Quantitative Finance, 2020
Vol. 20, No. 9, 1513–1530, https://doi.org/10.1080/14697688.2020.1737328

High-dimensional index tracking based on the adaptive elastic net

LIANJIE SHU *†, FANGQUAN SHI† and GUOLIANG TIAN‡

†Faculty of Business, University of Macau, Macau, People's Republic of China
‡Department of Statistics and Data Science, Southern University of Science and Technology, ShenZhen, People's Republic of China

(*Received 1 January 2019; accepted 24 February 2020; published online 20 April 2020*)

When a portfolio consists of a large number of assets, it generally incorporates too many small and illiquid positions and needs a large amount of rebalancing, which can involve large transaction costs. For financial index tracking, it is desirable to avoid such atomized, unstable portfolios, which are difficult to realize and manage. A natural way of achieving this goal is to build a tracking portfolio that is sparse with only a small number of assets in practice. The cardinality constraint approach, by directly restricting the number of assets held in the tracking portfolio, is a natural idea. However, it requires the pre-specification of the maximum number of assets selected, which is rarely practicable. Moreover, the cardinality constrained optimization problem is shown to be NP-hard. Solving such a problem will be computationally expensive, especially in high-dimensional settings. Motivated by this, this paper employs a regularization approach based on the adaptive elastic-net (Aenet) model for high-dimensional index tracking. The proposed method represents a family of convex regularization methods, which nests the traditional Lasso, adaptive Lasso (Alasso), and elastic-net (Enet) as special cases. To make the formulation more practical and general, we also take the full investment condition and turnover restrictions (or transaction costs) into account. An efficient algorithm based on coordinate descent with closed-form updates is derived to tackle the resulting optimization problem. Empirical results show that the proposed method is computationally efficient and has competitive out-of-sample performance, especially in high-dimensional settings.

Keywords: Index tracking; Sparsity; Cardinality; Lasso

1. Introduction

Index tracking aims at determining an optimal portfolio that replicates the performance of a target index (or benchmark). There are two ways to do this: full replication and partial replication. In full replication, all constituents of the target index are held in the same proportions in the tracking portfolio. However, too many positions in the tracking portfolio often result in high administrative and transaction costs. It is usually inconvenient and impractical to maintain a full replication. For this reason, fund managers may want to use only a subset of the index constituents to accurately track the index level. Therefore, a good partial index tracking model should be able to determine which stocks should be selected in the tracking portfolio (Montfort *et al.* 2008).

A benchmark index can be tracked on the basis of both cointegration and correlation. The cointegration-based approach aims at finding long-term relationships between the subset of constituents and the index, while the correlation-based approach aims at minimizing the tracking error. Different measures of the tracking error have been defined. In the majority of studies, the tracking error is defined as the variance of the difference between tracking portfolio return and index return, while it is sometimes defined as the mean absolute deviation of the tracking portfolio return from the index (Rudolf *et al.* 1999, Consiglio and Zenios 2001). Both approaches have been widely studied. The early work on the comparison of both approaches can be found in Alexander and Dimitriu (2005). Some recent comparisons have been made by Acosta-González *et al.* (2015) and Sant'Anna *et al.* (2017). This paper will focus on the correlation-based approach for sparse index tracking.

Depending on whether the model can simultaneously determine the subset of stocks and the optimal portfolio weights, partial index tracking models based on correlation can be grouped into two categories: a sequential approach and a unified approach. In the first category, the task is divided into two steps, namely, step (i): selection of the subset of assets, and step (ii): capital allocation, i.e. distribution of the available capital among the selected assets. A sample of research in

*Corresponding author. Email: ljshu@umac.mo, ljshu@um.edu.mo

this category includes Jansen and Van Dijk (2002) and Montfort *et al.* (2008). They used some heuristic approaches based on stock market capitalization and correlations between the stocks and the index to select a small number of stocks. Then the optimal weights are obtained by minimizing the tracking error under additional restrictions on transaction costs. Note that a sequential approach is employed in this category to optimize the capital allocation among the selected assets. Therefore, it is not clear how optimal is the resulting tracking portfolio (Benidis *et al.* 2018).

The second category aims at unifying the above two steps, i.e. to select the subset of assets and to determine the optimal weights of the selected assets simultaneously. A typical model in this category consists of minimizing a given tracking error measure while imposing the cardinality constraint, i.e. limiting the maximum number of assets held in the portfolio. However, the cardinality constrained index tracking problem is shown to be NP-hard in Ruiz-Torrubiano and Suárez (2009). This presents computational challenges, especially in high-dimensional settings, due to the non-differentiability and non-convexity of the search space.

Many approaches have been proposed to deal with the cardinality constraint. One common approach is to transform the above optimization problem with the cardinality constraint into a mixed-integer programing problem (Takeda *et al.* 2013). When the problem size is small, it can be solved using standard commercial solvers in a reasonable time (Scozzari *et al.* 2013). However, when the problem size becomes large, computing the exact solution to such a problem could take a very long time or even be impossible. It is usually hard for the standard solvers to find an optimal solution within practical time limits when the dimension is relatively large.

In order to speed up the solution process with the cardinality constraint, some heuristic algorithms have been proposed. For example, Gilli and Këllezi (2002) proposed using threshold accepting heuristic for this problem. Maringer and Oyewumi (2007) used differential evolution heuristics to solve the cardinality constrained index tracking problem. Krink *et al.* (2009) proposed a hybrid algorithm based on differential evolution and combinatorial search for index tracking, denoting as DECS-IT. Scozzari *et al.* (2013) employed the DECS-IT algorithm to solve the index tracking problem with the cardinality constraint and European Union Directive UCITS (Undertaking for Collective Investments in Transferable Securities) rules. The UCITS rules are additional regulatory requirements for a collective fund, which require that the sum of all asset weights exceeding 5% must be smaller than 40%. Recently, both Guastaroba and Speranza (2012) and Takeda *et al.* (2013) formulated a mixed-integer problem. They respectively proposed kernel search and a greedy algorithm to speed up the solution procedure. Also, many researchers have proposed genetic algorithms. A sample of research in this category includes Beasley *et al.* (2003), Shapcott (1992), Chang *et al.* (2000), Oh *et al.* (2005), Ruiz-Torrubiano and Suárez (2009), and Strub and Trautmann (2019). However, these heuristic search algorithms are not able to guarantee optimality of the solution. In general, they only find an approximate solution to the optimization problem.

Some other works involve the use of certain functions to approximate the non-convex cardinality constraint in order to simplify the solution process. For example, Fastrich *et al.* (2014) considered the use of the q-norm ($q = 1/2$) constraint as a variation on the cardinality constraint. Benidis *et al.* (2018) used a logarithm-based function to approximate the cardinality constraint and derived a fast and efficient algorithm for high-dimensional index tracking.

In addition to the use of cardinality constraints to restrict the maximum number of assets in the portfolio, another feasible solution consists of using a regularization model for obtaining optimal sparse portfolios which originates from high-dimensional regression analysis. Owing to their good property for generating sparse solutions, some Lasso-based regularization methods have been recently proposed to tackle the sparse index tracking problem. For example, Wu *et al.* (2014) proposed using Lasso for index tracking with no-short selling constraints. Analogously, Wu and Yang (2014) and Yang and Wu (2016) respectively considered the nonnegative elastic-net (Enet) and nonnegative adaptive Lasso (Alasso) for sparse index tracking. However, there is one important limitation associated with these studies. They are discussed under no-short selling constraints only, and no other practical constraints are considered. Therefore, the above formulations have limited practical applicability. The classical full investment condition, which restricts the sum of weights of the selected assets to be exactly one, is simply ignored in these studies. By omitting the full investment condition, the traditional Lasso method can easily achieve different levels of sparsity by tuning the regularization parameter. However, with full investment and no-short selling constraints considered together, the Lasso method is ineffective in promoting sparsity, even if the regularization parameter is very large. This is because the Lasso penalty, ℓ_1 norm on the portfolio weights, equals a constant value of one under these two constraints. Therefore, the Lasso penalty cannot shrink many of the portfolio weights towards zero, regardless of the tuning parameter. This will be shown later in this paper, which has been also mentioned in Fastrich *et al.* (2014) and Benidis *et al.* (2018).

In order to widen the applications of Lasso-based methods for sparse index tracking, this paper proposes a general regularization approach based on the adaptive elastic-net (Aenet) model. The proposed method represents an important class of convex regularization methods, which includes the traditional Lasso, adaptive Lasso, and elastic-net methods as special cases. Moreover, to make the formulation more practical and flexible, we will take the full investment condition into account as well as transaction costs. In a real-world setup, managers have to take into account transaction, administrative, and other costs. Although the standard solvers can be used to solve the resulting optimization problem, they may not be computationally efficient, especially in high-dimensional settings. Yen and Yen (2014) showed that the coordinate descent algorithm is much faster than the standard solvers when the solution vector in high-dimensional settings is expected to be sparse. For this reason, we develop a coordinate descent algorithm to solve the optimization problem for high-dimensional sparse index tracking.

2. The Aenet for index tracking

In this section, we first briefly review the traditional index tracking with a cardinality constraint. Then the regularization method based on Aenet is reviewed. Next, we extend the Aenet model to sparse index tracking. Finally, the effects of regularization parameters are discussed.

2.1. Index tracking with the cardinality constraint

One of the most commonly used tracking error measures is the mean squared error. Under the full investment and no-short selling constraints, the index tracking problem is defined as

$$\min_{w} \quad \frac{1}{n}\|y - Xw\|_2^2, \tag{1}$$

$$\text{s.t.} \quad \sum_{j=1}^{p} w_j = 1, \tag{2}$$

$$w_j \geq 0, \tag{3}$$

where $y = [y_1, \ldots, y_n]'$ is the $n \times 1$ vector of index returns, $X = (x_{ij})_{n\times p}$ is the $n \times p$ matrix of returns on the p index constituents in n time periods, and $w = [w_{1,\ldots}, w_p]'$ is the $p \times 1$ weight vector to be determined for minimizing the tracking error.

Sparse index tracking aims at reproducing the performance of a given index with a smaller number of constituents. One of the most common approaches to achieve this goal is to impose a cardinality constraint on the tracking portfolio (Fastrich *et al.* 2014)

$$\min_{w} \quad \frac{1}{n}\|y - Xw\|_2^2, \tag{4}$$

$$\text{s.t.} \quad \sum_{j\in\Im} w_j = 1, \tag{5}$$

$$w_j \geq 0, \tag{6}$$

$$\#\Im \leq K_{\max}, \tag{7}$$

where the constraint $\#\Im \leq K_{\max}$ restricts number of active positions $\#\Im$ no more than $K_{\max}$, where $\Im = \{i \in \{1, 2, \ldots, p\} \mid w_i > 0\}$. The parameter $K_{\max}$ directly controls the sparsity of the portfolio. Smaller values of $K_{\max}$ generate sparser solutions.

When the cardinality constraint (7) is imposed, the mixed 0–1 integer programing can be applied (Canakgoz and Beasley 2009). If the number of 0–1 variables is relatively small, the exact solution to the optimization problem can be obtained by using standard optimization software (e.g. CPLEX or Gurobi). However, the standard software solutions commonly fail if the number of integer variables is very large. In this case, computing the exact solution to such a problem could take a very long time or even be impossible. Although some heuristic algorithms have been proposed for this purpose, the solution is only suboptimal. In view of the good property of obtaining sparse solution and computation efficiency, it is natural to extend the regularization method originating from high-dimensional regression analysis to sparse index tracking.

2.2. The adaptive elastic-net regression

The Lasso estimator proposed by Tibshirani (1996) is given by

$$\hat{\beta}(\text{Lasso}) = \arg\min_{\beta} \|y - X\beta\|_2^2 + \lambda\|\beta\|_1, \tag{8}$$

where $y = [y_1, \ldots, y_n]'$ is the response vector, X is the $n \times p$ centered predictor matrix, and $\|\beta\|_1 = \sum_{j=1}^{p}$ is the ℓ_1 penalty of β. The ℓ_1 penalty in Lasso can regularize the least squares loss function and shrink some coefficients to zero at the same time. The entire Lasso solution paths can be computed by the LARS algorithm (Efron *et al.* 2004). These nice properties make Lasso very popular in variable selection.

However, Lasso has some limitations. Fan and Li (2001) pointed out that the Lasso estimator has noticeably large bias and may not have the oracle property† due to the bias problem. Zou (2006) proved that Lasso could be inconsistent for variable selection unless the predictor matrix satisfies a rather strong condition. To overcome the inconsistency issue, Zou (2006) proposed the following Alasso estimator

$$\hat{\beta}(\text{Alasso}) = \arg\min_{\beta} \|y - X\beta\|_2^2 + \lambda \sum_{j=1}^{p} \hat{v}_j|\beta_j|, \tag{9}$$

where $\hat{v}_j$s are the adaptive weights. In practice, it is suggested to compute $\hat{v}_j$ by $\hat{v}_j = (|\hat{\beta}_j^{\text{init}}|)^{-\tau}$, where τ is a positive constant, and $\hat{\beta}_j^{\text{init}}$ is an initial consistent estimate of β. For simplicity, τ is popularly set as $\tau = 1$.

Moreover, the presence of collinearity can severely degrade the performance of Lasso, which is often encountered in high-dimensional data. Collinearity makes the Lasso solution paths unstable. To stabilize the solution paths and improve the prediction accuracy of Lasso, Zou and Hastie (2005) proposed the Enet estimator by introducing an extra ℓ_2 penalty, given by

$$\hat{\beta}(\text{Enet}) = \arg\min_{\beta} \|y - X\beta\|_2^2 + \lambda_1 \sum_{j=1}^{p} |\beta_j| + \lambda_2\|\beta\|_2^2. \tag{10}$$

Notice that the Alasso and Enet estimators improve Lasso in two different directions. The Alasso overcomes the inconsistency problem while the Enet improves the stability of the solution paths, compared to Lasso. It is also natural to combine the ideas of the Alasso and Enet in order to obtain an even better method. For this reason, Zou and Zhang (2009) proposed the Aenet regression based on

$$\hat{\beta}(\text{Aenet}) = \left(1 + \frac{\lambda_2}{n}\right)\left\{\arg\min_{\beta} \|y - X\beta\|_2^2 + \lambda_1 \sum_{j=1}^{p} \hat{v}_j|\beta_j| + \lambda_2\|\beta\|_2^2\right\}, \tag{11}$$

where λ_1 and λ_2 are the regularization parameters on the ℓ_1 and ℓ_2 penalties.

† The oracle property means that the underlying estimator is asymptotically equivalent to the ideal estimator obtained only with signal variables without penalization. Therefore, an oracle estimator must satisfy consistency in variable selection and parameter estimation. See Fan and Li (2001) for more detailed descriptions.

2.3. Index tracking based on the Aenet penalty

The sparse index tracking problem with the Aenet penalty can be formulated as:

$$\min_{w} \quad \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{p} x_{ij}w_j\right)^2 + \lambda_1 \sum_{j=1}^{p} \hat{v}_j|w_j| + \lambda_2\|w\|_2^2 + \lambda_c \sum_{j=1}^{p} |w_j - \overline{w}_j| \tag{12}$$

$$\text{s.t.} \quad w'e = 1, \tag{13}$$

$$w_j \geq 0, \tag{14}$$

where e is a vector of ones. Constraints (13) and (14) are the budget and no-short selling constraints, respectively. The term $\lambda_1 \sum_{j=1}^{p} \hat{v}_j|w_j|$ is the weighted ℓ_1 penalty, which controls the sparsity of the portfolio weights. In this paper, we set the adaptive weight as $\hat{v}_j = (|\hat{\beta}_j^{\text{init}}|)^{-\tau}$, where $\hat{\beta}_j^{\text{init}}$ is chosen as the solution to index tracking problems (1)–(3). The term $\lambda_2\|w\|_2^2$ is the ℓ_2 penalty, i.e. $\lambda_2\|w\|_2^2 = \lambda_2 \sum_{j=1}^{p} w_j^2$. The term $\lambda_c \sum_{j=1}^{p} |w_j - \overline{w}_j|$ is the turnover penalty, where $\overline{w}_j$ is portfolio weight of asset j in previous time period. The turnover penalty is equivalent to the turnover constraint discussed by Scozzari *et al.* (2013), which sets an upper bound on the total change in the portfolio composition between two consecutive periods. As the transaction cost is often positively related to turnover, the turnover constraint allows for a control of the transaction cost when updating the tracking portfolio in time. It is interesting to notice that the turnover penalty shares the same form as the fused lasso penalty (Tibshirani *et al.* 2005), which encourages the sparsity in the differences of $w_j - \overline{w}_j$.

The formulation (12)–(14) is very general in the sense that it can include many existing models as special cases. It is straightforward that the Aenet penalty reduces to the Lasso penalty when $\lambda_2 = 0$ and $\tau = 0$, the Alasso penalty when $\lambda_2 = 0$, and the Enet penalty when $\tau = 0$. When $\lambda_c = 0$, the turnover restriction is relaxed, implying that the transaction cost is ignored in the optimization. Compared to the models of Wu *et al.* (2014), Wu and Yang (2014), and Yang and Wu (2016), our formulation can take the additional constraints into consideration, including the full investment and turnover constraints. Therefore, our model is more general and flexible.

To help us understand the effects of different penalties and constraints on the optimal solution in index tracking, we first look at the optimal solution to the following index tracking problem with the full investment constraint only:

$$\min_{w} \quad \frac{1}{n}\|y - Xw\|_2^2, \tag{15}$$

$$\text{s.t.} \quad \sum_{j=1}^{p} w_j = 1. \tag{16}$$

Denoting $A_{p\times p} = X'X/n$ and $B = y'X/n$, the optimization problem (15)–(16) can be written as

$$\min_{w} \quad w'Aw - 2Bw, \tag{17}$$

$$\text{s.t.} \quad \sum_{j=1}^{p} w_j = 1. \tag{18}$$

It is essentially a quadratic programing problem with the optimal solution

$$w^* = A^{-1}B' + cA^{-1}e,$$

where $c = (1 - \mathrm{e}'A^{-1}B')/\mathrm{e}'A^{-1}e$, and e is a $p \times 1$ vector of ones.

The following propositions illustrate the effects of no-short selling constraints and different penalties on the optimal solution to problem (12)–(14). Details of the proofs are shown in the Appendix. Denote $\widetilde{g}$ as the sub-gradient of $\|w - \overline{w}\|_1$. When w_i is greater, smaller than $\overline{w}_i$ or equal to $\overline{w}_i$, the ith element in $\widetilde{g}$ is 1, -1 or between $[-1, 1]$, respectively. Also, let $\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_p)$ be the vector of Lagrange multipliers for the nonnegativity constraints, and λ_1, λ_2 and λ_c be the Lagrange multipliers for the adaptive lasso constraint, the ℓ_2 norm constraint, and the turnover constraint, respectively. Proposition 1 shows that the no-short selling constraint affects matrix A in the optimal solution, while vector B remains unchanged. Proposition 2 shows that the ℓ_2 penalty affects matrix A, while the ℓ_1 penalty affects vector B in the optimal solution.

PROPOSITION 1 *Effect of the no-short selling constraints: The solution to problem* (1)–(3) *is the same as the solution to problem* (17)–(18) *when* A *is replaced by* $\widetilde{A} = A - (\gamma\,\mathrm{e}' + e\gamma')/2$ *and* B *remains unchanged.*

PROPOSITION 2 *Effect of the adaptive elastic-net penalty: The solution to problem* (12)–(14) *is the same as the solution to problem* (17)–(18) *when* A *is replaced by* $\widetilde{A} = A - (\gamma\,\mathrm{e}' + e\gamma')/2 + \lambda_2 I$ *and* B *is replaced by* $\widetilde{B} = B - 1/2\lambda_c\tilde{g} - 1/2\lambda_1\hat{v}'$.

2.4. Effects of the regularization parameters

The Aenet method depends on two tuning parameters: λ_1 and λ_2. To gain some understanding of effects of regularization parameters on the solution paths, we use a small data set as an example. In particular, the price data for the Dow Jones Utility Average Index (DJU) index with 10 stocks is used. A subset of its constituents during the period from 4/1/2016 to 30/12/2016 is downloaded from Yahoo Finance. We compute the weekly log-returns on the DJU index and its components. The total sample size is 50.

Figures 1(a)–(d) plot the solution paths to problem (12)–(14) for the Lasso, Alasso, Enet, and Aenet methods with $\lambda_c = 0$, respectively. A pronounced observation from figure 1(a) is that the number of stocks selected by using Lasso is always equal to 9, regardless of the choice of λ_1. This clearly illustrates that the Lasso method is ineffective in promoting sparsity under the full investment and no-short selling constraints. Similarly, the Enet method based on the combination of the ℓ_1 and ℓ_2 penalties cannot achieve the desired sparsity, as can be seen from figure 1(c).

In contrast, both Alasso and Aenet can achieve the desired level of sparsity. As can be seen from figure 1(b), the number of stocks selected by using Alasso decreases with λ_1,

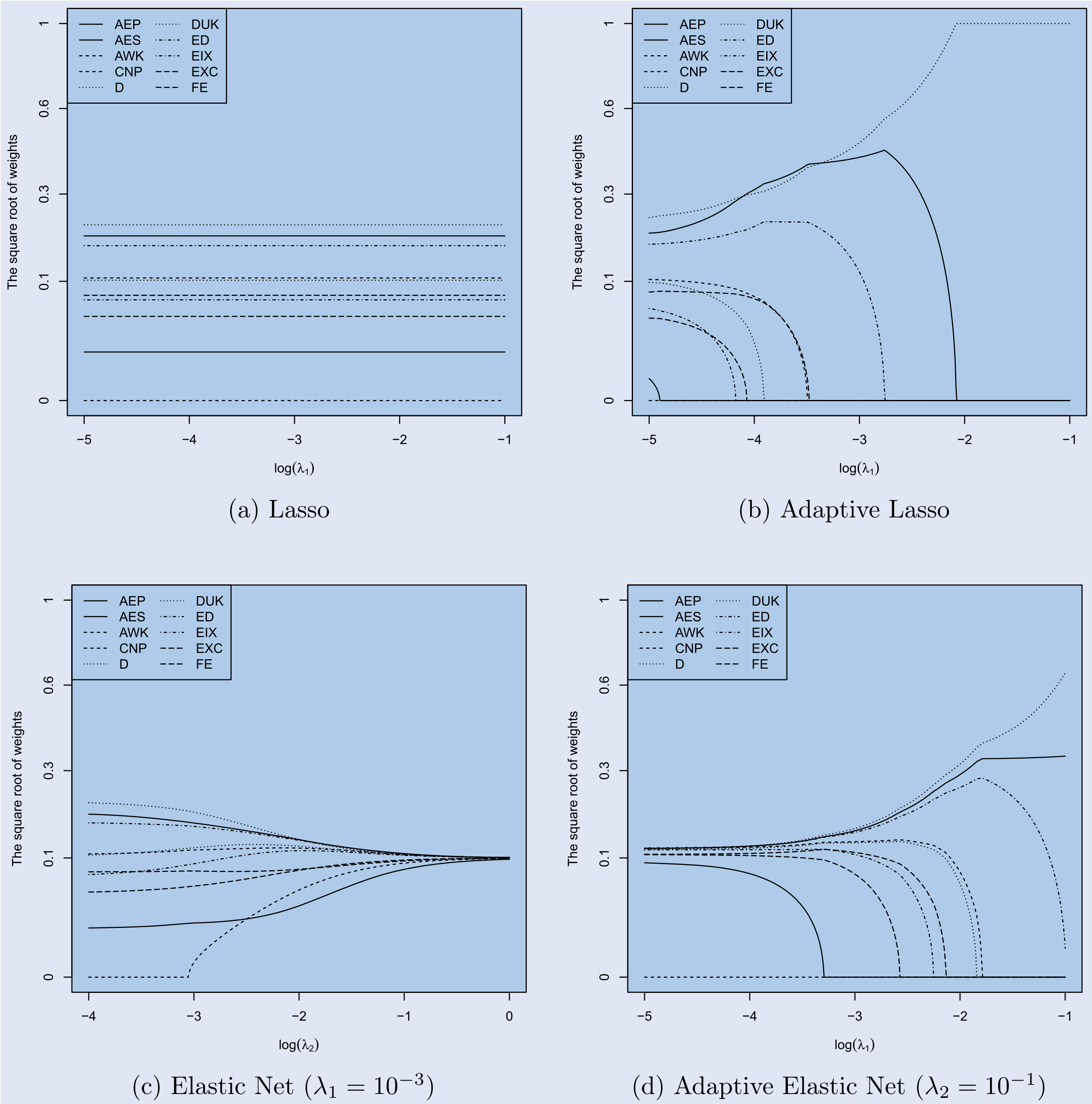


Figure 1. The solution paths for the Lasso, Alasso, Enet, and Aenet methods.

which can even reach one when $log(\lambda_1) > -2$. This observation illustrates the advantage of using the weighted ℓ_1 penalty over the standard ℓ_1 penalty to yield sparse portfolio weights in index tracking.

It is well known that the parameter λ_2 can be used to stabilize the solution paths, especially when there is multicollinearity present. By comparing figure 1(d) with figure 1(b), one can observe that the solution path of stock AEP based on Aenet has less turning points than that based on Alasso. From figure 1(b), three turning points can be observed on the solution path of AEP stock over the interval $\lambda_1 \in [10^{-4}, 10^{-2}]$. From figure 1(d), one can observe only one turning point on the solution path of AEP stock over the interval $\lambda_1 \in [10^{-4}, 10^{-1}]$. In this sense, the regularization parameter λ_2 can smooth the solution paths.

3. The coordinate descent algorithm

In this section, a coordinate descent algorithm is developed to solve the high-dimensional sparse index tracking problem (12)–(14). By rescaling the coefficient from $1/n$ to $1/2$, the sparse index tracking problem can be rewritten as

$$\min_w \quad \frac{1}{2}\sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{p} x_{ij}w_j\right)^2 + \lambda_c \sum_{j=1}^{p}|w_j - \overline{w}_j| + \lambda_1 \sum_{j=1}^{p}\hat{v}_j w_j + \lambda_2 \sum_{j=1}^{p} w_j^2,$$

$$\text{s.t.} \quad w'e = 1, \qquad w_j \geq 0. \tag{19}$$

The optimization problem (19) is essentially the same as

$$\min_w \frac{1}{2}\sum_{i=1}^n \left(y_i - \sum_{j=1}^p x_{ij}w_j\right)^2 + g_1(w) + \lambda_c \sum_{j=1}^p |w_j - \overline{w}_j| + \lambda_1 \sum_{j=1}^p \hat{v}_j w_j + \lambda_2 \sum_{j=1}^p w_j^2 + \sum_{j=1}^p g_2(w_j), \tag{20}$$

where

$$g_1(w) = \begin{cases} 0, & \text{if } w'e = 1, \\ \infty, & \text{otherwise.} \end{cases} \tag{21}$$

and

$$g_2(w_j) = \begin{cases} 0, & \text{if } w_j \geq 0, \\ \infty, & \text{otherwise.} \end{cases} \tag{22}$$

Let $f_0(w) = \frac{1}{2}\sum_{i=1}^n (y_i - \sum_{j=1}^p x_{ij}w_j)^2 + g_1(w)$ and $f_j(w_j) = \lambda_c|w_j - \overline{w}_j| + \lambda_1 \hat{v}_j w_j + \lambda_2 w_j^2 + g_2(w_j)$. The objective function in (20) can then be rewritten as

$$\min_w f(w) = \min_w \left\{ f_0(w) + \sum_{j=1}^p f_j(w_j) \right\}. \tag{23}$$

Tseng (2001) showed that a coordinate descent method can be applied to minimize $f(w)$ if some regularity conditions hold for $f_0(w)$, and if $f_j(w_j)$ is additively separable for $j = 1, 2, \ldots, p$. It can be shown that $f_0(w)$ satisfies the sufficient conditions, which are needed for applying Theorem 5.1 of Tseng (2001). Moreover, $\sum_{j=1}^p f_j(w_j)$ are convex functions of w, and $\sum_{j=1}^p f_j(w_j)$ meets the additive separability. Therefore, the global minimum of $f(w)$ can be solved by using a coordinate-wise descent algorithm.

Let γ_0 be the Lagrange multiplier associated with the full investment constraint. As defined above, let γ_i $(i = 1, \ldots, p)$ be the Lagrange multipliers associated with the no-short selling constraints. The solution strategy is to first update w for a fixed γ_0, and then to update γ_0 via the full investment constraint using the updated w. The Lagrangian corresponding to the optimization problem in (19) can be written as

$$\begin{aligned} L(w, \gamma; \lambda_1, \lambda_2) &= \frac{1}{2}\sum_{i=1}^n \left(y_i - \sum_{j=1}^p x_{ij}w_j\right)^2 + \lambda_c \sum_{j=1}^p |w_j - \overline{w}_j| + \lambda_1 \sum_{j=1}^p \hat{v}_j w_j \\ &\quad + \lambda_2 \sum_{j=1}^p w_j^2 - \gamma_0(w'e - 1) - \sum_{j=1}^p \gamma_j w_j \\ &= \frac{1}{2}\sum_{i=1}^n \left(y_i - \sum_{j=1}^p x_{ij}w_j\right)^2 + \sum_{j=1}^p \left(\lambda_c|w_j - \overline{w}_j| + \lambda_1 \hat{v}_j w_j + \lambda_2 w_j^2 - \gamma_0 w_j\right) \\ &\quad + \gamma_0 - \sum_{j=1}^p \gamma_j w_j. \end{aligned} \tag{24}$$

Let $\widetilde{y}_i^{(j)} = \sum_{k \neq j} x_{ik}\widetilde{w}_k$, $d_j = \sum_{i=1}^n x_{ij}(y_i - \widetilde{y}_i^{(j)})$, and $c_j = \sum_{i=1}^n x_{ij}^2$. The Karush-Kuhn-Tucker (KKT) conditions for (24) are

$$c_j w_j - d_j + 2\lambda_2 w_j - \gamma_0 - \gamma_j + \hat{v}_j\lambda_1 = -\lambda_c, \quad \text{if } w_j > \overline{w}_j, \tag{25}$$

$$|c_j w_j - d_j + 2\lambda_2 w_j - \gamma_0 - \gamma_j + \hat{v}_j\lambda_1| \leq \lambda_c, \quad \text{if } w_j = \overline{w}_j, \tag{26}$$

$$c_j w_j - d_j + 2\lambda_2 w_j - \gamma_0 - \gamma_j + \hat{v}_j\lambda_1 = \lambda_c, \quad \text{if } w_j < \overline{w}_j, \tag{27}$$

$$w'e = 1, \tag{28}$$

$$w_j \geq 0, \tag{29}$$

$$\gamma_i \geq 0, \tag{30}$$

$$\gamma_i w_i = 0. \tag{31}$$

The condition $\gamma_i w_i = 0$ is the complementary slackness condition. For a fixed γ_0, we can use the following formula to update each weight w_j:

$$w_j \leftarrow \begin{cases} \dfrac{d_j + \gamma_0 - \hat{v}_j\lambda_1 - \lambda_c}{c_j + 2\lambda_2}, & \text{if } \overline{w}_j < \dfrac{d_j + \gamma_0 - \hat{v}_j\lambda_1 - \lambda_c}{c_j + 2\lambda_2}, \\ \overline{w}_j, & \text{if } \dfrac{d_j + \gamma_0 - \hat{v}_j\lambda_1 - \lambda_c}{c_j + 2\lambda_2} \leq \overline{w}_j \leq \dfrac{d_j + \gamma_0 - \hat{v}_j\lambda_1 + \lambda_c}{c_j + 2\lambda_2}, \\ \dfrac{d_j + \gamma_0 - \hat{v}_j\lambda_1 + \lambda_c}{c_j + 2\lambda_2}, & \text{if } 0 \leq \dfrac{d_j + \gamma_0 - \hat{v}_j\lambda_1 + \lambda_c}{c_j + 2\lambda_2} < \overline{w}_j, \\ 0, & \text{if } \dfrac{d_j + \gamma_0 - \hat{v}_j\lambda_1 + \lambda_c}{c_j + 2\lambda_2} < 0. \end{cases} \tag{32}$$

However, the updated weights may not satisfy the full investment constraint. In order to meet this constraint, we need to update γ_0. Let $S_u = \{j : w_j > \overline{w}_j \geq 0\}$, $S_m = \{j : w_j = \overline{w}_j > 0\}$, and $S_l = \{j : 0 < w_j < \overline{w}_j\}$. Then

$$w'e = \sum_{j \in S_u} \frac{d_j + \gamma_0 - \hat{v}_j\lambda_1 - \lambda_c}{c_j + 2\lambda_2} + \sum_{j \in S_m} \overline{w}_j + \sum_{j \in S_l} \frac{d_j + \gamma_0 - \hat{v}_j\lambda_1 + \lambda_c}{c_j + 2\lambda_2} = 1. \tag{33}$$

Solving equation (33) for γ_0 leads to the following updating formula for γ_0:

$$\gamma_0 \leftarrow \left[\sum_{j \in S_u \cup S_l} \frac{1}{c_j + 2\lambda_2}\right]^{-1} \left[1 - \sum_{i \in S_m} \overline{w}_j - \sum_{j \in S_u \cup S_l} \frac{d_j - \hat{v}_j\lambda_1}{c_j + 2\lambda_2}\right.$$

$$- \sum_{j \in S_l} \frac{\lambda_c}{c_j + 2\lambda_2} + \sum_{j \in S_u} \frac{\lambda_c}{c_j + 2\lambda_2} \Bigg]. \tag{34}$$

To implement the algorithm, the initial value of each weight is set to be $w_1^{(0)} = \cdots = w_p^{(0)} = 1/p$, and the initial value of γ_0 is set to be $\gamma_0^{(0)} > \max_{1 \le j \le p} \hat{v}_j \lambda_1 + \lambda_c$. The algorithm first updates $w_1, w_2, \ldots,$ and w_p, and then uses the updated weights to update γ_0. The procedure is repeated until both w and γ_0 have converged. The procedures of the algorithm are summarized as follows.

ALGORITHM 1 Coordinate descent update for the high-dimensional tracking portfolio

(1) Fix λ_1, λ_2, λ_c, and τ at some constant levels;
(2) Compute the values of $\hat{\beta}^{\text{init}}$ and $\overline{w}$ before rebalancing;
(3) Initialize $w^{(0)} = p^{-1}e$ and $\gamma_0^{(0)} > \max_{1 \le j \le p} \hat{v}_j \lambda_1 + \lambda_c$;
(4) For $j = 1, \ldots, p$, and $m > 0$, update each weight w_j sequentially using the following form:

$$w_j^{(m)} \leftarrow \begin{cases} \dfrac{d_j^{(m)} + \gamma_0^{(m-1)} - \hat{v}_j \lambda_1 - \lambda_c}{c_j + 2\lambda_2}, & \text{if } \overline{w}_j < \dfrac{d_j^{(m)} + \gamma_0^{(m-1)} - \hat{v}_j \lambda_1 - \lambda_c}{c_j + 2\lambda_2}, \\ \overline{w}_j, & \text{if } \dfrac{d_j^{(m)} + \gamma_0^{(m-1)} - \hat{v}_j \lambda_1 - \lambda_c}{c_j + 2\lambda_2} \le \overline{w}_j \le \dfrac{d_j^{(m)} + \gamma_0^{(m-1)} - \hat{v}_j \lambda_1 + \lambda_c}{c_j + 2\lambda_2}, \\ \dfrac{d_j^{(m)} + \gamma_0^{(m-1)} - \hat{v}_j \lambda_1 + \lambda_c}{c_j + 2\lambda_2}, & \text{if } 0 < \dfrac{d_j^{(m)} + \gamma_0^{(m-1)} - \hat{v}_j \lambda_1 + \lambda_c}{c_j + 2\lambda_2} < \overline{w}_j, \\ 0, & \text{if } \dfrac{d_j^{(m)} + \gamma_0^{(m-1)} - \hat{v}_j \lambda_1 + \lambda_c}{c_j + 2\lambda_2} \le 0, \end{cases} \tag{35}$$

where $d_j^{(m)} = \sum_{i=1}^n x_{ij}(y_i - \sum_{k<j} x_{ik} w_k^{(m)} - \sum_{k>j} x_{ik} w_k^{(m-1)})$.

(5) For $m > 0$, update γ_0 using the following formula

$$\gamma_0^{(m)} \leftarrow \Bigg[\sum_{j \in S_u^{(m)} \cup S_l^{(m)}} \frac{1}{c_j + 2\lambda_2} \Bigg]^{-1} \times \Bigg[1 - \sum_{i \in S_m^{(m)}} \overline{w}_j - \sum_{j \in S_u^{(m)} \cup S_l^{(m)}} \frac{d_j - \hat{v}_j \lambda_1}{c_j + 2\lambda_2} - \sum_{j \in S_l^{(m)}} \frac{\lambda_c}{c_j + 2\lambda_2} + \sum_{j \in S_u^{(m)}} \frac{\lambda_c}{c_j + 2\lambda_2} \Bigg]. \tag{36}$$

where $S_u^{(m)} = \{j : w_j^{(m)} > \overline{w}_j \ge 0\}$, $S_m^{(m)} = \{j : w_j = \overline{w}_j > 0\}$, and $S_l^{(m)} = \{j : 0 < w_j^{(m)} < \overline{w}_j\}$.

(6) Repeat Steps (4) and (5) until both $w^{(m)}$ and $\gamma_0^{(m)}$ converge.

4. Experimental setup

4.1. Data sets and performance metrics

To evaluate the out-of-sample performance of the proposed method, we consider three stock market indices and their constituents: FTSE 100 (UK), S&P 100 (USA), and Nikkei 225 (Japan). The prices of these indices and their constituents from period 01/01/2006 to 31/12/2010 are downloaded from Yahoo Finance. The stocks that have more than 5 consecutively missing prices are removed from the data sets. Other missing prices are imputed by the linear interpolation approach. We compute the daily log-return

$$x_{t,j} = \log\left(\frac{P_{t,j}}{P_{t-1,j}}\right), \quad t = 1, \ldots, T,$$

where T is the total number of periods in a data set, and $P_{t,j}$ is the daily price of asset j in day t. Table 1 summarizes the descriptive statistics of the index returns for these data sets. The index returns exhibit typical negative skewness and fat tails.

The performance of a tracking portfolio is often assessed based on the following criteria: (i) in-sample and out-of-sample tracking error (*TE*); (ii) in-sample and out-of-sample average active return (*AR*); (iii) out-of-sample tracking portfolio turnover (*TO*); and (iv) out-of-sample correlation with the index (*corr*). The tracking error measures how closely the tracking portfolio replicates the index. The active returns are the investment returns on the tracking portfolio that exceed the returns on the underlying index, which can be positive or negative. Positive active returns show that the tracking portfolio outperforms the index, while negative active returns show that the tracking portfolio underperforms the index. Correlation measures the dependence between the tracking portfolio and the index, which ranges between -1 and 1. The index tracking is appealing if its correlation is close to one. The turnover measures the stability of the tracking portfolio. Lower turnover means lower transaction cost.

In all the experiments, a moving time window procedure was employed to determine index tracking investment strategies. In particular, a training window of size T_{train} ($T_{\text{train}} < T$) is first selected to determine the optimal tracking portfolio. Then we hold it unchanged and evaluate its performance for the subsequent T_{test} out-of-sample trading days. In the end of this testing period, we need to redesign the new tracking portfolio. For this, we move the training window forward by T_{test} days, and use the last T_{train} days to design and the subsequent T_{test} days to evaluate the new portfolio. This scheme is shown in figure 2. Following the setting of Fastrich *et al.* (2014), we choose $T_{\text{train}} = 250$ days and $T_{\text{test}} = 21$ days. All the experiments were performed on a personal computer with a 3.60 GHz processor and 16.0 GB memory.

Based on the portfolio weights determined in the first training window, w_1, the out-of-sample return r_t^{os} at time t ($t = T_{\text{train}} + 1, \ldots, T_{\text{train}} + T_{\text{test}}$) is calculated as

$$r_t^{\text{os}} = \left[\sum_{j=1}^p w_{1,j} \prod_{i=T_{\text{train}}+1}^{t} (1 + x_{i,j}) \right] /$$

Table 1. Descriptive statistics of the indices' daily log-returns (%).

	T	No. of Stocks	Mean	Standard deviation	Skewness	Kurtosis	Min	Max
FTSE 100	1258	82	0.003	1.470	– 0.090	9.983	– 9.265	9.384
S&P 100	1258	91	– 0.002	1.518	– 0.183	11.949	– 9.186	10.655
Nikkei 225	1216	198	– 0.038	1.799	– 0.377	10.577	– 12.111	13.235

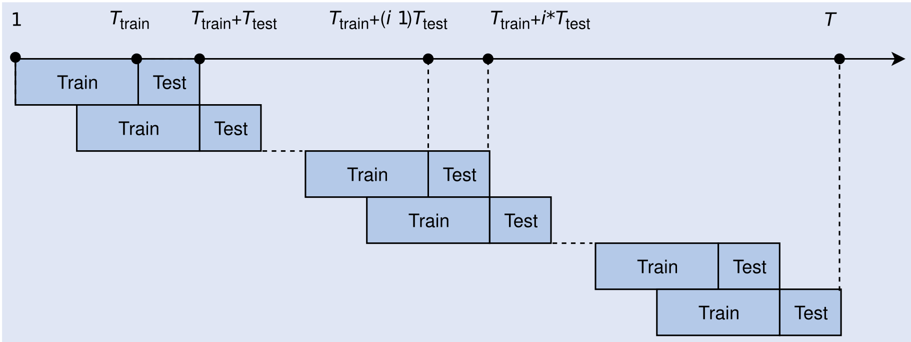


Figure 2. Illustration of the rolling training and testing windows.

$$\left[\sum_{j=1}^{p} w_{1,j} \prod_{i=T_{\text{train}}+1}^{t-1} (1+x_{i,j})\right] - 1. \tag{37}$$

The out-of-sample average active return and tracking error for the first testing window are given by

$$AR_1^{\text{os}} = \frac{1}{T_{\text{test}}} \sum_{t=T_{\text{train}}+1}^{T_{\text{train}}+T_{\text{test}}} (r_t^{\text{os}} - y_t), \tag{38}$$

and

$$TE_1^{\text{os}} = \sqrt{\frac{1}{T_{\text{test}}} \sum_{t=T_{\text{train}}+1}^{T_{\text{train}}+T_{\text{test}}} (r_t^{\text{os}} - y_t)^2}, \tag{39}$$

respectively, where. The active return and tracking error for the subsequent training and test windows can be computed in the same way.

Define $N = (T - T_{\text{train}})/T_{\text{test}}$ as the total number of rolling windows. Based on the sequence of w_i for $i = 1, \ldots, N$, the turnover is computed by

$$TO = \frac{1}{N-1} \sum_{i=1}^{N-1} \sum_{j=1}^{p} (|w_{i+1,j} - w_{i+,j}|), \tag{40}$$

where $w_{i+1,j}$ is the desired weight of asset j at the $(i+1)$th window (after rebalancing), and $w_{i+,j}$ denotes the weight of asset j at the $(i+1)$th window before rebalancing, given by

$$w_{i+,j} = \prod_{m=i\times T_{\text{train}}+1}^{i\times T_{\text{train}}+T_{\text{test}}} (1+x_{m,j}) w_{i,j} \Big/ \left(\sum_{j=1}^{p} \prod_{m=i\times T_{\text{train}}+1}^{i\times T_{\text{train}}+T_{\text{test}}} (1+x_{m,j}) w_{i,j}\right).$$

The turnover measure can be interpreted as the average percentage of wealth traded across the p available assets over the $N-1$ trading periods.

In the first training window, the in-sample returns r_t^{is} for $t = 1, \ldots, T_{\text{train}}$, can be computed as

$$r_t^{\text{is}} = \sum_{j=1}^{p} w_{1,j} x_{t,j}. \tag{41}$$

The in-sample average active return and tracking error for the first training window can be similarly computed by

$$AR_1^{\text{is}} = \frac{1}{T_{\text{train}}} \sum_{t=1}^{T_{\text{train}}} (r_t^{\text{is}} - y_t), \tag{42}$$

and

$$TE_1^{\text{is}} = \sqrt{\frac{1}{T_{\text{train}}} \sum_{t=1}^{T_{\text{train}}} (r_t^{\text{is}} - y_t)^2}, \tag{43}$$

respectively. The subsequent in-sample active returns and tracking errors are computed in a similar way.

4.2. *Descriptives of the regularization parameters*

The Aenet approach involves the selection of two regularization parameters, λ_1 and λ_2. The general way to select the regularization parameters in index tracking is minimizing the tracking error by using the K-fold cross validation. It consists of the following procedures: (1) Randomly split the indexing data (index returns y and the constituent returns X) into K approximately equal groups, with n_k observations in the kth group. The data in the kth group ($y^{(k)}$ and $X^{(k)}$) are used

Table 2. The tuning parameters of the Aenet approach selected by using cross-validation for each data set.

Data set	λ_1	λ_2
FTSE 100	$1.44 \cdot 10^{-6}$	$2.98 \cdot 10^{-3}$
S&P 100	$1.44 \cdot 10^{-6}$	$3.79 \cdot 10^{-3}$
Nikkei 225	$1.06 \cdot 10^{-5}$	$3.79 \cdot 10^{-2}$

Table 3. In-sample (IS) and Out-of-sample (OOS) tracking errors and active returns, correlation, and turnover.

	IS		OOS		IS		OOS		IS		OOS	
	Aenet	Card	Aenet	Card.	Aenet	Card	Aenet	Card	Aenet	Card	Aenet	Card
					Panel A: FTSE 100							
	13 to 28 stocks ($\bar{k}=18.88$)				26 to 42 stocks ($\bar{k}=33.33$)				41 to 53 stocks ($\bar{k}=47.46$)			
$(\lambda_1,\ \lambda_2)$	$(5.86\cdot 10^{-5},\ 1.07\cdot 10^{-3})$				$(1.22\cdot 10^{-5},\ 2.92\cdot 10^{-3})$				$(1.44\cdot 10^{-6},\ 2.98\cdot 10^{-3})$			
					Annualized tracking error in percent							
Mean	3.08	2.43	4.00	3.61	2.03	1.89	3.02	2.97	1.82	1.79	2.77	2.77
SD	0.48	0.60	1.96	1.80	0.71	0.75	1.74	1.61	0.77	0.78	1.61	1.66
t_{diff}	5.76***		1.01		0.98		0.16		0.16		−0.01	
					Annualized active return in percent							
Mean	4.47	4.41	−1.59	0.61	4.77	4.70	−0.52	−0.70	4.88	4.78	−0.87	−1.38
SD	4.14	2.98	14.76	12.18	2.57	2.47	12.22	12.09	2.19	2.32	11.43	11.47
t_{diff}	0.09		-0.8		0.14		0.07		0.22		0.22	
corr	0.989	0.994	0.980	0.984	0.996	0.997	0.989	0.990	0.997	0.997	0.991	0.991
TO	0.221	0.41			0.239	0.304			0.201	0.214		
					Panel B: S&P 100							
	14 to 31 stocks ($\bar{k}=20.65$)				24 to 42 stocks ($\bar{k}=33.00$)				53 to 69 stocks ($\bar{k}=60.67$)			
$(\lambda_1,\ \lambda_2)$	$(9.38\cdot 10^{-5},\ 5.21\cdot 10^{-3})$				$(2.66\cdot 10^{-5},\ 3.47\cdot 10^{-3})$				$(1.44\cdot 10^{-6},\ 3.79\cdot 10^{-3})$			
					Annualized tracking error in percent							
Mean	3.08	2.15	3.73	3.31	1.99	1.54	2.81	2.50	1.19	1.17	1.93	1.91
SD	0.40	0.41	1.69	1.37	0.43	0.54	1.42	1.31	0.64	0.66	1.35	1.37
t_{diff}	11.28***		1.35		4.61***		1.09		0.18		0.08	
					Annualized active return in percent							
Mean	3.18	3.99	−1.28	2.15	3.92	4.14	0.54	1.16	4.20	4.13	0.88	0.79
SD	3.34	2.41	13.54	13.35	2.48	2.10	10.21	10.14	1.82	1.86	8.07	7.83
t_{diff}	−1.37		−1.25		−0.47		−0.3		0.19		0.05	
corr	0.986	0.993	0.981	0.984	0.994	0.997	0.990	0.992	0.999	0.999	0.996	0.996
TO	0.282	0.705			0.285	0.522			0.206	0.235		
					Panel C: Nikkei 225							
	30 to 44 stocks ($\bar{k}=36.17$)				47 to 70 stocks ($\bar{k}=60.41$)				69 to 86 stocks ($\bar{k}=77.09$)			
$(\lambda_1,\ \lambda_2)$	$(8.44\cdot 10^{-5},\ 3.27\cdot 10^{-2})$				$(8.59\cdot 10^{-6},\ 6.21\cdot 10^{-3})$				$(1.06\cdot 10^{-5},\ 3.79\cdot 10^{-2})$			
					Annualized tracking error in percent							
Mean	2.83	1.91	3.70	3.44	1.59	1.42	2.89	2.79	1.40	1.28	2.63	2.65
SD	0.31	0.23	1.35	1.34	0.20	0.25	1.08	1.06	0.25	0.27	1.06	1.30
t_{diff}	16.2***		0.93		3.65***		0.43		2.37**		−0.08	
					Annualized active return in percent							
Mean	4.10	2.91	0.27	−0.56	3.07	2.59	−1.08	−2.14	2.51	2.36	−1.69	−1.87
SD	2.32	2.33	10.70	11.08	1.63	1.65	8.69	8.90	1.57	1.61	8.35	9.16
t_{diff}	2.45**		0.37		1.42		0.58		0.44		0.1	
corr	0.993	0.997	0.987	0.988	0.998	0.998	0.992	0.992	0.998	0.999	0.992	0.992
TO	0.426	0.794			0.441	0.573			0.398	0.465		

Note: This table reports results of the tracking portfolios based on the Aenet and cardinality (Card) constraint approach for each data set. Three pairs of λ_1 and λ_2 are selected to obtain tracking portfolios with different average numbers of active stocks, $\bar{k}$. The t-statistic of the differences in the tracking errors and active returns based on the two approaches is given by t_{diff}. The statistically significant differences with confidence levels 10%, 5%, and 1% are indicated by *, **, and ***, respectively. The correlation coefficient between the tracking portfolio returns and the benchmark return, and the average turnover are given by *corr* and *TO*, respectively.

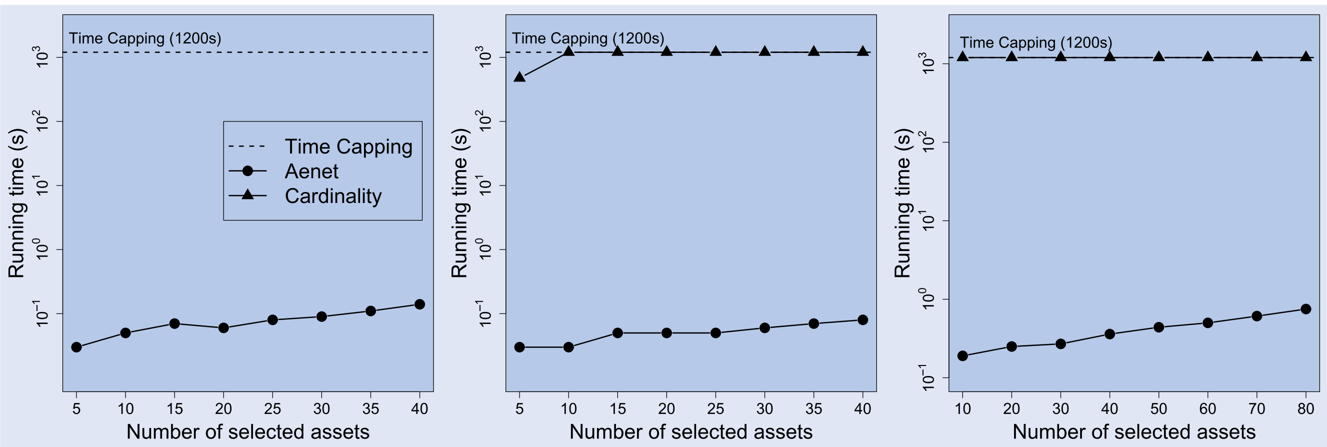
as the validation data, while the remaining $K-1$ groups of data ($y^{(-k)}$ and $X^{(-k)}$) are used as the training data; (2) Compute the tracking portfolio weights, $w^{(k)}_{(\theta)}$, for the given set of tuning parameters, $\theta=(\lambda_1,\lambda_2)$; (3) Compute the out-of-sample returns of the resulting tracking portfolio based on the validation data $r^{(k)}_{(\theta)}=X^{(-k)}w^{(k)}_{(\theta)}$; (4) Compute the out-of-sample tracking error $TE^{(k)}_{(\theta)}=\sqrt{(1/n_k)\sum_{i=1}^{n_k}(r^{(k)}_{\theta,i}-y^{(k)}_i)^2}$. The optimal regularization parameter can be obtained by minimizing the cross-validated tracking error, i.e. $\theta^*=\arg\min_\theta (1/K)\sum_{k=1}^{K} TE^{(k)}_{(\theta)}$.

The Aenet approach provides a closed-form update of the tracking portfolio weights for a given set of regularization parameters. It is very computationally efficient. Following the method of Fastrich *et al.* (2014), the regularization parameters for the Aenet approach is optimized for the first rolling window and remains constant over time. Table 2 lists the parameters selected for the Aenet approach by using the

Table 4. Descriptive statistics of daily out-of-sample returns and portfolio weights based on the Aenet and Cardinality approaches.

	Aenet	Card	Aenet	Card	Aenet	Card
	Panel A: FTSE 100					
	13 to 28 stocks ($\bar{k}$ = 18.88)		26 to 42 stocks ($\bar{k}$ = 33.33)		41 to 53 stocks ($\bar{k}$ = 47.46)	
	Descriptive statistics of daily out-of-sample portfolio returns (%)					
Skewness	0.0100	− 0.0410	− 0.0931	− 0.1208	− 0.1065	− 0.0934
Kurtosis	9.4090	8.7808	9.1916	9.2109	9.2275	9.2186
$VaR_{0.99}$	− 5.1755	− 4.9810	− 5.3173	− 5.2935	− 5.3154	− 5.3045
	Average values of maximum, median, minimum and HHI of portfolio active weights					
Max	0.1575	0.1269	0.1084	0.1117	0.0980	0.1096
Median	0.0463	0.0500	0.0225	0.0220	0.0140	0.0128
Min	0.0048	0.0219	0.0009	0.0068	0.0007	0.0019
HHI	0.0879	0.0696	0.0541	0.0500	0.0423	0.0455
	Panel B: S&P 100					
	14 to 31 stocks ($\bar{k}$ = 20.65)		24 to 42 stocks ($\bar{k}$ = 33.00)		53 to 69 stocks ($\bar{k}$ = 60.67)	
	Descriptive statistics of daily out-of-sample portfolio returns (%)					
Skewness	− 0.0535	− 0.1443	− 0.0827	− 0.1472	− 0.1337	− 0.1231
Kurtosis	9.9840	9.8361	10.0587	10.2069	10.3043	10.2986
$VaR_{0.99}$	− 5.4148	− 5.4244	− 5.4916	− 5.4701	− 5.3944	− 5.4089
	Average values of maximum, median, minimum and HHI of portfolio active weights					
Max	0.1404	0.1067	0.0955	0.0786	0.0532	0.0667
Median	0.0437	0.0467	0.0261	0.0279	0.0144	0.0132
Min	0.0041	0.0214	0.0014	0.0098	0.0009	0.0021
HHI	0.0801	0.0606	0.0488	0.0386	0.0242	0.0260
	Panel C: Nikkei 225					
	30 to 44 stocks ($\bar{k}$ = 36.17)		47 to 70 stocks ($\bar{k}$ = 60.41)		69 to 86 stocks ($\bar{k}$ = 77.09)	
	Descriptive statistics of daily out-of-sample portfolio returns (%)					
Skewness	− 0.4173	− 0.4214	− 0.4393	− 0.4419	− 0.4550	− 0.4536
Kurtosis	8.9769	9.0965	9.2533	9.2960	9.4207	9.3043
$VaR_{0.99}$	− 5.8186	− 5.7520	− 5.5854	− 5.5586	− 5.5837	− 5.5578
	Average values of maximum, median, minimum and HHI of portfolio active weights					
Max	0.0662	0.0578	0.0496	0.0480	0.0342	0.0444
Median	0.0247	0.0259	0.0148	0.0146	0.0124	0.0107
Min	0.0015	0.0110	0.0005	0.0037	0.0005	0.0019
HHI	0.0394	0.0322	0.0247	0.0216	0.0175	0.0184

Note: This table reports the skewness, the kurtosis, and the value at risk at 99% confidence level ($VaR_{0.99}$) of the out-of-sample portfolio returns.



(a) FTSE 100 (b) S&P 100 (c) Nikkei 225

Figure 3. Comparison of running time between the Aenet and cardinality constraint approaches for the first training window in each data set. (a) FTSE 100. (b) S&P 100. (c) Nikkei 225.

Table 5. In-sample (IS) and Out-of-sample (OOS) tracking errors and active returns, correlation, and turnover for the Lasso-based approaches under both the no-short selling and full investment constraints.

	Lasso		Enet		Alasso		Aenet	
	IS	OOS	IS	OOS	IS	OOS	IS	OOS
	Panel A: FTSE 100							
	$\bar{k} = 72.73$		$\bar{k} = 77.06$		$\bar{k} = 47.52$		$\bar{k} = 47.46$	
	Annualized tracking error in percent							
Mean	1.78	2.74	1.79	2.73	1.80	2.77	1.82	2.77
SD	0.78	1.63	0.78	1.62	0.77	1.63	0.77	1.61
	Annualized active return in percent							
Mean	4.74	− 1.27	4.80	− 1.29	4.70	− 0.96	4.88	− 0.87
SD	2.30	11.15	2.24	11.13	2.40	11.39	2.19	11.43
	Correlation (corr) and turnover (TO)							
corr	0.997	0.991	0.997	0.992	0.997	0.991	0.997	0.991
TO	0.189		0.181		0.212		0.201	
	Panel B: S&P 100							
	$\bar{k} = 80.23$		$\bar{k} = 82.06$		$\bar{k} = 60.73$		$\bar{k} = 60.67$	
	Annualized tracking error in percent							
Mean	1.15	1.85	1.16	1.84	1.18	1.94	1.19	1.93
SD	0.66	1.37	0.65	1.36	0.64	1.35	0.64	1.35
	Annualized active return in percent							
Mean	4.22	0.86	4.26	0.87	4.16	0.91	4.20	0.88
SD	1.85	7.69	1.84	7.68	1.88	7.96	1.82	8.07
	Correlation (corr) and turnover (TO)							
corr	0.999	0.997	0.999	0.997	0.999	0.996	0.999	0.996
TO	0.195		0.182		0.226		0.206	
	Panel C: Nikkei 225							
	$\bar{k} = 172.09$		$\bar{k} = 188.85$		$\bar{k} = 77.30$		$\bar{k} = 77.09$	
	Annualized tracking error in percent							
Mean	1.22	2.39	1.72	17.43	1.33	2.70	1.40	2.63
SD	0.27	1.09	0.22	100.59	0.23	1.06	0.25	1.06
	Annualized active return in percent							
Mean	2.25	− 1.84	1.72	59.80	2.57	− 1.32	2.51	− 1.69
SD	1.61	7.92	1.50	424.98	1.59	8.45	1.57	8.35
	Correlation (corr) and turnover (TO)							
corr	0.999	0.994	0.997	0.981	0.999	0.992	0.998	0.992
TO	0.350		0.186		0.445		0.398	

5-fold cross validation for the first training window in each data set.

5. Comparison results

In this section, we first compare the Aenet approach with the cardinality constraint approach for sparse index tracking. Then we compare the Aenet approach with the traditional Lasso-based approaches. Finally, the effects of transaction cost are discussed.

5.1. Comparison with the cardinality approach

Table 3 compares the in-sample and out-of-sample tracking errors, active returns, correlation, and turnover between the Aenet and cardinality constraint approaches for each data set. In addition to the setting of parameters listed in table 2, two more settings of (λ_1, λ_2) are also considered for the Aenet approach in order to generate tracking portfolios of different sizes. The parameters, λ_1 and λ_2, for the Aenet approach to construct a portfolio with a pre-specified number of k stocks are determined as follows. First, for each given value of λ_2, an appropriate value of λ_1 for generating a portfolio with k stocks can be obtained. One can increase λ_1 when the number of stocks selected is more than k, and decrease λ_1 when the number of stocks selected is less than k. Second, we search the appropriate λ_1 values for a given sequence of λ_2 values. Finally, we optimize the combination (λ_1, λ_2) by minimizing the 5-fold cross-validated tracking error for the first training window for a pre-specified k value. For example, for the FTSE 100 data set, the optimal parameters for the Aenet approach determined in this way are $\lambda_1 = 5.86 \times 10^{-5}$ and $\lambda_2 = 1.07 \times 10^{-3}$, aimed at generating tracking portfolios with 15 stocks in the first training window, while the combination of $\lambda_1 = 1.22 \times 10^{-5}$ and $\lambda_2 = 2.92 \times 10^{-3}$ aims

Table 6. In-sample (IS) and Out-of-sample (OOS) tracking errors and active returns, and correlation for the Lasso-based approaches under the no-short selling constraints only.

	Lasso		Alasso		Enet		Aenet	
	IS	OOS	IS	OOS	IS	OOS	IS	OOS
Panel A: FTSE 100								
	$k = 40$		$k = 40$		$k = 40$		$k = 40$	
Annualized tracking error in percent								
Mean	3.31	4.69	1.93	2.96	3.36	4.77	1.93	2.96
SD	0.83	2.91	0.77	1.79	0.78	3.00	0.77	1.78
Annualized active return in percent								
Mean	2.74	− 2.88	4.55	− 0.84	2.71	− 2.74	4.54	− 0.76
SD	3.79	15.29	2.66	12.28	3.78	15.63	2.64	12.15
Correlation (corr)								
corr	0.990	0.981	0.996	0.990	0.990	0.981	0.997	0.990
Panel B: S&P 100								
	$k = 40$		$k = 40$		$k = 40$		$k = 40$	
Annualized tracking error in percent								
Mean	3.83	5.36	1.78	2.69	3.81	5.22	1.80	2.71
SD	1.36	4.33	0.59	1.57	1.31	3.61	0.59	1.61
Annualized active return in percent								
Mean	2.76	− 0.21	3.46	− 0.01	2.67	− 0.37	3.41	0.06
SD	4.07	14.21	2.63	9.72	4.07	14.36	2.65	9.64
Correlation (corr)								
corr	0.987	0.976	0.996	0.992	0.987	0.976	0.996	0.992
Panel C: Nikkei 225								
	$k = 80$		$k = 80$		$k = 80$		$k = 80$	
Annualized tracking error in percent								
Mean	2.79	4.90	1.35	2.75	2.87	4.96	1.35	2.75
SD	0.43	3.89	0.23	1.23	0.40	3.90	0.23	1.23
Annualized active return in percent								
Mean	3.04	− 1.93	2.60	− 0.97	3.05	− 1.97	2.62	− 1.02
SD	2.66	14.63	1.68	8.14	2.68	14.43	1.67	8.20
Correlation (corr)								
corr	0.995	0.980	0.999	0.992	0.995	0.980	0.999	0.992

at generating tracking portfolios with 30 stocks. Analogously, two sets of parameters are selected for the Aenet approach in other data sets to generate tracking portfolios of different stocks. Notice that the cardinality constraint approach requires the pre-specification of the maximum number of assets selected, $k_{\max}$, which is rarely known in practice. For fair comparisons, we set $k_{\max}$ to be equal to the number of active weights found with the Aenet approach in the different windows.

As expected, the out-of-sample tracking error is larger than the in-sample tracking error for both approaches under all the data sets. Although the in-sample tracking error of the cardinality constraint approach can be significantly smaller than that of the Aenet approach, the differences in the average out-of-sample tracking error are not statistically significant. Also, the two approaches yield portfolios with similar out-of-sample active returns and correlations with the benchmark. Compared to the cardinality constraint approach, a pronounced advantage of the Aenet approach is its low turnover of the tracking portfolio, especially in the data set Nikkei 225 with a relatively large number of stocks, as can be seen from table 3.

Table 4 reports the skewness, kurtosis and value at risk (VaR) of the out-of-sample portfolio returns, and the maximum, median, minimum, and the Herfindahl–Hirschman Index (HHI) of the portfolio weights. The skewness, kurtosis, and VaR at the 99% confidence level for both approaches are similar. The HHI measures the concentration level of the tracking portfolio, computed $HHI = \sum_{i=1}^{p} w_i^2$. The larger the HHI, the more concentrated (or less diversified) the tracking portfolio. As can be seen from table 4, the Aenet approach produces higher HHI values than the cardinality approach when a small number of stocks are selected, but lower HHI values when a relatively large number of stocks are selected in the tracking portfolio.

The above comparisons show the slightly better performance of the Aenet approach in terms of low turnover and more diversification when the tracking portfolio invests in a relatively large number of stocks, compared to the cardinality constraint approach. In addition to this, the Aenet approach also enjoys the pronounced advantage of being computationally efficient. To illustrate this, figure 3 plots the running time of both approaches against the number of stocks selected for the first rolling window. To speed up the solution procedure

(a) FTSE 100

(b) S&P 100

(c) Nikkei 225

Figure 4. Wealth paths of tracking portfolios for the Lasso-based approaches under the no-short selling and full investment constraints. (a) FTSE 100. (b) S&P 100. (c) Nikkei 225.

of the cardinality constraint approach, we set the maximum running time as 1200 seconds. Clearly, the cardinality constraint approach usually reaches the time capping of 1200 seconds, while the Aenet approach only requires less than one second to determine the optimal solution. These good observations indicate the promising applications of the proposed approach in high-dimensional sparse index tracking where the computational time is a practical concern.

5.2. Comparison with the Lasso-based methods

Table 5 reports the in-sample and out-of-sample tracking errors, active returns, correlation and turnover among the existing Lasso-based approaches, including Lasso, Alasso, Enet, and Aenet. As discussed above, the Lasso is ineffective in promoting sparsity under both full investment and no-short selling constraints, independent of the regularization parameter. Not surprisingly, the tracking portfolios based on the

Figure 5. Wealth paths of tracking portfolios for the Lasso-based approaches under the no-short selling constraints only. (a) FTSE 100. (b) S&P 100. (c) Nikkei 225.

Lasso approach are not sparse. For example, the average numbers of active stocks selected in the tracking portfolio based on the Lasso approach for data sets FTSE 100, S&P 100, and Nikkei 225 are 72, 80, and 172, respectively, which are about 90% of the stocks in the benchmark index. Also, the Enet approach is ineffective in promoting sparsity.

In contrast, both the Aenet and Alasso approaches are effective in promoting sparsity, which can provide a desired level of sparsity for the tracking portfolio by tuning the regularization parameters. Although both approaches lead to the nearly equivalent tracking portfolios with negligible differences in the out-of-sample tracking error and correlation with the benchmark, Aenet has smaller turnover than Alasso. This implies that Aenet leads to a more stable tracking portfolio than Alasso.

Notice that without the full investment constraint, the Lasso approach can produce sparse solutions. The studies of Wu *et al.* (2014), Wu and Yang (2014) and Yang and

Table 7. In-sample (IS) and Out-of-sample (OOS) tracking errors and active returns, correlation, turnover, and average number of active stocks for the Aenet approach with different values of λ_c for various data sets.

	IS	OOS	IS	OOS	IS	OOS	IS	OOS	IS	OOS
					Panel A: FTSE 100					
	$\overline{\Delta} = 51.85$		$\overline{\Delta} = 43.33$		$\overline{\Delta} = 31.59$		$\overline{\Delta} = 15.5$		$\overline{\Delta} = 3.57$	
λ_c	0		$5.00 \cdot 10^{-5}$		$1.50 \cdot 10^{-4}$		$7.00 \cdot 10^{-4}$		$1.00 \cdot 10^{-1}$	
					Annualized tracking error in percent					
Mean	1.83	2.80	1.84	2.78	1.87	2.74	2.04	2.73	3.20	3.04
SD	0.77	1.61	0.77	1.61	0.75	1.57	0.73	1.50	1.71	1.49
					Annualized active return in percent					
Mean	4.93	− 0.95	4.93	− 1.00	5.03	− 0.91	5.47	− 0.52	5.98	− 2.46
SD	2.19	11.54	2.17	11.32	2.07	10.80	1.94	10.40	4.06	10.23
					Correlation (corr), turnover (TO), and average stocks ($\bar{k}$)					
corr	0.997	0.991	0.997	0.991	0.997	0.992	0.996	0.992	0.992	0.989
TO	0.202		0.153		0.109		0.054		0.014	
$\bar{k}$	47.36		47.51		47.62		45.21		36.38	
					Panel A: S&P 100					
	$\overline{\Delta} = 64.96$		$\overline{\Delta} = 50.2$		$\overline{\Delta} = 37.26$		$\overline{\Delta} = 15.22$		$\overline{\Delta} = 2.98$	
λ_c	0		$4.00 \cdot 10^{-5}$		$1.00 \cdot 10^{-4}$		$5.00 \cdot 10^{-4}$		$1.00 \cdot 10^{-1}$	
					Annualized tracking error in percent					
Mean	1.20	1.94	1.21	1.94	1.24	1.93	1.44	1.96	2.67	2.19
SD	0.64	1.36	0.64	1.35	0.63	1.34	0.60	1.27	1.85	1.20
					Annualized active return in percent					
Mean	4.23	0.82	4.27	0.71	4.36	0.60	4.84	0.29	4.66	− 0.81
SD	1.84	8.15	1.82	8.22	1.79	8.22	1.91	8.13	3.10	8.09
					Correlation (corr), turnover (TO), and average stocks ($\bar{k}$)					
corr	0.999	0.996	0.999	0.996	0.999	0.996	0.998	0.996	0.995	0.994
TO	0.208		0.148		0.109		0.048		0.016	
$\bar{k}$	60.83		61.00		60.68		57.72		50.96	
					Panel A: Nikkei 225					
	$\overline{\Delta} = 89.45$		$\overline{\Delta} = 71.64$		$\overline{\Delta} = 49.18$		$\overline{\Delta} = 20.89$		$\overline{\Delta} = 14.91$	
λ_c	0		$2.00 \cdot 10^{-4}$		$6.00 \cdot 10^{-4}$		$5.00 \cdot 10^{-3}$		$1.00 \cdot 10^{-1}$	
					Annualized tracking error in percent					
Mean	1.43	2.63	1.45	2.58	1.52	2.53	2.12	2.62	2.48	2.80
SD	0.24	1.04	0.24	1.00	0.23	0.99	0.41	1.05	0.68	1.24
					Annualized active return in percent					
Mean	2.59	− 1.79	2.73	− 1.77	3.00	− 2.05	4.24	− 2.25	5.09	− 1.34
SD	1.47	8.51	1.44	8.26	1.41	8.28	1.92	9.20	2.37	9.96
					Correlation (corr), turnover (TO), and average stocks ($\bar{k}$)					
corr	0.998	0.993	0.998	0.993	0.998	0.993	0.997	0.993	0.996	0.993
TO	0.393		0.296		0.204		0.090		0.066	
$\bar{k}$	76.00		76.42		76.80		77.16		75.73	

Note: This table reports the in-sample and out-of-sample tracking errors and active returns of the tracking portfolios based on the Aenet approach for each data set. The parameters λ_1 and λ_2 are selected by minimizing the cross-validated tracking errors using the first training window in each data set. Different values of λ_c are considered, leading to the tracking portfolios with different average numbers of the stocks to be rebalanced, $\overline{\Delta}$.

Wu (2016) are based on the setting without the full investment constraint. It is thus of interest to compare the tracking performance among the Lasso-based approaches under this setting. Table 6 reports the in-sample and out-of-sample tracking errors, active returns, correlation and turnover for the Lasso-based approaches under no full investment constraint. Only the no-short selling constraints are considered. For illustration, we set 40 as the number of active stocks in the tracking portfolio for data sets FTSE 100 and S&P 100, and 80 for the data set Nikkei 225. As can be seen from table 6, both the Aenet and Alasso approaches have very competitive tracking performance, and both outperform the Lasso and Enet approaches in term of out-of-sample tracking errors.

The wealth paths corresponding to the different approaches shown in tables 5 and 6 are also provided. Figure 4 shows the wealth growth over the trading horizon among the Lasso, Enet, Alasso, and Aenet approaches under the case with both the no-short selling and full investment constraints. There is negligible difference among these methods for data sets FTSE100 and S&P 100. However, for the data set Nikkei 225 with a relatively large number of stocks, Aenet delivers slightly higher growth than Lasso. Figure 5 further shows the wealth growth over the trading horizon among the above approaches under the no-short selling constraints only. In this case, the wealth growth curve based on Aenet is much closer to that of the benchmark as compared to the Lasso approach, due to the smaller out-of-sample tracking error. The Aenet approach can provide better wealth growth than the Lasso approach for the data set FTSE 100 but slightly worse wealth growth for the data set S&P 100.

5.3. *Effects of the turnover restriction*

The above comparisons are under the ideal case without restriction on turnover, i.e. the regularization parameter $\lambda_c = 0$. However, in reality, it is practical to take into account the transaction cost as well as other costs. The transaction cost depends on the trading volume, which is often a monotonic increasing function of the latter. To shed some insights into effects of the transaction cost on sparse index tracking by using the Aenet method, it is also interesting to investigate the effect of the regularization parameter λ_c. In the following empirical analysis, we set λ_1 and λ_2 the same as the values listed in table 2. The first training window is used to construct the initial portfolio for each data set.

Table 7 reports the tracking performance based on the Aenet method with different regularization parameter λ_c under various data sets. The parameter λ_c discourages the changes in portfolio weights between two consecutive time periods. To examine this, define Δ as the number of stocks to be rebalanced in period $t+1$, expressed as $\Delta = ||w_{t+1} - w_{t+}||_0$, where w_{t+} is the weights before rebalancing. As can be seen from table 7, the average number of stocks rebalanced across the rolling windows significantly decreases with λ_c as expected. Moreover, it is not surprising to see that the turnover decreases with λ_c.

The turnover restriction also has important effects on both the in-sample and out-of-sample tracking errors. From table 7, one can observe an increasing tendency of the in-sample tracking error with λ_c for each data set. For example, in the data set Nikkei 225, the average in-sample tracking error increases from 1.43% to 2.48% as λ_c increases from 0 to 0.1. This is because as λ_c increases, the biases in estimating the portfolio weights would increase. This leads to a large deviation of the in-sample tracking portfolio return from the index return, which in turn increases the in-sample tracking error.

However, the effect of λ_c on the out-of-sample tracking error is more complex. The out-of-sample tracking error does not linearly increase with λ_c, but tends to have a parabolic shape. That is, the out-of-sample tracking errors can be larger for both a larger λ_c and a smaller λ_c. For example, in Nikkei 225, the out-of-sample tracking error (2.53%) is the smallest when $\lambda_c = 6 \times 10^{-4}$. The out-of-sample tracking errors with $\lambda_c = 2 \times 10^{-4}$ and $\lambda_c = 10^{-1}$ are both larger than the out-of-sample tracking error with $\lambda_c = 6 \times 10^{-4}$.

6. Conclusions

A straight way to achieve sparsity in index tracking is to minimize a given tracking error measure by imposing the cardinality constraint, i.e. limiting the maximum number of assets held in the tracking portfolio. However, the resulting problem with the cardinality constraint is NP-hard. Solving such a problem will be computationally expensive, especially when the problem size is large. Compared to the cardinality constraint approach, the regularization approach is more computationally efficient. This paper develops a general regularization method to perform sparse index tracking in high-dimensional settings based on the Aenet penalty, which represents a family of convex regularization methods. The traditional regularization methods are often discussed under no-short selling constraints only. In order to make the formulation more general, this paper also takes other practical conditions into account such as the full investment constraint and a restriction on turnover. A coordinate descent algorithm with closed-form updates is developed to solve the resulting optimization problem.

The empirical results show that the Aenet approach requires much less running time to solve the sparse index tracking problem than the cardinality constraint approach. Although both the Aenet and cardinality approaches yield nearly the same out-of-sample tracking errors, the former can achieve lower turnover and more diversification than the latter when the tracking portfolio invests in a relatively large number of stocks.

It is important to note that the traditional methods regularized with Lasso and Enet penalties are ineffective in promoting sparsity under no-short selling and full investment conditions. The proposed method and Alasso are however effective in promoting sparsity. Compared to the Alasso method, the proposed method makes use of an additional ℓ_2 penalty, which can stabilize the solution paths. In general, a tracking portfolio based on the Aenet approach has smaller turnover than that based on the Alasso approach.

Acknowledgments

The authors would like to thank the editor, managing editor, and two referees for their valuable comments.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

Prof. Shu's work was funded in part by the Research Committee under the grant MYRG2018-00087-FBA, and by the Science and Technology Development Fund, Macau SAR (File No. FDCT/0064/2018/A2). Prof. Tian's work was supported by the grant from the National Natural Science Foundation of China (No. 11771199).

ORCID

Lianjie Shu http://orcid.org/0000-0001-6860-698X

References

Acosta-González, E., Armas-Herrera, R. and Fernández-Rodríguez, F., On the index tracking and the statistical arbitrage choosing the stocks by means of cointegration: The role of stock picking. *Quant. Finance*, 2015, **15**(6), 1075–1091.

Alexander, C. and Dimitriu, A., Indexing and statistical arbitrage. *J. Portfolio Manag.*, 2005, **31**(2), 50–63.

Beasley, J.E., Meade, N. and Chang, T.J., An evolutionary heuristic for the index tracking problem. *Eur. J. Oper. Res.*, 2003, **148**(3), 621–643.

Benidis, K., Feng, Y. and Palomar, D.P., Sparse portfolios for high-dimensional financial index tracking. *IEEE Trans. Signal Process.*, 2018, **66**, 155–170.

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

Chang, T.J., Meade, N., Beasley, J.E. and Sharaiha, Y.M., Heuristics for cardinality constrained portfolio optimisation. *Comput. Oper. Res.*, 2000, **27**(13), 1271–1302.

Consiglio, A. and Zenios, S.A., Integrated simulation and optimization models for tracking international fixed income indices. *Math. Program.*, 2001, **89**(2), 311–339.

Efron, B., Hastie, T., Johnstone, I. and Tibshirani, R., Least angle regression. *Ann. Stat.*, 2004, **32**(2), 407–451.

Fan, J. and Li, R., Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Am. Stat. Assoc.*, 2001, **96**(456), 1348–1360.

Fastrich, B., Paterlini, S. and Winker, P., Cardinality versus q-norm constraints for index tracking. *Quant. Finance*, 2014, **14**(11), 2019–2032.

Gilli, M. and Këllezi, E., The threshold accepting heuristic for index tracking. In *Financial Engineering, E-commerce and Supply Chain, Applied Optimization*, edited by P.M. Pardalos and V.K. Tsitsiringos, Vol. 70, 2002 (Springer: Boston, MA).

Guastaroba, G. and Speranza, M.G., Kernel search: An application to the index tracking problem. *Eur. J. Oper. Res.*, 2012, **217**(1), 54–68.

Jansen, R. and Van Dijk, R., Optimal benchmark tracking with small portfolios. *J. Portfolio Manag.*, 2002, **28**(2), 33–39.

Krink, T., Mittnik, S. and Paterlini, S., Differential evolution and combinatorial search for constrained index-tracking. *Ann. Oper. Res.*, 2009, **172**(1), 153–176.

Maringer, D. and Oyewumi, O., Index tracking with constrained portfolios. *Intell. Syst. Account. Fin. Manag.*, 2007, **15**(1–2), 57–71.

Montfort, K.v., Visser, E. and van Draat, L.F., Index tracking by means of optimized sampling. *J. Portfolio Manag.*, 2008, **34**(2), 143–152.

Oh, K.J., Kim, T.Y. and Min, S., Using genetic algorithm to support portfolio optimization for index fund management. *Expert Syst. Appl.*, 2005, **28**(2), 371–379.

Rudolf, M., Wolter, H.J. and Zimmermann, H., A linear model for tracking error minimization. *J. Bank. Financ.*, 1999, **23**(1), 85–103.

Ruiz-Torrubiano, R. and Suárez, A., A hybrid optimization approach to index tracking. *Ann. Oper. Res.*, 2009, **166**(1), 57–71.

Sant'Anna, L.R., Filomena, T.P. and Caldeira, J.F., Index tracking and enhanced indexing using cointegration and correlation with endogenous portfolio selection. *Q. Rev. Econ. Finance*, 2017, **65**, 146–157.

Scozzari, A., Tardella, F., Paterlini, S. and Krink, T., Exact and heuristic approaches for the index tracking problem with UCITS constraints. *Ann. Oper. Res.*, 2013, **205**(1), 235–250.

Shapcott, J., Index tracking: genetic algorithms for investment portfolio selection. Edinburgh Parallel Computing Centre, 1992.

Strub, O. and Trautmann, N., A two-stage approach to the UCITS-constrained index-tracking problem. *Comput. Oper. Res.*, 2019, **103**, 167–183.

Takeda, A., Niranjan, M., Gotoh, J y. and Kawahara, Y., Simultaneous pursuit of out-of-sample performance and sparsity in index tracking portfolios. *Comput. Manag. Sci.*, 2013, **10**(1), 21–49.

Tibshirani, R., Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B*, 1996, **58**, 267–288.

Tibshirani, R., Saunders, M., Rosset, S., Zhu, J. and Knight, K., Sparsity and smoothness via the fused lasso. *J. R. Stat. Soc. Ser. B*, 2005, **67**(1), 91–108.

Tseng, P., Convergence of a block coordinate descent method for nondifferentiable minimization. *J. Optim. Theory. Appl.*, 2001, **109**(3), 475–494.

Wu, L. and Yang, Y., Nonnegative elastic net and application in index tracking. *Appl. Math. Comput.*, 2014, **227**, 541–552.

Wu, L., Yang, Y. and Liu, H., Nonnegative-lasso and application in index tracking. *Comput. Stat. Data. Anal.*, 2014, **70**, 116–126.

Yang, Y. and Wu, L., Nonnegative adaptive lasso for ultra-high dimensional regression models and a two-stage method applied in financial modeling. *J. Stat. Plan. Inference*, 2016, **174**, 52–67.

Yen, Y.M. and Yen, T.J., Solving norm constrained portfolio optimization via coordinate-wise descent algorithms. *Comput. Stat. Data. Anal.*, 2014, **76**, 737–759.

Zou, H., The adaptive lasso and its oracle properties. *J. Am. Stat. Assoc.*, 2006, **101**(476), 1418–1429.

Zou, H. and Hastie, T., Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B*, 2005, **67**(2), 301–320.

Zou, H. and Zhang, H.H., On the adaptive elastic-net with a diverging number of parameters. *Ann. Stat.*, 2009, **37**(4), 1733–1751.

Appendix: Proofs

Proof of Proposition 1. Under both the no-short selling and full investment conditions, the KKT conditions for problem (1)–(3) are

$$2Aw - 2B - \gamma - \gamma_0 e = 0, \tag{A1}$$

$$\gamma_j \geq 0, \tag{A2}$$

$$w_j \geq 0, \tag{A3}$$

$$w_j \gamma_j = 0. \tag{A4}$$

where $\gamma = (\gamma_1, \ldots, \gamma_p)'$ are the Lagrange multipliers for the non-negativity constraints, and γ_0 is the multiplier for the full investment constraint.

To show that w is the solution to the index tracking problem (17)–(18) when A is replaced by $\widetilde{A}$ and B remains unchanged, it suffices to verify the first order condition. Denote $\widetilde{A} = A - (\gamma\, \mathrm{e}' + e\gamma')/2$, the first-order condition in equation (A1) can be written as

$$2Aw - 2B - \gamma - \gamma_0 e = 2\widetilde{A}w + (\gamma\, \mathrm{e}' + e\gamma')w - 2B - \gamma - \gamma_0 e. \tag{A5}$$

Based on the complementary slackness condition $w_i\gamma_i = 0$ for all i, $e\gamma' w = 0$. Moreover, according to the full investment condition, we have $\gamma\, \mathrm{e}' w = \gamma$. Equation (A5) reduces to

$$2Aw - 2B - \gamma - \gamma_0 e = 2\widetilde{A}w - 2B - \gamma_0 e,$$

which coincides with the first-order condition for the index tracking problem (17)–(18). That is, w solves the unconstrained index tracking problem (17)–(18) when A is replaced by $\widetilde{A}$ and B remains unchanged. This proves Proposition 1. ■

Proof of Proposition 2. There exists a Lagrange multiplier λ_c such that the solution to problem (12)–(14) coincides with the solution to

$$\min_{w} \quad w'Aw - 2(B - 1/2\lambda_c\tilde{g})w + \lambda_1\hat{v}'w + \lambda_2 w'w \tag{A6}$$

$$\text{s.t.} \quad w'e = 1, \tag{A7}$$

$$w_j \geq 0. \tag{A8}$$

The KKT conditions for problem (A6)–(A8) are

$$2(A + \lambda_2 I)w - 2(B - 1/2\lambda_c\tilde{g} - 1/2\lambda_1\hat{v}') - \gamma = \gamma_0 e, \tag{A9}$$

$$\gamma_j \geq 0, \tag{A10}$$

$$w_j \geq 0, \tag{A11}$$

$$w_j \gamma_j = 0. \tag{A12}$$

Denote $\widetilde{A} = A - (\gamma\, \mathrm{e}' + e\gamma')/2 + \lambda_2 I)$ and $\widetilde{B} = B - 1/2\lambda_c \tilde{g} - 1/2\lambda_1 \hat{v}'$. To show that w is the solution to the index tracking problem (17)–(18) when A is replaced by $\widetilde{A}$ and B is replaced by $\widetilde{B}$, it suffices to verify the first-order condition

$$\begin{aligned}
2\widetilde{A}w - 2B &= 2(A + \lambda_2 I)w - (\gamma\, \mathrm{e}' + e\gamma')w \\
&\quad - 2(B - 1/2\lambda_c \tilde{g} - 1/2\lambda_1 \hat{v}') \\
&= 2(A + \lambda_2 I)w - \gamma\, \mathrm{e}'w - 2(B - 1/2\lambda_c \tilde{g} - 1/2\lambda_1 \hat{v}') \\
&= 2(A + \lambda_2 I)w - \gamma - 2(B - 1/2\lambda_c \tilde{g} - 1/2\lambda_1 \hat{v}') \\
&= \gamma_0 e.
\end{aligned} \tag{A13}$$

The second equality follows from the complementary slackness condition $w_i \gamma_i = 0$ for all i. The third equality holds because $\mathrm{e}'w = 1$, and the last equality follows from equation (A9). The fact that $2\widetilde{A}w - 2B = \gamma_0 e$ coincides with the first-order condition for the index tracking problem (17)–(18) when A is replaced by $\widetilde{A}$ and B is replaced by $\widetilde{B}$. This proves Proposition 2. ■