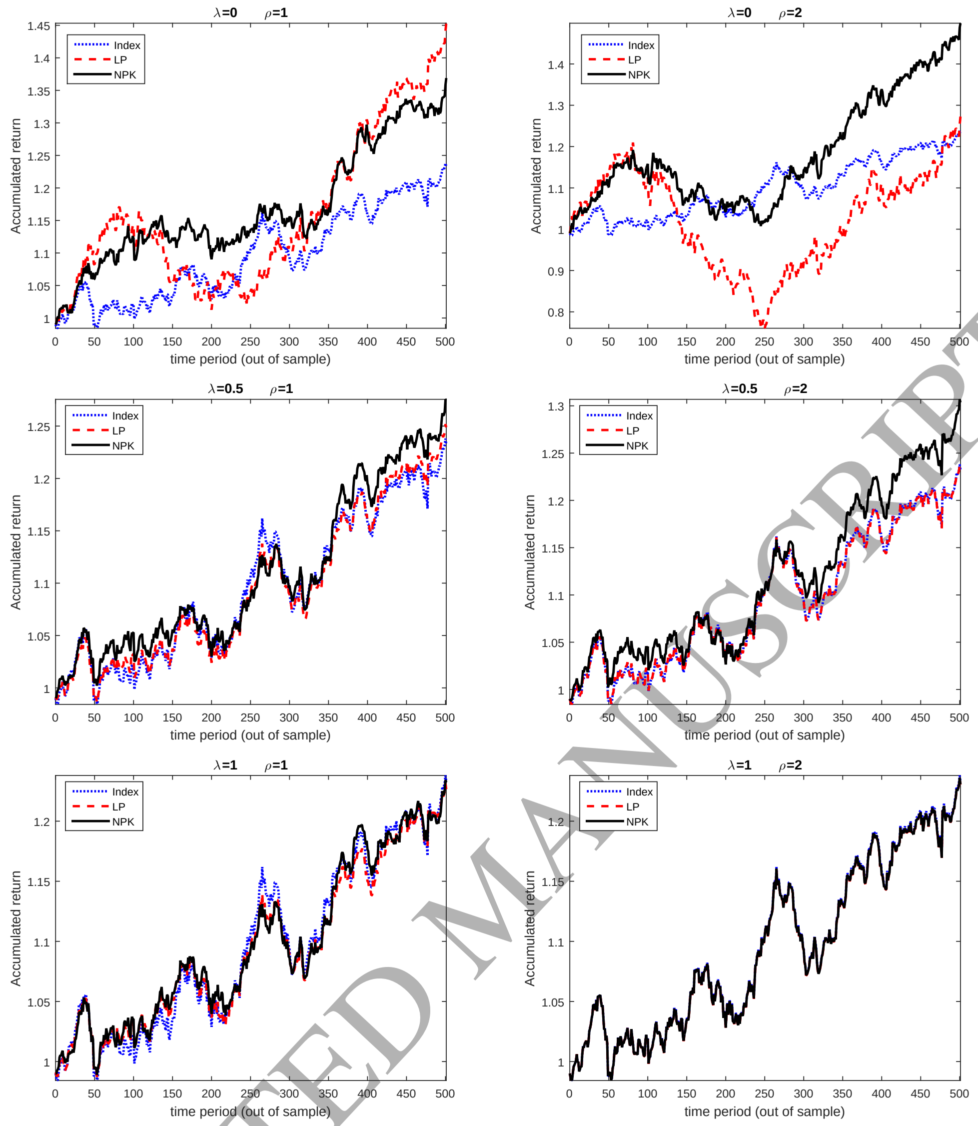


Figure 3. Accumulated returns of index tracking models when the markets rises: out of sample analysis. In this figure, we apply the (static) optimized tracking portfolio obtained from training sample to show the accumulated performance in the out-of-sample period. We show the out-of-sample accumulated returns of AIM ($\lambda = 0$), EIM ($\lambda = 0.5$) and IRM ($\lambda = 1$) when the market rises.

5. Empirical test

5.1. Data and descriptive statistics

In Section 5, we empirically test the proposed NPK index tracking model with the CVaR constraint. We choose the S&P 500 index and the FTSE 100 index as the benchmark indices. The S&P 500 and FTSE 100 index daily price data are sourced from the

CRSP database, and constituents' daily price data are obained from Thomson Reuters Datastream. We test the model performance when it is suffering from extreme market downside risk. To best achieve our goal, we concentrate on period from 2 April 2007 to 31 December 2012, spanning the global and post-global financial crisis. Figure 4 presents the S&P 500 and FTSE 100 price time series over our sample.

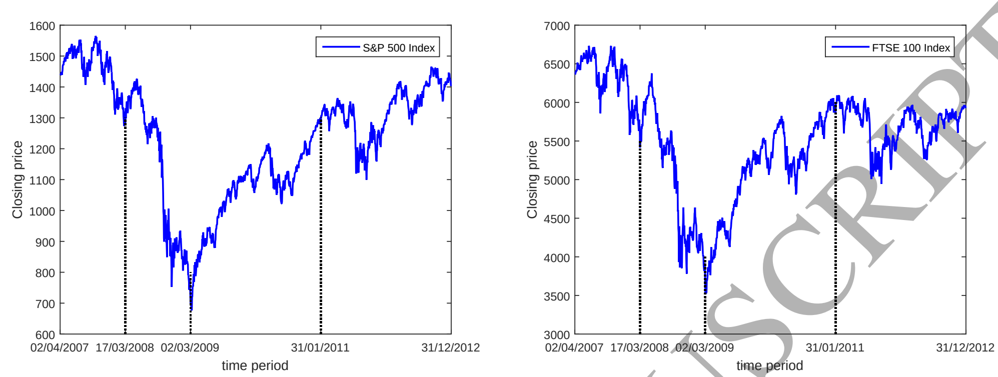


Figure 4. S&P 500 and FTSE 100 price time series from 2007 to 2012. In this figure, we plot time series of S&P 500 and FTSE 100 indices respectively. Our sample spans from 2nd April 2007 to 31st December 2012. We divide our sample to bearish market and bullish market. The bearish period is from 2nd April 2007 to 2nd March 2009, and the bullish period is from 3rd March 2009 to 31st December 2012. During the bearish period, the estimation sub-period spans from 2nd April 2007 to 17th March 2008; and the investment sub-period spans from 18th March 2008 to 2nd March 2009. During the bullish period, the estimation sub-period spans from 3rd March 2009 to 31st January 2011; and the investment sub-period spans from 1st February 2011 to 31st December 2012.

From April 2007 to March 2009, the global financial crisis impacts American and British stock markets, sending them into deep recession (we refer to this period as the bearish market). Following this, the market climbs upwards from March 2009 to December 2012 (we call this the bullish market period). Therefore, these two periods create an ideal laboratory-like market environment in which to test the performance of the index tracking models controlling CVaR. We obtain 1500 daily returns (in percentage), where 500 returns correspond to the bearish market period (2 April 2007-2 March 2009) and the other 1000 returns correspond to the bullish market period (3 March 2009-31 December 2012). Next, we split the 500 daily returns in the bearish market period into two parts. The first 250 daily returns are training samples in the estimation sub-period (2 April 2007-17 March 2008) and the other 250 returns are test samples in the investment sub-period (18 March 2008-2 March 2009). Similarly, we split 1000 daily returns in the bullish market period (3 March 2009-31 December 2012) into two parts. The first 500 daily returns are training samples in the estimation sub-period (3 March 2009-31 January 2011) and the other 500 daily returns are test samples in the investment sub-period (1 February 2011-31 December 2012).

In table 4, Panel A shows in detail the descriptive statistics for the S&P 500 and FTSE 100 index returns. Skewness and kurtosis show that index returns perform biased, high peaks and heavy tail characteristics. We point out that, from March 2008 to March 2009 (investment sub-period), the markets display huge volatility and jump risk with maximum variance (7.28) and maximum CVaR (9.41) for the S&P 500 and maximum variance (5.54) and maximum CVaR (8.73) for the FTSE 100. Panel B shows the descriptive statistics of the betas of each benchmark index constituent, without missing values in the examination periods. In a bearish market, the betas of S&P 500 constituents range from 0.068 to 2.780, whereas in a bullish market, the betas of S&P 500 constituents range from 0.280 to 2.852. For FTSE 100 constituents, betas range from 0.356 to 1.869 in a bearish market and from 0.381 to 2.142 in a bullish market.

INSERT TABLE 4 HERE

5.2. Tracking stocks selection

We need to select a subset of constituents in order to construct an optimized tracking portfolio. A popular method in the literature is to assign each constituent to a binary variable, with the intention to add cardinality constraints to construct mixed-integer programming problems. The drawback with this method is that when the number of constituents is huge, it is necessary to introduce numerous binary variables, leading to difficulties in solving this model. Worse, in addition to cardinality constraints, the NPK estimator of CVaR as a nonlinear constraint means that there is no efficient way to solve the model. In this paper, guided by Ling et al. (2014), we select a subset of constituents to construct tracking portfolios to avoid getting trapped in tedious and inefficient calculations (Canakgoz and Beasley, 2009). Specifically, we carry out an unbiased beta criterion (Ling et al., 2014), i.e., we calculate the betas of constituents in the estimation sub-periods and then choose stocks with betas close to 1. In this study, we select 100 stocks out of the S&P 500 constituents and 20 stocks out of the FTSE 100 constituents.

5.3. Measurement indicators

We assume an investor with standardized wealth 1 initially allocates zero weights on each stock at time 0, $a_{i,0} = 0$, $i = 1, 2, ..., n$. This investor imposes costs on each stock $c_i = 0.01$ and a total cost $c = 0.1$. The proportional costs for buying and selling stocks are $\delta_i^b = \delta_i^s = 0.01$. The loss probability α is 0.01. We allow short selling $l_i = -\infty$, $u_i = \infty$ and set different ρ values to test our models.

With respect to the AIM ($\lambda = 0$), we assume the tracking portfolio returns are $r_{p,t} = \mathbf{a}'\mathbf{r}_t$. We define two indicators to measure the performance of the NPK and LP models. These indicators are average excess return $\Delta R = \frac{1}{T}\sum_{t=1}^{T}(r_{p,t} - r_{I,t})$ and Sharpe ratio $SR = \frac{\Delta R}{\sqrt{Var}}$, where $Var = \frac{1}{T-1}\sum_{t=1}^{T}(r_{p,t} - \bar{r}_p)^2$, $\bar{r}_p = \frac{1}{T}\sum_{t=1}^{T} r_{p,t}$. With respect to the IRM ($\lambda = 1$), we define the tracking error $TE = \frac{1}{T}\sum_{t=1}^{T}|r_{p,t} - r_{I,t}|$ to measure the performance of the NPK and LP models. With respect to the EIM ($\lambda = 0.5$), we use $\Delta R, SR$ and TE to evaluate its performance.

We use the historical simulation method to calculate the CVaR and to measure the performance of the tracking portfolio on the downside risk control as follows: first, we sort the tracking portfolio returns $r_{p,t}$ from the lowest value to the highest. We denote the sequential estimators of $\{r_{p,t}\}_{t=1}^{T}$ by $r_{(1)} \leqslant r_{(2)} \leqslant \ldots \leqslant r_{(T)}$, then we have the sample percentile estimator of $VaR = -r_{[T\alpha]}$ and the sample weighted average estimator of $CVaR = -\frac{\sum_{t=1}^{T} r_{p,t} I(r_{p,t} \leqslant -VaR)}{\sum_{t=1}^{T} I(r_{p,t} \leqslant -VaR)}$, where $[T\alpha]$ means a maximum integer not greater than $T\alpha$. $I(\cdot)$ is an indicator function (Dowd, 2001).

5.4. Empirical results

In Section 5.4, we document the empirical results of the NPK index tracking models. For the sake of space, we only present the results from the investment sub-period.

In Table 5 we compare the NPK AIM and the LP AIM ($\lambda = 0$) in the American and British markets during investment sub-periods, where Panel A presents results for a bearish market and Panel B presents results for a bullish market. First, we find that in both bearish and bullish markets, the CVaRs from the NPK AIM are significantly less than the CVaRs from the LP AIM. This finding implies that the NPK AIM can more effectively control downside risk when compared with the LP AIM, in both bearish and bullish markets. Next, we document that the NPK AIM can obtain higher excess returns ΔR than the LP AIM in most cases. After being adjusted by standard deviations, the NPK AIM can deliver higher Sharpe ratios SR in all cases. Although in bullish markets, the LP AIM can earn greater excess returns relative to the S&P 500, the Sharpe ratios of the LP AIM are lower than those of the NPK AIM, owing to greater variance. We conclude that the NPK AIM captures higher excess returns and delivers higher Sharpe ratios with better downside risk controlling. Regarding redcomputing time, it is evident that LP AIM requires less time than NPK AIM, which is consistent with our finding in the Simulation section that LP method has advantages in solving AIM in terms of computing time. Figure 5 depicts accumulated returns trends for NPK AIM and LP

AIM against the S&P 500 and FTSE 100 indices, in both bearish and bullish markets. In the American stock market, when the S&P 500 index drops, the LP AIM generates higher accumulated returns at an early stage but the NPK AIM dominates LP AIM as the benchmark index continues to fall. When S&P 500 rises in a bullish market, LP AIM performs better than NPK AIM in terms of accumulated returns, but is subject to greater volatility. In this case too, NPK AIM delivers higher Sharpe ratios (see Table 5). In the British market, when the FTSE 100 index falls, LP AIM falls with the benchmark index but NPK AIM performs quite stably. In such a case, NPK AIM delivers higher excess returns. When FTSE 100 index rises in a bullish market, we show that NPK AIM climbs more quickly than LP AIM does and that NPK AIM yields higher excess returns.

INSERT TABLE 5 HERE

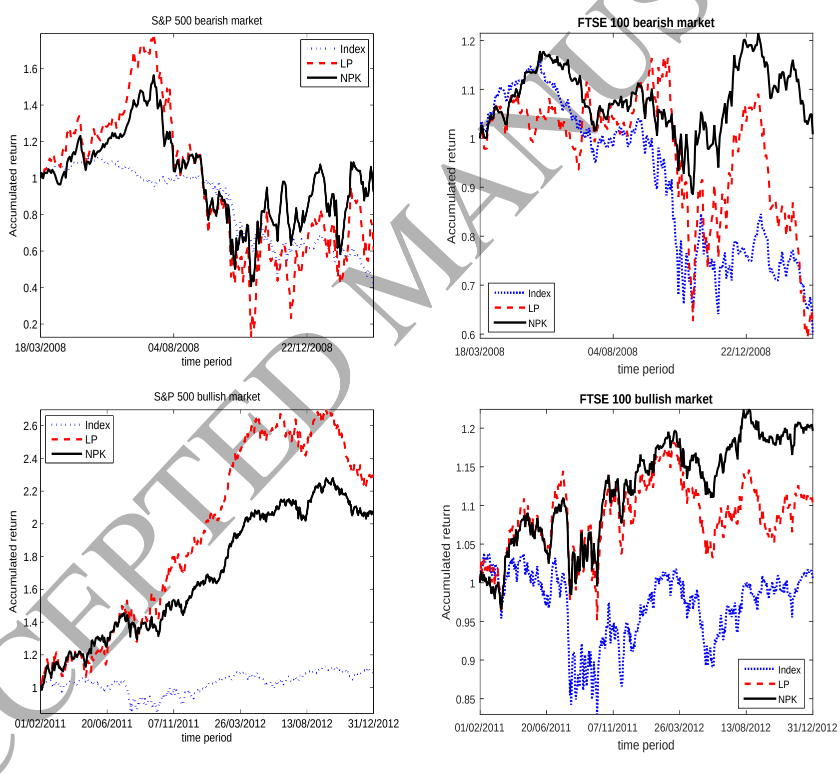


Figure 5. S&P 500 and FTSE 100 accumulated returns of the AIM ($\lambda = 0$). In this figure, we depict the accumulated returns of the NPK and the LP AIM ($\lambda = 0$). In the case of the S&P 500 in a bearish (bullish) market, we choose $\rho = 2.9$ ($\rho = 2.3$); in the case of the FTSE 100 in a bearish (bullish) market, we choose $\rho = 3$ ($\rho = 2.3$).

In Table 6 we compare the NPK IRM and LP IRM ($\lambda = 1$). We show that the CVaR values from NPK IRM are less than those from LP IRM in most cases. Therefore, NPK IRM outperforms LP IRM in terms of controlling downside risk. Furthermore, we find that when the downside risk constraint ρ increases, the tracking portfolios' CVaR values increase and the tracking errors decrease for both the NPK and LP methods. This finding means that the stricter downside risk constraint can cause greater tracking errors and when we loosen the downside risk control, the tracking portfolios will follow the benchmark index more closely. Specifically, in a bearish market, it is evident that when the downside risk constraint $\rho <= 2.9$ (case S&P 500) or $\rho <= 4.0$ (case FTSE 100), the NPK IRM generates greater tracking errors than LP IRM. The reason for this, we argue, is that the LP IRM cannot effectively control downside risk and the tracking portfolio must adhere closely to market turns. In contrast, the NPK IRM deviates from the falling benchmark index with greater tracking errors because it protects the tracking portfolio from suffering downside risk. Notice that when the downside risk constraint is loosened to some extent (e.g., $\rho >= 3.1$ in the American market), or furthermore, if we remove the downside risk constraint ($\rho = \infty$), NPK IRM produces smaller tracking errors in the American market or quite close tracking errors in the British market. In the bullish market, when the S&P 500 index rises, it is evident that tracking errors of NPK IRM are greater than those of the LP IRM, but the CVaR values are less than those for LP IRM. This might be because NPK IRM prevents the tracking portfolio from reversing with the benchmark index and therefore induces greater tracking errors. After removing the downside risk constraint ($\rho = \infty$), NPK IRM has lower tracking errors than LP IRM has. When the FTSE 100 index rises, we find similar results when the downside risk constraint $\rho <= 2.5$. However, when the downside risk constraint is loosened to $\rho >= 2.7$, LP IRM performs better in terms of controlling downside risk and tracking errors. After removing the downside risk constraint ($\rho = \infty$), the tracking errors of the two models are quite close. It is evident that LP IRM requires less time than NPK IRM, possibly because, in this empirical test, the sample size is smaller relative to a given portfolio size, such that the LP method performs better in time consumption. The computing time of the two methods are within seconds, which, in reality, would produce little difficulty.

INSERT TABLE 6 HERE

In Figure 6, we illustrate the accumulated returns of the NPK IRM and LP IRM in both the American and British markets during the investment sub-periods. This can provide a more direct impression of the downside risk controlling performance of NPK IRM and LP IRM in both bearish and bullish market environments. When the

S&P 500 index falls, both the NPK and LP IRMs' tracking portfolios drop with the benchmark index, but LP IRM falls with S&P index to a larger extent. When the FTSE 100 index crashes, we see that the NPK IRM tracking portfolio does not jump in the way the LP IRM does. This suggests that NPK IRM has larger tracking errors but yields greater accumulated returns, because NPK IRM controls downside risk more effectively than LP IRM. When the S&P 500 or FTSE 100 increases, both the NPK IRM and LP IRM increase with the benchmark index, but NPK IRM increases to a larger extent with better accumulated returns and larger tracking errors. We notice that NPK IRM falls less than LP IRM does when either S&P 500 or FTSE 100 index reverses, due to the better downside risk control. This might be the reason why NPK IRM has larger tracking errors in the empirical test. Thus, we conclude that NPK IRM outperforms LP IRM in terms of controlling downside risk and, therefore, displays larger tracking errors when the downside risk constraint is strict.

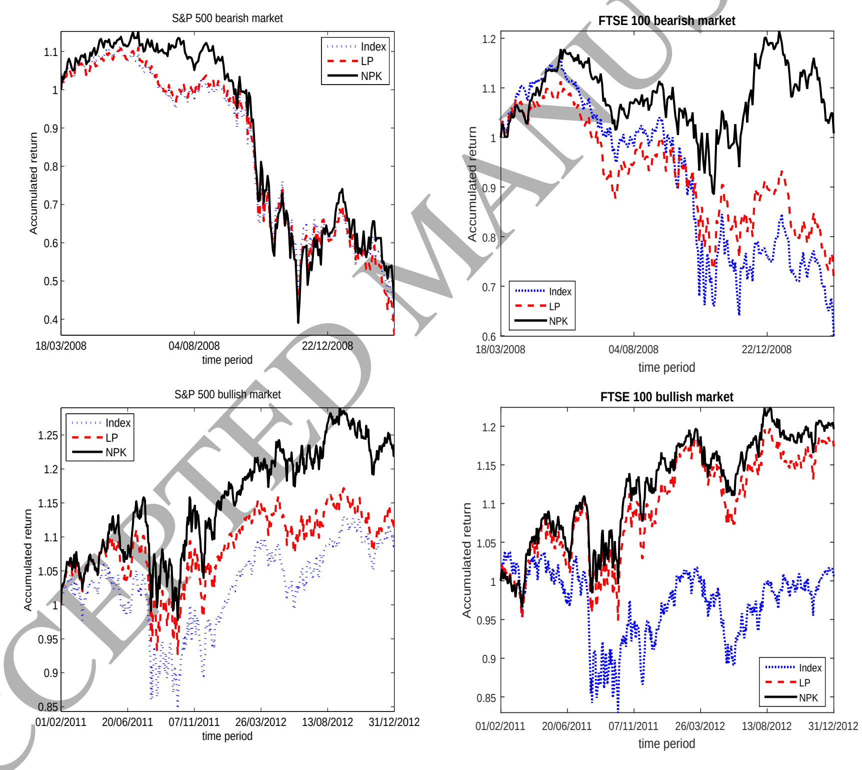
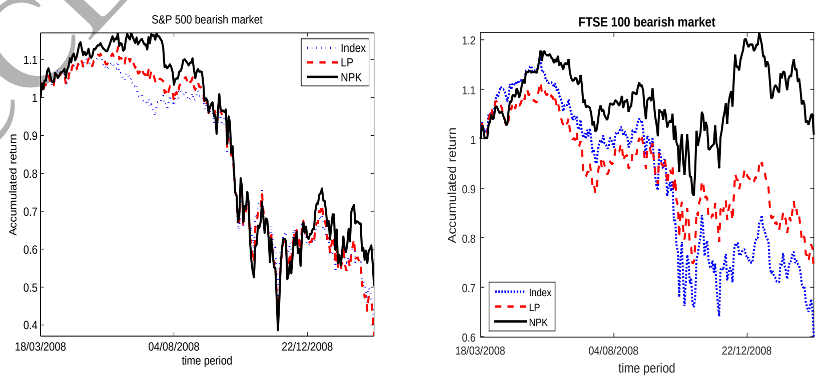


Figure 6. S&P 500 and FTSE 100 accumulated returns of the IRM ($\lambda = 1$). In this figure, we depict the accumulated returns of the NPK and the LP IRM ($\lambda = 1$). In the case of the S&P 500 in a bearish (bullish) market, we choose $\rho = 2.3$ ($\rho = 2.3$); in the case of the FTSE 100 in a bearish (bullish) market, we choose $\rho = 3$ ($\rho = 2.3$).

Table 7 presents empirical results for EIM. Consistent with results in Tables 5 and 6, NPK EIM controls downside risk more effectively than LP EIM does. The CVaR values of NPK EIM are lower than those of LP EIM for the S&P 500 and FTSE 100 indices, in both bearish and bullish markets. EIM aims to capture greater excess returns when tracking the benchmark index. Similar to the results in Table 5, we show that NPK EIM outperforms LP EIM in terms of excess returns. After being adjusted to standard deviations, NPK EIM also performs better than LP EIM in Sharpe ratios. These results are consistent across both American and British markets and in both bearish and bullish environments. With respect to tracking errors, we find that, like the results in Table 6, when the downside risk constraint is strict, the NPK EIM generates greater tracking errors and prevents the tracking portfolio from suffering downside risk. After removing the downside risk constraint ($\rho = \infty$), we find the tracking errors of NPK EIM to be less than or close to those of LP EIM. With regard to the computing time, it is evident that LP IRM requires less time than NPK IRM, possibly because the sample size in this empirical test is small relative to the portfolio size, such that the LP method outperforms NPK in time consumption. The computing time of the two methods is less than 8 seconds, which should provide little difficulty in reality. Figure 7 presents the accumulated returns of the NPK EIM and LP EIM in the American and British markets, during both bullish and bearish investment sub-periods. The results are similar to those in Figure 6. LP EIM falls further than NPK EIM when markets crash, because the LP method cannot effectively control downside risk like NPK does. However, when markets rise, NPK EIM climbs more quickly than LP EIM does. Therefore, NPK EIM yields higher accumulated returns. We conclude that NPK EIM outperforms LP EIM in terms of controlling downside risk and obtaining excess returns.

INSERT TABLE 7 HERE



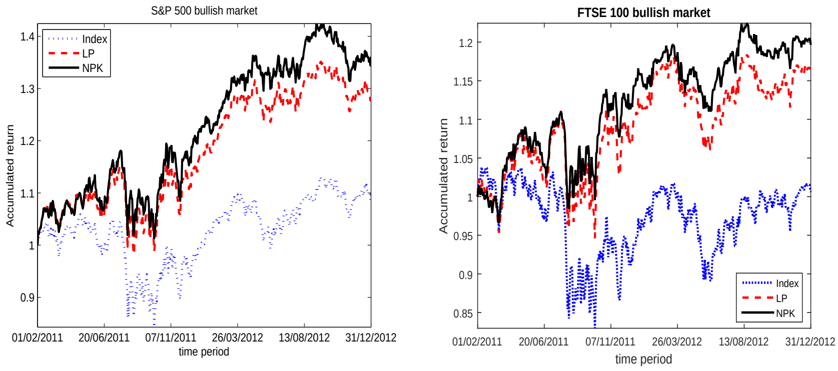


Figure 7. S&P 500 and FTSE 100 accumulated returns of the EIM ($\lambda = 0.5$). In this figure, we depict the accumulated returns of the NPK and the LP EIM ($\lambda = 0.5$). In the case of the S&P 500 in a bearish (bullish) market, we choose $\rho = 2.3$ ($\rho = 2.3$); in the case of the FTSE 100 in a bearish (bullish) market, we choose $\rho = 3$ ($\rho = 2.3$).

6. Conclusions

In this paper, we study three classes of index tracking models-the EIM, IRM and AIM-under the NPK framework, which uses a higher-order original moment to measure tracking error. In particular, these three models impose CVaR constraints to protect the tracking portfolio from market downside risks and impose other realistic constraints, such as transaction costs and investment proportion constraints. In theory, we show that the model with the CVaR constraint is a convex optimization model. Moreover, we derive NPK index tracking models with the CVaR constraint that do not rely on assumptions for asset distribution information. Compared with the LP method, our proposed NPK method has two advantages. First, NPK has smooth properties, which are helpful for optimizing index tracking models. Second, NPK mitigates some of the computational difficulties of the LP method, where the number of decision variables and constraints increases dramatically with an increase in sample size.

Numerical simulations show that the NPK method outperforms the LP method in terms of estimation accuracy. In simulated market environments, NPK displays better performance in terms of both controlling downside risk and obtaining excess returns. Regarding computational efficiency, we have mixed findings. Sample size tests show that NPK models save more time with the increase in sample size. Portfolio size tests show that, NPK EIM performs better than LP EIM; NPK IRM outperforms LP IRM with a moderate portfolio size; LP AIM requires less time than NPK AIM.

Finally, we empirically study the performance of the NPK and LP models in both the US and British stock markets, in both bullish and bearish environments. We adopt the un-biased beta method to select stocks and then use these stocks to obtain an optimized tracking portfolio. Out-of-sample tests show that NPK outperforms LP in terms of controlling downside risk and obtaining excess returns. Specifically, NPK AIM can obtain higher excess returns and Sharpe ratios; NPK IRM prevent tracking portfolios from jumping with market crashes; and NPK EIM not only controls downside risk more effectively, but also yields higher excess returns.

In this study, we provide a framework to study the NPK index tracking models with downside risk constraint. In the future, we intend to examine the effects of other downside risk measures, such as VaR, lower semi-variance, and lower partial moments, for our index tracking models. In addition, we will study the model under an expected utility framework with downside risk constraints.

References

Acerbi, C., Tasche, D., 2002. On the coherence of expected shortfall. Journal of Banking and Finance 26, 1487–1503.

Alexander, G.J., Baptista, A.M., 2004. A comparison of VaR and CVaR constraints on portfolio selection with the mean-variance model. Management Science 50, 1261–1273.

Alexander, G.J., Baptista, A.M., 2008. Active portfolio management with benchmarking: Adding a value-at-risk constraint. Journal of Economic Dynamics and Control 32, 779–820.

Alexander, G.J., Baptista, A.M., 2010. Active portfolio management with benchmarking: A frontier based on alpha. Journal of Banking and Finance 34, 2185–2197.

Anagnostopoulos, K.P., Mamanis, G., 2010. A portfolio optimization model with three objectives and discrete variables. Computers and Operations Research 37, 1285–1297.

Andreev, A., Kanto, A., 2004. Conditional value-at-risk estimation using non-integer values of degrees of freedom in student's t-distribution. The Journal of Risk 7, 55–61.

Appel, I.R., Gormley, T.A., Keim, D.B., 2016. Passive investors, not passive owners. Journal of Financial Economics 121, 111–141.

Artzner, P., Delbaen, F., Eber, J.M., Heath, D., 1999. Coherent measures of risk. Mathematical Finance 9, 203–228.

Basel Committee on Banking Supervision, 2013. Fundamental review of the trading book: A revised market risk framework. Consultative Document, Bank for International Settlements, http://www.bis.org/publ/bcbs265.pdf .

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.

Brown, D.B., Smith, J.E., 2011. Dynamic portfolio optimization with transaction costs: Heuristics and dual bounds. Management Science 57, 1752–1770.

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.

Christoffersen, S.E., Simutin, M., 2017. On the demand for high-beta stocks: Evidence from mutual funds. The Review of Financial Studies 30, 2596–2620.

Clarke, R.G., Krase, S., Statman, M., 1994. Tracking errors, regret, and tactical asset allocation. The Journal of Portfolio Management 20, 16–24.

Coleman, T.F., Li, Y., Henniger, J., 2006. Minimizing tracking error while restricting the number of assets. The Journal of Risk 8, 33–55.

Filippi, C., Guastaroba, G., Speranza, M., 2016. A heuristic framework for the bi-objective enhanced index tracking problem. Omega 65, 122–137.

Franks, E.C., 1992. Targeting excess-of-benchmark returns. The Journal of Portfolio Management 18, 6–12.

Guastaroba, G., Mansini, R., Ogryczak, W., Speranza, M.G., 2016. Linear programming models based on omega ratio for the enhanced index tracking problem. European Journal of Operational Research 251, 938–956.

Guastaroba, G., Speranza, M.G., 2012. Kernel search: An application to the index tracking problem. European Journal of Operational Research 217, 54–68.

Haugen, R.A., Baker, N.L., 1990. Dedicated stock portfolios. The Journal of Portfolio Management 16, 17–22.

Hirschberger, M., Steuer, R.E., Utz, S., Wimmer, M., Qi, Y., 2013. Computing the nondominated surface in tri-criterion portfolio selection. Operations Research 61, 169–183.

Hodges, S.D., 1976. Problems in the application of portfolio selection models. Omega 4, 699–709.

Jorion, P., 1997. Value at risk: the new benchmark for controlling market risk. McGraw-Hill.

Konno, H., Yamazaki, H., 1991. Mean-absolute deviation portfolio optimization model and its applications to tokyo stock market. Management Science 37, 519–531.

Kotz, S., Kozubowski, T., Podgorski, K., 2012. The Laplace distribution and generalizations: a revisit with applications to communications, economics, engineering, and finance. Springer Science and Business Media.

Krokhmal, P., Palmquist, J., Uryasev, S., 2002. Portfolio optimization with conditional value-at-risk objective and constraints. Journal of Risk 4, 43–68.

Landsman, Z.M., Valdez, E.A., 2003. Tail conditional expectations for elliptical distributions. North American Actuarial Journal 7, 55–71.

Larsen Jr, G.A., Resnick, B.G., 1998. Empirical insights on indexing. The Journal of Portfolio Management 25, 51–60.

Li, Q., Racine, J.S., 2007. Nonparametric econometrics: theory and practice. Princeton University Press.

Lin, Z., Bai, Z., 2010. Probability Inequalities, Springer.

Ling, A., Sun, J., Yang, X., 2014. Robust tracking error portfolio selection with worst-case downside risk measures. Journal of Economic Dynamics and Control 39, 178–207.

Markowitz, H., 1952. Portfolio selection. The Journal of Finance 7, 77–91.

Palomba, G., Riccetti, L., 2012. Portfolio frontiers with restrictions to tracking error volatility and value at risk. Journal of Banking and Finance 36, 2604–2615.

Pflug, G.C., 2000. Probabilistic constrained optimization, Springer.

Rockafellar, R.T., Uryasev, S., 2000. Optimization of conditional value-at-risk. Journal of Risk 2, 21–42.

Rockafellar, R.T., Uryasev, S., 2002. Conditional value-at-risk for general loss distributions. Journal of Banking and Finance 26, 1443–1471.

Rohweder, H.C., 1998. Implementing stock selection ideas: Does tracking error optimization do any good? The Journal of Portfolio Management 24, 49–59.

Roll, R., 1992. A mean/variance analysis of tracking error. The Journal of Portfolio Management 18, 13–22.

Roman, D., Mitra, G., Zverovich, V., 2013. Enhanced indexation based on second-order stochastic dominance. European Journal of Operational Research 228, 273–281.

Rudolf, M., Wolter, H.J., Zimmermann, H., 1999. A linear model for tracking error minimization. Journal of Banking and Finance 23, 85–103.

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

Valle, C., Meade, N., Beasley, J., 2014. Absolute return portfolios. Omega 45, 20–41.

Wang, M., Xu, C., Xu, F., Xue, H., 2012. A mixed 0–1 lp for index tracking problem with cvar risk constraints. Annals of Operations Research 196, 591–609.

Wang, M.Y., 1999. Multiple-benchmark and multiple-portfolio optimization. Financial Analysts Journal 55, 63–72.

Wu, L.C., Chou, S.C., Yang, C.C., Ong, C.S., 2007. Enhanced index investing based on goal programming. The Journal of Portfolio Management 33, 49–56.

Yao, D.D., Zhang, S., Zhou, X.Y., 2006. Tracking a financial benchmark using a few assets. Operations Research 54, 232–246.

Yao, H., Li, Y., Benson, K., 2015. A smooth non-parametric estimation framework for safety-first portfolio optimization. Quantitative Finance 15, 1865–1884.

Yao, H., Li, Z., Lai, Y., 2013. Mean–CVaR portfolio selection: A nonparametric estimation framework. Computers and Operations Research 40, 1014–1022.

Zhao, S., Lu, Q., Han, L., Liu, Y., Hu, F., 2015. A mean-CVaR-skewness portfolio optimization model based on asymmetric Laplace distribution. Annals of Operations Research 226, 727–739.

Table 1. Simulation results of the IRM

In this table, we present the simulation results of the IRM($\lambda = 1$). mse are the mean square errors of OF estimators. $\Delta\% = ((mse^{lp} - mse^{npk})/mse^{lp} \times 100\%$. $freq$ is the frequency with which the NPK model outperforms the LP model. z^* is the test statistic. AT is the average computational time (in seconds). T is the sample size. n is the portfolio size. Superscripts npk and lp refer to the NPK and LP methods, respectively. For sample size tests, we set $n = 10$, $\alpha = 0.01$, $\rho = 3$; the number of samplings is 1000, and we do not allow short selling. For portfolio size tests, we set $T = 1500$, $\alpha = 0.01$, $\rho = 3$; the number of samplings is 500, and we do not allow short selling. Panels A, B, and C show the results under normal, t, and AL distributions, respectively.

Sample size test								Portfolio size test							
Panel A. Normal distribution															
T	mse^{lp}	mse^{npk}	$\Delta\%$	$freq$	z^*	AT^{lp}	AT^{npk}	n	mse^{lp}	mse^{npk}	$\Delta\%$	$freq$	z^*	AT^{lp}	AT^{npk}
250	0.251	0.191	24.01	1.000	31.623	0.02	0.11	20	0.141	0.119	15.18	1.000	22.361	0.54	0.25
1500	0.235	0.203	13.64	1.000	31.623	0.46	0.13	100	0.214	0.188	12.36	1.000	22.361	3.43	1.61
2500	0.231	0.204	11.74	1.000	31.623	0.95	0.18	200	0.171	0.147	14.09	1.000	22.361	7.23	5.56
3500	0.230	0.206	10.55	1.000	31.623	1.62	0.24	300	0.167	0.143	14.38	1.000	22.361	12.23	15.43
5000	0.230	0.208	9.50	1.000	31.623	3.03	0.35	400	0.152	0.129	15.18	1.000	22.361	19.21	31.14
Panel B. t distribution ($m = 5$)															
T	mse^{lp}	mse^{npk}	$\Delta\%$	$freq$	z^*	AT^{lp}	AT^{npk}	n	mse^{lp}	mse^{npk}	$\Delta\%$	$freq$	z^*	AT^{lp}	AT^{npk}
250	0.011	0.005	52.31	0.703	12.839	0.02	0.13	20	0.011	0.004	64.05	1.000	22.361	0.57	0.27
1500	0.004	0.001	79.22	0.874	23.654	0.48	0.13	100	0.033	0.019	41.89	1.000	22.361	3.29	1.88
2500	0.004	0.001	82.14	0.952	28.587	1.03	0.18	200	0.036	0.022	40.62	1.000	22.361	7.27	6.57
3500	0.003	0.001	81.09	0.982	30.484	1.74	0.24	300	0.042	0.026	38.65	1.000	22.361	12.92	19.55
5000	0.003	0.001	79.30	0.997	31.433	3.23	0.35	400	0.048	0.031	36.59	1.000	22.361	20.33	39.22
Panel C. AL distribution															
T	mse^{lp}	mse^{npk}	$\Delta\%$	$freq$	z^*	AT^{lp}	AT^{npk}	n	mse^{lp}	mse^{npk}	$\Delta\%$	$freq$	z^*	AT^{lp}	AT^{npk}
250	0.022	0.006	72.07	0.953	28.650	0.02	0.12	20	0.024	0.013	46.89	1.000	22.361	0.56	0.24
1500	0.015	0.006	61.24	1.000	31.623	0.48	0.13	100	0.044	0.028	35.98	1.000	22.361	3.23	1.57
2500	0.014	0.006	55.40	1.000	31.623	1.02	0.18	200	0.048	0.032	34.17	1.000	22.361	7.36	5.24
3500	0.014	0.007	51.50	1.000	31.623	1.71	0.24	300	0.050	0.033	34.12	1.000	22.361	12.57	21.44
5000	0.014	0.007	47.23	1.000	31.623	3.19	0.34	400	0.047	0.031	34.85	1.000	22.361	18.95	38.85

Table 2. Simulation results of the AIM

In this table, we present the simulation results of the AIM($\lambda = 0$). mse are the mean square errors of OF estimators. $\Delta\% = ((mse^{lp} - mse^{npk})/mse^{lp} \times 100\%$. $freq$ is the frequency with which the NPK model outperforms the LP model. z^* is the test statistic. AT is the average computational time (in seconds). T is the sample size. n is the portfolio size. Superscripts npk and lp refer to the NPK and LP methods, respectively. For sample size tests, we set $n = 10$, $\alpha = 0.05$, $\rho = 3$; the number of samplings is 1000, and we allow short selling. For portfolio size tests, we set $T = 1500$, $\alpha = 0.05$, $\rho = 3$; the number of samplings is 500, and we do not allow short selling. Panels A, B, and C show the results under normal, t, and AL distributions, respectively.

Sample size test								Portfolio size test							
Panel A. Normal distribution															
T	mse^{lp}	mse^{npk}	$\Delta\%$	$freq$	z^*	AT^{lp}	AT^{npk}	n	mse^{lp}	mse^{npk}	$\Delta\%$	$freq$	z^*	AT^{lp}	AT^{npk}
250	0.058	0.035	39.61	0.834	21.124	0.01	0.18	20	0.003	0.003	1.57	0.866	16.368	0.39	0.35
1500	0.004	0.003	22.36	0.632	8.348	0.31	0.19	100	0.005	0.005	1.96	0.886	17.262	1.14	12.13
2500	0.002	0.002	16.87	0.609	6.894	0.72	0.25	200	0.078	0.077	0.40	0.912	18.425	2.80	29.90
3500	0.002	0.002	12.13	0.570	4.427	1.26	0.32	300	0.100	0.099	1.48	0.944	19.856	4.75	53.31
5000	0.001	0.001	8.84	0.550	3.162	2.53	0.44	400	0.116	0.115	0.89	0.936	19.499	7.35	82.43
Panel B. t distribution ($m = 5$)															
T	mse^{lp}	mse^{npk}	$\Delta\%$	$freq$	z^*	AT^{lp}	AT^{npk}	n	mse^{lp}	mse^{npk}	$\Delta\%$	$freq$	z^*	AT^{lp}	AT^{npk}
250	0.079	0.052	34.04	0.864	23.021	0.01	0.16	20	0.004	0.003	11.01	0.566	2.952	0.35	0.28
1500	0.006	0.005	16.53	0.680	11.384	0.30	0.18	100	0.004	0.004	1.28	1.000	22.361	1.14	10.13
2500	0.003	0.003	13.64	0.643	9.044	0.72	0.23	200	0.009	0.009	1.28	1.000	22.361	2.91	26.56
3500	0.002	0.002	10.95	0.599	6.261	1.28	0.30	300	0.012	0.012	0.70	1.000	22.361	5.29	50.27
5000	0.001	0.001	11.03	0.616	7.336	2.57	0.41	400	0.005	0.005	1.67	0.950	20.125	8.07	80.11
Panel C. AL distribution															
T	mse^{lp}	mse^{npk}	$\Delta\%$	$freq$	z^*	AT^{lp}	AT^{npk}	n	mse^{lp}	mse^{npk}	$\Delta\%$	$freq$	z^*	AT^{lp}	AT^{npk}
250	0.076	0.046	39.61	0.854	22.389	0.01	0.17	20	0.034	0.034	0.50	1.000	22.361	0.37	0.29
1500	0.005	0.004	20.32	0.626	7.969	0.30	0.18	100	0.050	0.050	0.60	1.000	22.361	1.12	10.09
2500	0.003	0.003	17.01	0.598	6.198	0.72	0.24	200	0.056	0.056	0.30	1.000	22.361	2.85	26.62
3500	0.002	0.002	14.89	0.602	6.451	1.26	0.30	300	0.068	0.068	0.20	1.000	22.361	5.42	51.18
5000	0.001	0.001	8.59	0.533	2.087	2.52	0.41	400	0.071	0.070	0.40	1.000	22.361	8.48	80.97

Table 3. Simulation results of the EIM

In this table, we present the simulation results of the EIM ($\lambda = 0.5$). mse are the mean square errors of OF estimators. $\Delta\% = ((mse^{lp} - mse^{npk})/mse^{lp} \times 100\%$. $freq$ is the frequency with which the NPK model outperforms the LP model. z^* is the test statistic. AT is the average computational time (in seconds). T is the sample size. n is the portfolio size. Superscripts npk and lp refer to the NPK and LP methods, respectively. For sample size tests, we set $n = 10$, $\alpha = 0.05$, $\rho = 3$; the number of samplings is 1000, and we allow short selling. For portfolio size tests, we set $T = 1500$, $\alpha = 0.05$, $\rho = 3$; the number of samplings is 500, and we allow short selling. Panels A, B, and C show the results under normal, t, and AL distributions, respectively.

Sample size test								Portfolio size test							
Panel A. Normal distribution															
T	mse^{lp}	mse^{npk}	$\Delta\%$	$freq$	z^*	AT^{lp}	AT^{npk}	n	mse^{lp}	mse^{npk}	$\Delta\%$	$freq$	z^*	AT^{lp}	AT^{npk}
250	0.038	0.026	31.32	0.999	31.560	0.02	0.10	20	0.020	0.016	20.19	1.000	22.361	0.59	0.24
1500	0.029	0.023	19.03	1.000	31.623	0.43	0.13	100	0.033	0.025	22.36	1.000	22.361	9.96	1.81
2500	0.028	0.023	16.32	1.000	31.623	0.87	0.17	200	0.072	0.061	15.04	1.000	22.361	31.95	29.68
3500	0.028	0.024	14.60	1.000	31.623	1.53	0.22	300	0.090	0.079	12.51	1.000	22.361	53.91	52.25
5000	0.028	0.024	13.04	1.000	31.623	2.91	0.30	400	0.101	0.090	10.39	1.000	22.361	82.05	82.27
Panel B. t distribution ($m = 5$)															
T	mse^{lp}	mse^{npk}	$\Delta\%$	$freq$	z^*	AT^{lp}	AT^{npk}	n	mse^{lp}	mse^{npk}	$\Delta\%$	$freq$	z^*	AT^{lp}	AT^{npk}
250	0.019	0.009	51.17	0.952	28.587	0.02	0.10	20	0.006	0.003	47.62	1.000	22.361	0.60	0.25
1500	0.009	0.005	43.85	1.000	31.623	0.43	0.12	100	0.011	0.006	46.98	1.000	22.361	9.70	2.23
2500	0.009	0.005	39.58	1.000	31.623	0.88	0.15	200	0.036	0.026	28.21	1.000	22.361	30.74	26.31
3500	0.008	0.005	36.35	1.000	31.623	1.53	0.21	300	0.045	0.034	23.37	1.000	22.361	55.45	50.32
5000	0.008	0.006	33.16	1.000	31.623	2.91	0.29	400	0.060	0.050	17.49	1.000	22.361	82.89	79.13
Panel C. AL distribution															
T	mse^{lp}	mse^{npk}	$\Delta\%$	$freq$	z^*	AT^{lp}	AT^{npk}	n	mse^{lp}	mse^{npk}	$\Delta\%$	$freq$	z^*	AT^{lp}	AT^{npk}
250	0.025	0.013	47.48	0.998	31.496	0.02	0.10	20	0.014	0.009	36.22	1.000	22.361	0.61	0.22
1500	0.017	0.011	33.64	1.000	31.623	0.42	0.11	100	0.040	0.027	32.80	1.000	22.361	10.34	1.62
2500	0.017	0.012	29.13	1.000	31.623	0.88	0.14	200	0.071	0.053	25.82	1.000	22.361	30.22	26.27
3500	0.016	0.012	26.47	1.000	31.623	1.54	0.18	300	0.097	0.076	21.94	1.000	22.361	53.82	48.98
5000	0.016	0.012	23.78	1.000	31.623	2.92	0.26	400	0.105	0.083	20.82	1.000	22.361	78.25	78.27

Table 4. S&P 500 and FTSE 100 index returns - descriptive statistics

In this table, we illustrate the divisions of market environments, estimation and investment sub-periods, and corresponding descriptive statistics from the S&P 500 and FTSE 100 indices. Panel A presents the descriptive statistics of index returns. Panel B presents the descriptive statistics of the constituents' betas. We remove the constituents with missing values in our tests.

Panel A. Descriptive statistics of index returns.

Market conditions	Sample period		Sample size	Mean	Variance	CVaR	Skewness	Kurtosis
S&P 500 Bearish market	Estimation sub-period	2 April 2007-17 March 2008	250	-0.044	1.318	3.131	-0.289	3.633
	Investment sub-period	18 March 2008-2 March 2009	250	-0.240	7.282	9.412	-0.041	5.905
S&P 500 Bullish market	Estimation sub-period	3 March 2009-31 January 2011	500	0.121	1.650	3.948	0.235	6.245
	Investment sub-period	1 February 2011-31 December 2012	500	0.021	1.395	4.922	-0.514	7.518
FTSE 100 Bearish market	Estimation sub-period	2 April 2007-17 March 2008	250	-0.062	1.773	4.911	-0.258	4.907
	Investment sub-period	18 March 2008-2 March 2009	250	-0.160	5.535	8.722	0.089	6.526
FTSE 100 Bullish market	Estimation sub-period	3 March 2009-31 January 2011	500	0.096	1.426	3.280	0.135	4.657
	Investment sub-period	1 February 2011-31 December 2012	500	0.001	1.262	3.992	-0.237	4.897

Panel B. Descriptive statistics of betas of constituents.

Market conditions	Sample period		Portfolio size	Mean	Median	Min	Max	Variance
S&P 500 Bearish market	Estimation sub-period	2 April 2007-17 March 2008	457	1.018	0.998	0.068	2.691	0.108
	Investment sub-period	18 March 2008-2 March 2009	457	1.091	1.029	0.319	2.780	0.164
S&P 500 Bullish market	Estimation sub-period	3 March 2009-31 January 2011	471	1.172	1.122	0.280	2.852	0.275
	Investment sub-period	1 February 2011-31 December 2012	471	1.099	1.079	0.322	2.213	0.125
FTSE 100 Bearish market	Estimation sub-period	2 April 2007-17 March 2008	89	1.024	0.950	0.356	1.869	0.104
	Investment sub-period	18 March 2008-2 March 2009	89	0.932	0.822	0.432	1.848	0.127
FTSE 100 Bullish market	Estimation sub-period	3 March 2009-31 January 2011	92	0.984	0.887	0.381	2.122	0.217
	Investment sub-period	1 February 2011-31 December 2012	92	1.011	0.965	0.402	2.142	0.162

Table 5. AIM ($\lambda = 0$) empirical test results

In this table, we compare the performance of the NPK AIM with the LP AIM in the investment sub-periods. Consistent with parameter settings in a numerical simulation section, we set $\alpha = 0.01, \mathbf{a}_{i,0} = 0, c_i = 0.01, c = 0.1, l_i = -\infty, u_i = \infty, \delta_i^b = \delta_i^s = 0.01, i = 1, 2, , n$. The initial wealth is standardized to be 1. ρ indicates the downside risk constraint. Panel A presents the results in a bearish market and Panel B presents the results in a bullish market.

Panel A. AIM ($\lambda = 0$) performance in the bearish market (investment sub-period)

	S&P 500								FTSE 100								
	NPK				LP					NPK				LP			
ρ	CVaR	ΔR	SR	time	CVaR	ΔR	SR	time	ρ	CVaR	ΔR	SR	time	CVaR	ΔR	SR	time
2.3	20.142	0.168	0.034	7.539	27.001	0.208	0.029	0.053	3.0	5.430	0.164	0.080	2.173	12.250	0.009	0.003	0.016
2.5	21.878	0.183	0.034	7.935	25.866	0.144	0.020	0.062	3.2	5.761	0.164	0.076	2.104	11.385	0.094	0.028	0.016
2.7	22.763	0.202	0.036	8.187	25.083	0.088	0.012	0.067	3.4	6.092	0.164	0.072	2.162	11.429	0.089	0.026	0.018
2.9	23.166	0.209	0.036	8.694	26.225	0.083	0.011	0.066	3.6	6.512	0.150	0.064	0.966	12.393	0.087	0.023	0.016
3.1	23.448	0.198	0.033	9.897	27.225	0.071	0.009	0.062	3.8	7.690	0.126	0.050	1.115	13.950	0.105	0.025	0.016
3.3	23.856	0.178	0.029	8.981	28.717	0.017	0.002	0.067	4.0	8.157	0.118	0.044	1.151	15.329	0.114	0.025	0.016

Panel B. AIM ($\lambda = 0$) performance in the bullish market (investment sub-period)

	S&P 500								FTSE 100								
	NPK				LP					NPK				LP			
ρ	CVaR	ΔR	SR	time	CVaR	ΔR	SR	time	ρ	CVaR	ΔR	SR	time	CVaR	ΔR	SR	time
2.3	5.357	0.195	0.106	6.973	7.964	0.240	0.092	0.142	2.3	3.230	0.038	0.042	2.242	4.244	0.019	0.016	0.038
2.5	5.419	0.209	0.106	7.224	8.583	0.255	0.093	0.171	2.5	3.473	0.041	0.042	2.264	4.893	0.007	0.006	0.041
2.7	5.456	0.222	0.107	6.420	9.115	0.250	0.088	0.142	2.7	3.717	0.044	0.042	2.281	6.947	-0.002	-0.001	0.043
2.9	5.762	0.227	0.105	8.016	9.267	0.261	0.088	0.160	2.9	4.236	0.041	0.035	1.414	7.710	0.006	0.003	0.042
3.1	6.143	0.234	0.103	8.837	9.289	0.278	0.090	0.145	3.1	5.139	0.030	0.022	1.387	8.389	0.014	0.007	0.044
3.3	6.577	0.239	0.101	8.336	9.795	0.295	0.093	0.155	3.3	5.958	0.030	0.019	1.917	8.743	0.019	0.009	0.043

Table 6. IRM ($\lambda = 1$) empirical test results

In this table, we compare the performance of the NPK IRM with the LP IRM in the investment sub-periods. Consistent with parameter settings in a numerical simulation section, we set $\alpha = 0.01, \mathbf{a}_{i,0} = 0, c_i = 0.01, c = 0.1, l_i = -\infty, u_i = \infty, \delta_i^b = \delta_i^s = 0.01, i = 1, 2, , n$. The initial wealth is standardized to be 1. ρ indicates the downside risk constraint, and $\rho = \infty$ means that we remove the downside risk constraint. Panel A presents the results in a bearish market and Panel B presents the results in a bullish market.

Panel A. IRM ($\lambda = 1$) performance in the bearish market (investment sub-period)

	S&P 500							FTSE 100					
	NPK			LP				NPK			LP		
ρ	CVaR	TE	time	CVaR	TE	time	ρ	CVaR	TE	time	CVaR	TE	time
2.3	9.560	0.866	6.640	9.685	0.382	0.116	3.0	5.430	0.915	2.167	6.065	0.744	0.029
2.5	9.683	0.659	6.580	10.117	0.413	0.129	3.2	5.761	0.929	2.059	6.123	0.666	0.032
2.7	9.959	0.519	7.114	10.317	0.419	0.139	3.4	6.091	0.949	2.101	6.426	0.614	0.034
2.9	10.145	0.433	8.223	10.435	0.397	0.143	3.6	6.709	0.831	0.779	6.731	0.579	0.032
3.1	10.257	0.387	7.403	10.483	0.393	0.139	3.8	6.977	0.670	0.616	6.853	0.546	0.028
3.3	10.311	0.375	8.517	10.483	0.393	0.115	4.0	7.202	0.586	0.720	7.239	0.519	0.032
∞	10.320	0.374	6.047	10.483	0.393	0.131	∞	7.748	0.500	0.544	7.552	0.500	0.027

Panel B. IRM ($\lambda = 1$) performance in the bullish market (investment sub-period)

	S&P 500							FTSE 100					
	NPK			LP				NPK			LP		
ρ	CVaR	TE	time	CVaR	TE	time	ρ	CVaR	TE	time	CVaR	TE	time
2.3	4.242	0.351	6.427	4.377	0.261	0.345	2.3	3.229	0.419	2.256	3.602	0.318	0.101
2.5	4.381	0.305	7.643	4.454	0.236	0.395	2.5	3.473	0.417	2.249	3.534	0.244	0.111
2.7	4.366	0.266	8.745	4.521	0.217	0.395	2.7	3.717	0.419	2.314	3.479	0.208	0.110
2.9	4.464	0.232	8.012	4.558	0.191	0.400	2.9	3.821	0.343	1.155	3.595	0.201	0.104
3.1	4.552	0.201	8.622	4.604	0.174	0.422	3.1	3.829	0.228	0.938	3.720	0.200	0.103
3.3	4.625	0.173	9.307	4.675	0.163	0.414	3.3	3.856	0.202	0.970	3.720	0.200	0.100
∞	4.787	0.134	6.938	4.794	0.142	0.403	∞	3.913	0.200	0.594	3.720	0.200	0.061

Table 7. EIM ($\lambda = 0.5$) empirical test results

In this table, we compare the performance of the NPK EIM with the LP EIM in the investment sub-periods. Consistent with parameter settings in a numerical simulation section, we set $\alpha = 0.01, \mathbf{a}_{i,0} = 0, c_i = 0.01, c = 0.1, l_i = -\infty, u_i = \infty, \delta_i^b = \delta_i^s = 0.01, i = 1, 2, , n$. The initial wealth is standardized to be 1. ρ indicates the downside risk constraint, and $\rho = \infty$ means that we remove the downside risk constraint. Panel A presents the results in a bearish market and Panel B presents the results in a bullish market.

Panel A. EIM ($\lambda = 0.5$) performance in the bearish market (investment sub-period)

	S&P 500										FTSE 100										
	NPK					LP						NPK					LP				
ρ	CVaR	ΔR	SR	TE	time	CVaR	ΔR	SR	TE	time	ρ	CVaR	ΔR	SR	TE	time	CVaR	ΔR	SR	TE	time
2.3	9.976	0.041	0.013	1.115	6.115	10.056	-0.012	-0.004	0.620	0.166	3.0	5.430	0.164	0.080	0.915	2.109	6.083	0.053	0.027	0.764	0.031
2.5	9.997	0.022	0.007	0.870	6.159	10.086	-0.028	-0.009	0.617	0.218	3.2	5.761	0.164	0.076	0.929	2.142	6.345	0.058	0.029	0.706	0.035
2.7	10.152	0.000	0.000	0.690	7.396	10.193	-0.046	-0.015	0.598	0.220	3.4	6.093	0.164	0.072	0.950	2.142	6.549	0.048	0.024	0.644	0.034
2.9	10.273	-0.026	-0.009	0.572	7.200	10.457	-0.057	-0.019	0.571	0.203	3.6	6.699	0.126	0.056	0.839	0.849	6.703	0.045	0.022	0.582	0.034
3.1	10.303	-0.044	-0.015	0.503	7.860	10.457	-0.057	-0.019	0.571	0.225	3.8	6.919	0.076	0.035	0.690	0.715	6.953	0.046	0.022	0.538	0.035
3.3	10.291	-0.046	-0.015	0.489	8.063	10.457	-0.057	-0.019	0.571	0.191	4.0	7.156	0.057	0.026	0.599	0.751	7.160	0.040	0.019	0.504	0.035
∞	10.291	-0.046	-0.015	0.489	7.151	10.457	-0.057	-0.019	0.571	0.191	∞	7.863	0.029	0.013	0.488	0.462	7.405	0.039	0.018	0.490	0.021

Panel B. EIM ($\lambda = 0.5$) performance in the bullish market (investment sub-period)

	S&P 500										FTSE 100										
	NPK					LP						NPK					LP				
ρ	CVaR	ΔR	SR	TE	time	CVaR	ΔR	SR	TE	time	ρ	CVaR	ΔR	SR	TE	time	CVaR	ΔR	SR	TE	time
2.3	4.345	0.052	0.042	0.417	6.107	4.464	0.038	0.033	0.313	0.481	2.3	3.229	0.038	0.042	0.419	2.309	3.704	0.030	0.029	0.320	0.102
2.5	4.367	0.043	0.035	0.358	5.514	4.501	0.033	0.030	0.284	0.684	2.5	3.473	0.041	0.042	0.417	2.242	3.682	0.033	0.032	0.253	0.121
2.7	4.380	0.038	0.031	0.310	5.523	4.549	0.028	0.025	0.254	0.529	2.7	3.717	0.044	0.042	0.490	2.234	3.700	0.029	0.027	0.215	0.136
2.9	4.499	0.033	0.028	0.269	7.477	4.605	0.027	0.024	0.236	0.534	2.9	3.870	0.041	0.037	0.343	1.139	3.752	0.027	0.025	0.205	0.116
3.1	4.595	0.030	0.025	0.235	6.765	4.652	0.029	0.025	0.231	0.538	3.1	3.954	0.033	0.029	0.234	1.266	3.976	0.026	0.023	0.209	0.111
3.3	4.693	0.027	0.023	0.207	7.062	4.708	0.026	0.022	0.208	0.591	3.3	4.015	0.028	0.024	0.208	1.130	4.070	0.025	0.022	0.212	0.106
∞	4.828	0.018	0.015	0.163	5.266	4.831	0.016	0.013	0.175	0.470	∞	4.081	0.026	0.022	0.212	0.503	4.070	0.025	0.022	0.212	0.062

Appendix

Appendix A: Proof of Theorem 2

Proof: When γ is odd, according to Eq. (7), we have

$$\begin{aligned}
&\hat{\mathrm{E}}\left[|X|^{\gamma}\right] = \int_{-\infty}^{\infty} |x|^{\gamma}\, \hat{p}(x) dx \\
&= -\int_{-\infty}^{0} x^{\gamma} \frac{1}{Tb} \sum_{t=1}^{T} g\left(\frac{x-x_t}{b}\right) dx + \int_{0}^{\infty} x^{\gamma} \frac{1}{Tb} \sum_{t=1}^{T} g\left(\frac{x-x_t}{b}\right) dx \\
&= -\frac{1}{T} \sum_{t=1}^{T} \int_{-\infty}^{-\frac{x_t}{b}} (by+x_t)^{\gamma} g(y)\, dy + \frac{1}{T} \sum_{t=1}^{T} \int_{-\frac{x_t}{b}}^{\infty} (by+x_t)^{\gamma} g(y)\, dy \\
&= -\frac{1}{T} \sum_{t=1}^{T} \int_{-\infty}^{-\frac{x_t}{b}} \left(\sum_{i=0}^{\gamma} C_{\gamma}^{i} x_t^{\gamma-i} b^i y^i\right) g(y)\, dy + \frac{1}{T} \sum_{t=1}^{T} \int_{-\frac{x_t}{b}}^{\infty} \left(\sum_{i=0}^{\gamma} C_{\gamma}^{i} x_t^{\gamma-i} b^i y^i\right) g(y)\, dy \\
&= -\frac{1}{T} \sum_{t=1}^{T} \left(\sum_{i=0}^{\gamma} C_{\gamma}^{i} x_t^{\gamma-i} b^i \int_{-\infty}^{-\frac{x_t}{b}} y^i g(y)\, dy\right) + \frac{1}{T} \sum_{t=1}^{T} \left(\sum_{i=0}^{\gamma} C_{\gamma}^{i} x_t^{\gamma-i} b^i (-1)^i \int_{-\infty}^{\frac{x_t}{b}} y^i g(y)\, dy\right) \\
&= \frac{1}{T} \sum_{t=1}^{T} \left(\sum_{i=0}^{\gamma} C_{\gamma}^{i} x_t^{\gamma-i} b^i \left(-F_i(-\tfrac{x_t}{b}) + (-1)^i F_i(\tfrac{x_t}{b})\right)\right),
\end{aligned}$$

where $F_i(z) = \int_{-\infty}^{z} y^i g(y)\, dy$. When γ is even, according to Eq. (7), we have

$$\begin{aligned}
\hat{\mathrm{E}}\left[|X|^{\gamma}\right] &= \int_{-\infty}^{\infty} |x|^{\gamma}\, \hat{p}(x) dx = \int_{-\infty}^{\infty} x^{\gamma} \hat{p}(x) dx = \int_{-\infty}^{\infty} x^{\gamma} \frac{1}{Tb} \sum_{t=1}^{T} g\left(\frac{x-x_t}{b}\right) dx \\
&= \frac{1}{Tb} \sum_{t=1}^{T} \int_{-\infty}^{\infty} x^{\gamma} g\left(\frac{x-x_t}{b}\right) dx = \frac{1}{T} \sum_{t=1}^{T} \int_{-\infty}^{\infty} (by+x_t)^{\gamma} g(y)\, dy \\
&= \frac{1}{T} \sum_{t=1}^{T} \int_{-\infty}^{\infty} \left(\sum_{i=0}^{\gamma} C_{\gamma}^{i} x_t^{\gamma-i} b^i y^i\right) g(y)\, dy = \frac{1}{T} \sum_{t=1}^{T} \left(\sum_{i=0}^{\gamma} C_{\gamma}^{i} x_t^{\gamma-i} b^i M_i\right),
\end{aligned}$$

where $M_i = \int_{-\infty}^{\infty} y^i g(y)\, dy = \lim_{z\to\infty} F_i(z)$.

This completes the proof.

Appendix B: Proof of Theorem 3

Proof: $g(z) = \varphi(z) \triangleq \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}z^2}$ is the Gauss kernel function, by the formula of integration by parts, as follows:

$$\begin{aligned}
F_i(z) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} y^i \mathrm{e}^{-\frac{1}{2}y^2} dy = -\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} y^{i-1} d\mathrm{e}^{-\frac{1}{2}y^2} \\
&= -\frac{1}{\sqrt{2\pi}} z^{i-1} \mathrm{e}^{-\frac{1}{2}z^2} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \mathrm{e}^{-\frac{1}{2}y^2} dy^{i-1} \\
&= -\frac{1}{\sqrt{2\pi}} z^{i-1} \mathrm{e}^{-\frac{1}{2}z^2} + (i-1) \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} y^{i-2} \mathrm{e}^{-\frac{1}{2}y^2} dy \\
&= -z^{i-1} \varphi(z) + (i-1) F_{i-2}(z).
\end{aligned}$$

This completes the proof.

Appendix C: Simulation sample generation procedures

C.1. Normal distribution

Generating the d-dimension normal distribution $N_d(\mu, \Sigma)$ samples:

(a) Decompose Σ via the Cholesky factorization to obtain a lower triangle matrix A such that $\Sigma = AA'$;

(b) Generate multivariate standard normal distribution $N_d(\mathbf{0}, I_d)$ variate $\mathbf{X}$, I_d is d-dimension identity matrix;

(c) Set $\mathbf{Y} = A\mathbf{X} + \mu$;

(d) Return $\mathbf{Y}$.

C.2. t-distribution

Generating the d-dimension t-distribution $t(\mu, \Sigma, m)$ samples:

(a) Generate a chi-square distribution random variate Z with degrees of freedom m;

(b) Independently of Z , generate multivariate normal $N_d(\mathbf{0}, \Sigma)$ variate $\mathbf{X}$;

(c) Set $\mathbf{Y} = \mathbf{X}\sqrt{\frac{m}{Z}} + \mu$;

(d) Return $\mathbf{Y}$.

C.3. AL distribution

In order to generate the sample from AL distribution, we first introduce Lemma 6.

Lemma 6 *(Kotz et al. 2012): Let* $\mathbf{Y} = (Y_1, Y_2, ..., Y_d)' \sim AL_d(\mu, \Sigma)$ *, then there exists a random vector* $\mathbf{X} \sim N_d(\mathbf{0}, \Sigma)$ *, and an exponentially distributed random variable* Z *with mean 1, independent of* $\mathbf{X}$*, such that* $\mathbf{Y} = \mu Z + \sqrt{Z}\mathbf{X}$.

Generating the d-dimension asymmetric Laplace distribution $AL_d(\mu, \Sigma)$ samples:

(a) Generate a standard exponential variate Z with mean 1;

(b) Independently of Z , generate multivariate normal $N_d(\mathbf{0}, \Sigma)$ variate $\mathbf{X}$;

(c) Set $\mathbf{Y} = \mu Z + \sqrt{Z}\mathbf{X}$;

(d) Return $\mathbf{Y}$.