Sparse Index Tracking Portfolio with Sector Neutrality

Yuezhang Che, Shuyan Chen and Xin Liu *

School of Statistics and Management, Shanghai University of Finance and Economics, Shanghai 200433, China; 2017211592@163.sufe.edu.cn (Y.C.); schen@163.shufe.edu.cn (S.C.)
* Correspondence: liu.xin@mail.shufe.edu.cn or xliu246@uwo.ca

Abstract: As a popular passive investment strategy, a sparse index tracking strategy has advantages over a full index replication strategy because of higher liquidity and lower transaction costs. Sparsity and nonnegativity constraints are usually assumed in the construction of portfolios in sparse index tracking. However, none of the existing studies considered sector risk exposure of the portfolios that prices of stocks in one sector may fall at the same time due to sudden changes in policy or unexpected events that may affect the whole sector. Therefore, sector neutrality appeals to be critical when building a sparse index tracking portfolio if not using full replication. The statistical approach to sparse index tracking is a constrained variable selection problem. However, the constrained variable selection procedure using Lasso fails to produce a sparse portfolio under sector neutrality constraints. In this paper, we propose a high-dimensional constrained variable selection method using TLP for building index tracking portfolios under sparsity, sector neutrality and nonnegativity constraints. Selection consistency is established for the proposed method, and the asymptotic distribution is obtained for the sparse portfolio weights estimator. We also develop an efficient iteration algorithm for the weight estimation. We examine the performance of the proposed methodology through simulations and an application to the CSI 300 index of China. The results demonstrate the validity and advantages of our methodology.

Keywords: constrained variable selection; high-dimensional variable selection; sparse index tracking; sector neutrality; TLP; ADMM algorithm

MSC: 62H12; 62F12; 62F30; 62P05

1. Introduction

Despite various types and goals, portfolio managing strategies can be classified into two main groups according to styles, namely, active and passive strategies. The former ones attempt to beat the market by exploiting market inefficiency, whereas the latter ones prefer to follow the market. However, the majority of actively managed funds do not outperform the market in the long run [1]. On the contrary, passive funds provide market-level profits without taking an active risk. Passive funds become more popular in recent years. U.S. equity index fund assets have surpassed the assets of their actively managed counterparts for the first time, according to Morningstar’s latest fund flows report. This trend is also growing in other parts of the world.

The most commonly used passive investing strategy is index tracking, which aims to mimic the performance of a specified basket of underlying assets. Without exposure to active risk, an index tracking fund needs to follow the targeting index as closely as possible. The performance of an index tracker is measured by tracking error [2], which is defined as the divergence between the index tracker and the targeting index. Tracking error gives investors a sense of how “tight” the index tracker goes after the index. Smaller tracking error means less exposure to active risk.

A straightforward method to construct an index tracking portfolio is full replication, which is to buy appropriate quantities of all assets composing the index. Theoretically, the
performance of a full replication portfolio should match that of the targeting index perfectly. However, full replication portfolios often incur higher transaction and administration costs than portfolios of relatively fewer stocks. An alternative to full replication is sparse index tracking, which is to replicate the performance of an index by holding a representative sample of the securities in the index [3–5]. The challenge of building a sparse index tracking portfolio lies in the trade-off between transaction costs and tracking efficiency. A sparse index tracking portfolio manager attempts to hold as few securities as possible to reduce the transaction costs and eliminate potential illiquid stocks and curb the tracking error to maintain tracking efficiency at the same time. Sparse index portfolios are often constructed via portfolio optimization. Such an algorithm-driven portfolio aims to minimize the tracking error in both training samples and testing samples. Specifically, index tracking optimization needs to incorporate certain constraints, such as no short selling and balance among industrial sectors. The latter constraint aims at taming sector risk of sparse portfolios that a set of factors particular to a sector drags down the sector’s overall performance in financial risk management. Upon the occurrence of events affecting the entire sector, such as sudden policy changes or technology innovation, the stock prices of many of the companies in the same sector may undergo drastic changes simultaneously. Unbalanced allocation of assets among different sectors can result in the failure of index tracking for a portfolio. In Table 1, we present several periods when sector risk appears for China CSI300 Index. The table shows the performance of the index, the sector gainers and losers during the selected period. The large divergence between the index and the gainers or losers indicates that inappropriate exposure to sector risk can inflate the tracking error and undermine the portfolio performance. To manage the sector risk with index tracking, it is necessary to build a sector-neutral portfolio in which the total weight of stocks within each sector remains the same as that in the index. In particular, a sector-neutral strategy means not to overweight or underweight any given sector compared to its weight in the index, so as to ensure the performance of the portfolio will be stable and accordingly will be not affected by style switch in the market. Therefore, a sector-neutral portfolio can closely fit and help track the performance of a benchmark index, and essentially it turns out to be a passive investment strategy.

<table>
<thead>
<tr>
<th>Period</th>
<th>CSI300</th>
<th>Gainer</th>
<th>Loser</th>
</tr>
</thead>
<tbody>
<tr>
<td>October 2018–November 2018</td>
<td>−4.16%</td>
<td>Financials</td>
<td>Consumer Staple</td>
</tr>
<tr>
<td>February 2018–March 2018</td>
<td>−5.23%</td>
<td>Information</td>
<td>Financials</td>
</tr>
<tr>
<td>June 2017–August 2017</td>
<td>9.75%</td>
<td>Consumer Staple</td>
<td>Utilities</td>
</tr>
<tr>
<td>October 2015–December 2015</td>
<td>8.19%</td>
<td>Information</td>
<td>Energy</td>
</tr>
<tr>
<td>November 2012–January 2013</td>
<td>17.01%</td>
<td>Financials</td>
<td>Consumer Staple</td>
</tr>
</tbody>
</table>

To overcome the drawbacks of full replication and accommodate the sector risk in traditional sparse portfolios, we propose a novel method to construct sparse index-tracking portfolios, assuming that short positions are forbidden and no cash is allowed in the constructed portfolios. Finally, we formulate the sparse index tracking problem with sector neutrality as the minimization of the tracking error under constraints of nonnegativity, sparsity, and sector neutrality.

2. Literature Review

There is a vast amount of research on quantitative methods for index tracking. Ref. [6] studied the traditional Markowitz asset allocation problem and proposed an algorithm for designing a sparse index fund. This algorithm yields a locally optimal index portfolio with a fixed number of securities. Ref. [7] investigated the relation between error measures in statistical tracking and asset allocation restrictions expressed as admissible weight ranges. It addressed the relationship between tracking errors caused by active portfolio management and given tactical asset allocation ranges. Ref. [8] imposed the no-short-sale constraint
on the Markowitz mean-variance optimization and gave an insightful explanation and demonstration of why the constraints help. Ref. [9] considered maximizing the index fund’s tracking accuracy by rebalancing the composition of the index fund’s tracking portfolio in response to new market information and cash deposits and withdrawals from investors. Ref. [10] investigated the inclusion of portfolio liquidity constraints for the construction of index tracking portfolios and proposed two liquidity modelling approaches for index tracking.

Many statistical regularization methods have been applied to the index tracking problem recently. In general, a sparse index tracking portfolio can be constructed via variable selection, which removes non-informative features and yields sparse models, especially for the high-dimensional data, and consequently facilitates inference, interpretability and prediction. Ref. [11] formulated sparse index tracking problem as an optimization problem that minimizes the tracking error, subject to the number of selected assets less than or equal to a preset threshold. Ref. [12] formulated the index tracking as a regression problem whose objective was to minimise the tracking error and adds a $L_0$ penalty on weights corresponding to the amount to invest in each stock, then solved the optimization problem with stochastic neural networks. Ref. [13] investigated the applications of sparse auto-encoder deep-learning architectures with $L_1$ regularization in selecting representative stocks from the index constituents. Ref. [14] analyzed the constraints effect on covariance regularization for index tracking, and developed an $L_1$ and $L_2$ norm constrained minimum-variance portfolio. Ref. [15] reformulated the classical Markowitz mean-variance framework as a constrained least-squares regression problem and added a penalty function to construct a sparse index tracking portfolio. They emphasized that adding an $L_1$ penalty on weights to the objective function may bring several advantages, including promoting sparsity and stabilizing the out-of-sample performance. Ref. [16] offered deep mathematical insights into the utility approximations with the gross-exposure constraint and gave a theoretical justification for the empirical results by [8]. These approaches consist in solving index tracking problems through constraining portfolio norms, for example, the $L_1$ norm as implied in Lasso and the $L_2$ norm as imposed in ridge regression. Besides, many other variable selection methods introduced in various fields can be applied to the index tracking problems as well. Truncated $L_1$ penalty (TLP) [17] has advantages over Lasso [18] in that Lasso gives biased parameter estimates and possibly inconsistent variable selection. Intuitively, unlike Lasso which penalizes all variables, TLP does not penalize variables of large values, and thus enables us to incorporate more complicated constraints into optimization as in Section 3. We asserted in Section 1 that the novelty of our method lies in the construction of a sparse index tracking portfolio with exposure to sector risk. Sector neutrality can be achieved by constrained variable selection procedures. The constrained Lasso [19] introduced additional equality and inequality constraints to the traditional Lasso method for asset allocation, which generalized the work of [16]. However, with the constrained Lasso, the nonnegativity and sector neutrality constraints nullify the Lasso penalty and cause the method fails to give sparsity. More details will be discussed further in Section 3. Ref. [20] followed and added a ridge term into the objective function in constrained Lasso so that it may work in a high-dimensional case where sparsity is necessary.

Computationally solving a constrained variable selection problem is not an easy task, especially when complex penalty functions and constraints are used. The alternating direction method of multipliers (ADMM) algorithm [21] is an efficient and scalable optimization algorithm for convex optimization. Ref. [22] exploited the structure of distributed optimization framework and illustrated this framework via applications such as $L_1$ mean and variance filtering. The idea of the ADMM algorithm is to break the original optimization into iterations of easier problems. However, optimization with the TLP term is not a convex problem. To solve this nonconvex problem, ref. [17] presented an efficient algorithm based on a difference of convex (DC) decomposition [23] and a coordinate descent (CD) algorithm [24,25]. The DC method decomposes the nonconvex constraint into a difference of two convex functions to produce a sequence of approximating convex constraints. Thus
the original nonconvex optimization is broken down into a series of convex optimization problems. In minimization of a multivariate function, the CD algorithm iteratively minimizes the objective function with respect to each coordinate direction until convergence. This is an efficient algorithm for large-scale convex optimization problems.

In this paper, we propose a novel method to solve the sparse index tracking problem with sector neutrality. An error bound for variable selection is obtained for the method, via numerical simulations under different data generation settings. An application is given to index tracking in the Chinese stock market. Both the numerical experiments and application confirm the good performance of our method in general.

This paper is organized as follows. Section 3 formulates the sparse index tracking problem under sector neutrality as an optimization problem under constraints. Section 4 discusses the theory for high-dimensional constrained variable selection. Section 5 describes the algorithm for the optimization problem. Section 6 presents the results of the simulated experiments. Section 7 shows the application of the proposed method to index tracking portfolio construction. Summary and discussion are given in Section 8. Technical proofs and some tables are relegated to the Appendix A.

3. Methodology

From a statistical point of view, index tracking is a linear regression problem:

\[ Y = X\beta + \epsilon, \tag{1} \]

with the response \( Y \in \mathbb{R}^{n \times 1} \) and the covariates \( X \in \mathbb{R}^{n \times p} \), where \( n \) is the sample size and \( p \) is the dimension of covariates. \( \beta \in \mathbb{R}^{p} \) is the parameter of covariates \( X \), and \( \epsilon \) is the error term, we assume \( \epsilon \sim N(0, \sigma^2 I) \), where \( I \) is the identity matrix. In index tracking, \( Y \) and \( X \) represent the returns of an index and its constituents, respectively, and \( \beta \) is the weight vector of the index’s constituents.

Suppose that there are \( q \) sectors. We rewrite the covariates \( X = (X_1, \ldots, X_q) \), where \( X_i = (X_{i,1}, \ldots, X_{i,g_i}) \) are the covariates in the \( i \)-th sector, and \( g_i \) is the number of stocks in the \( i \)-th sector, \( i \in \{1, \ldots, q\} \). Accordingly, we write \( \beta = (\beta_1^T, \ldots, \beta_q^T)^T \), and \( \beta_i = (\beta_{i,1}, \ldots, \beta_{i,g_i})^T \).

For sparse index tracking, three categories of constraints are under consideration:

1. The sparsity constraint within sector: \( \sum_{i=1}^{q} \sum_{j=1}^{g_i} |\beta_{ij}| \leq K \);
2. The sector neutrality constraint: \( \sum_{j=1}^{g_i} \beta_{ij} = \omega_i \), for \( i \in \{1, \ldots, q\} \);
3. The nonnegativity constraint: \( \beta_{ij} \geq 0 \), for \( i \in \{1, \ldots, q\}, j \in \{1, \ldots, g_i\} \).

Here \( \omega_i \) are constants satisfying \( \sum_{i=1}^{q} \omega_i = 1 \). Each \( \omega_i \) is the sum of the original constituent weights in the \( i \)-th sector. These original non-sparse weights are pre-specified by portfolio managers and hence are known in advance from the definition of the index.

The penalty function \( f(\cdot) \) needs to be chosen carefully. The Lasso penalty is widely used in variable selection problems because of its ease of computation. Nevertheless, it turns out that the Lasso penalty fails to yield a sparse portfolio under sector neutrality and nonnegativity constraints for index tracking, as explained below. Under the nonnegativity constraint, the Lasso penalty \( \sum_{i=1}^{q} \sum_{j=1}^{g_i} |\beta_{ij}| \) becomes \( \sum_{i=1}^{q} \sum_{j=1}^{g_i} \beta_{ij} \), while the sector neutrality constraint makes this sum a constant: \( \sum_{i=1}^{q} \sum_{j=1}^{g_i} \beta_{ij} = \sum_{i=1}^{q} \omega_i = 1 \). Therefore, the Lasso penalty becomes invalid and gives no penalty on the portfolio size.
To solve this problem, we take $J(\cdot)$ to be the truncated $L_1$ penalty (TLP) function [17], which can achieve sparse portfolio selection under the sector neutrality and nonnegativity constraints. TLP can be thought of as a truncated version of the $L_1$ penalty approximating the $L_0$-function. As a piecewise linear function, TLP permits efficient computation and adaptive variable selection. TLP is defined as

$$J(|z|) = \min\left(\frac{|z|}{\tau}, 1\right),$$

where $\tau > 0$ is a tuning parameter controlling the degree of approximation. With the TLP function, the estimation of coefficients $\beta$ is obtained by minimization:

$$\min_\beta \text{TE}(\beta) + \lambda \sum_{i=1}^q \sum_{j=1}^{g_i} J(|\beta_{i,j}|). \quad (2)$$

where $\text{TE}(\beta)$ is the tracking error. This is the dual problem corresponding to the constrained primal problem

$$\min_\beta \text{TE}(\beta), \text{ subject to } \sum_{i=1}^q \sum_{j=1}^{g_i} J(|\beta_{i,j}|) \leq K.$$

The dual problem is not equivalent to the constrained primal problem in general. However, with the TLP as the penalty function, the two optimization problems are equivalent [17]. In the remainder of the paper, we will consider only the unconstrained dual problem for its computational advantages. Minimization of (2) reduces to a general weighted Lasso problem, which can be solved by many efficient algorithms [18,26,27].

The tracking error $\text{TE}(\beta)$ can assume many forms [28], for example, the empirical tracking error (ETE), defined as

$$\text{TE}(\beta) = \frac{1}{n} \|Y - X\beta\|_2^2. \quad (3)$$

With TLP and the empirical tracking error, we can rewrite (2) as

$$\min_\beta f(\beta) = n^{-1}\|Y - X\beta\|_2^2 + \lambda \sum_{i=1}^q \sum_{j=1}^{g_i} \min\left\{\frac{|\beta_{i,j}|}{\tau}, 1\right\},$$

subject to:

$$\sum_{j=1}^{g_i} \beta_{i,j} = \omega_i, \ i \in \{1,\ldots,q\},$$

$$\beta_{i,j} \geq 0, \ i \in \{1,\ldots,q\}, \ j \in \{1,\ldots,g_i\}. \quad (4)$$

4. Theory

Let us first introduce some notations. The true value of $\beta$ is denoted by $\beta^*$. Without loss of generality, we assume that the last component $\beta_{i,g_i}^*$ of $\beta_i^*$ is nonzero for $i \in \{1,\ldots,q\}$. By virtue of the sector neutrality constraint, we can solve for the last component $\beta_{i,g_i}$ in each sector:

$$\beta_{i,g_i} = \omega_i - \sum_{j=1}^{g_i-1} \beta_{i,j}, \ i \in \{1,\ldots,q\}. \quad (5)$$

We write

$$\beta_{-q} = (\beta_{1,1}, \ldots, \beta_{1,g_1-1}, \ldots, \beta_{q,1}, \ldots, \beta_{q,g_q-1})^T,$$

$$X_{-q} = (X_{1,1}, \ldots, X_{1,g_1-1}, \ldots, X_{q,1}, \ldots, X_{q,g_q-1}),$$

$$X_G = (X_{1,1} - X_{1,g_1}, \ldots, X_{1,g_1-1} - X_{1,g_1}, \ldots, X_{q,1} - X_{q,g_q}, \ldots, X_{q,g_q-1} - X_{q,g_q}).$$
For any subscripts set \( A \subset \{(i, j) : i \in \{1, \ldots, q\}, j \in \{1, \ldots, g_i - 1\} \} \), we use \( X_A \) to denote the matrix consisting of the columns of \( X_C \) with subscripts in \( A \); likewise, \( \beta_A \) comprises all the elements of \( \beta_{-q} \) with subscripts in \( A \).

Let

\[ A^* = \{(i, j) : \beta^*_{i,j} > 0, i \in \{1, \ldots, q\}, j \in \{1, \ldots, g_i - 1\} \} \]

be the set of subscripts of all nonzero true coefficients in \( \beta^*_{-q} \). We denote the cardinality of \( A^* \) as \( p_0 = |A^*| \), and the minimal value of nonzero true coefficients as \( \gamma_{\min} = \min \{ \beta^*_{i,j} : \beta^*_{i,j} > 0, i \in \{1, \ldots, q\}, j \in \{1, \ldots, g_i\} \} \).

By (5), the regression Equation (1) can be rewritten as

\[
Y - \sum_{i=1}^{q} \sum_{j=1}^{g_i-1} \beta^*_{i,j} X_{i,j} + e = \sum_{i=1}^{q} \sum_{j=1}^{g_i-1} \beta^*_{i,j} X_{i,j} + e
\]

By (5), the regression Equation (1) can be rewritten as

\[
Y - \sum_{i=1}^{q} \sum_{j=1}^{g_i-1} \beta^*_{i,j} X_{i,j} + e = \sum_{i=1}^{q} \sum_{j=1}^{g_i-1} \beta^*_{i,j} X_{i,j} + e.
\]

Define

\[
\hat{\beta}^{\text{ols}}_A = (X_A^\top X_A)^{-1} X_A^\top (Y - \sum_{i=1}^{q} w_i X_{i,g_i}).
\]

We expand \( \hat{\beta}^{\text{ols}}_A \) to be the oracle estimator \( \hat{\beta}^{\text{ols}} \) by letting \( \hat{\beta}^{\text{ols}}_{i,j} = 0 \) if \( \beta^*_{i,j} = 0 \), and obtaining \( \hat{\beta}^{\text{ols}} \) by Equation (5).

Let \( \hat{\beta}^{\text{lp}} \) be the global minimizer of (4), and let

\[ \hat{A} = \{(i, j) : \hat{\beta}^{\text{lp}}_{i,j} > 0, i \in \{1, \ldots, q\}, j \in \{1, \ldots, g_i - 1\} \}. \]

be the estimated set of nonzero coefficients.

We now give two key assumptions when establishing theoretical results regarding the error bound and asymptotic distribution of the coefficient estimator, and model selection consistency.

**Assumption 1.** For some constants \( d_0 > 0 \) and \( \alpha > 1 \), \( C_{\min} \geq d_0 \log p \), where

\[
C_{\min} = \inf_{\{\beta_A : |A| \leq \alpha |A^*|, A \neq A^* \}} \frac{|X_A \beta_A - X_A \beta_A|^2}{\max(|A^* \setminus A|, 1)}.
\]

**Assumption 2.** For some constants \( d_1 > 0 \) and \( d_2 > 0 \),

\[
d_1 \leq \eta_{\min} \left( \frac{1}{n} X^\top X \right) \leq \eta_{\max} \left( \frac{1}{n} X^\top X \right) \leq d_2,
\]

where \( \eta_{\min}(\cdot) \) and \( \eta_{\max}(\cdot) \) denote the minimum and maximum eigenvalues of a matrix, respectively.

**Theorem 1.** Assume that Assumption 1 holds, and \( 0 < \tau \leq \sqrt{\frac{L}{(n + 1) \eta_{\max}(\frac{1}{n} X^\top X)}} \). Then

\[
\Pr[\hat{\beta}^{\text{lp}} \neq \hat{\beta}^{\text{ols}}] \text{ is upper bounded by}
\]
(B) Asymptotic distribution: Let $F$ be the distribution function of $N(0, 1)$. Assume there are constants $a, a_C, a_\lambda, a_{p_0}$ and $a_{\beta}$ satisfying

$$a_{p_0} \geq 0, a_{\beta} \geq 0, a_{p_0} < a_C, a_\lambda < a_\gamma, a_{\beta} - 1 < a_\lambda, a_\lambda < a_C$$

and

$$\gamma_{\min} \asymp n^{1-a_\gamma}, C_{\min} \asymp n^{a_C}, \lambda \asymp n^{a_\lambda}, p_0 \asymp n^{a_{p_0}}, \log(p) \asymp n^{a_\beta},$$

where for two sequences $\{h_n\}$ and $\{g_n\}$ we say $h_n \asymp g_n$ if both sequences $\frac{h_n}{g_n}$ and $\frac{g_n}{h_n}$ are bounded. Also assume that the conditions of Theorem 1 and Assumption 2 hold. Then

(A) Selection consistency:

$$\Pr[\hat{A} \neq A^*] \leq \Pr[\hat{A} \neq \hat{A}^*] \to 0, \text{ as } n, p \to \infty.$$ 

(B) Asymptotic distribution: Let $F_n(t)$ denote the distribution function of

$$\sqrt{n} \left( \frac{X^\top A^\top - \beta^*_A}{\sqrt{n}} \right),$$

and $\Phi_{p_0}(t)$ the distribution function of $p_0$-dimensional standard multivariate normal distribution. We have

$$\lim_{n \to \infty} \sup_{t \in \mathbb{R}^{p_0}} |F_n(t) - \Phi_{p_0}(t)| = 0. \tag{7}$$

Remark 1. Theorems 1 and 2 stabilize the index tracking problem. They encourage the non-negativity and sector neutrality in constructing sparse index tracking portfolios in high dimensional cases and allow practical and empirical work with only a moderate size of training data. The proofs of Theorems 1 and 2 are displayed in Appendix A.1.

5. Computation

The minimization of (4) with TLP can be treated as a sequence of weighted Lasso problems [17] and can be solved iteratively. However, the sector neutrality constraint and the nonnegativity constraint in (4) are not easy to handle directly. Fortunately, the alternating direction method of multipliers (ADMM) algorithm [21,22] can be applied to solve the constrained minimization problem (4).

We are now to put our optimization problem in the ADMM framework. First, using the Lagrangian multiplier method, we transform the original optimization into minimization of the objective function

$$f(\beta) = n^{-1} \| Y - X\beta \|_2^2 + \lambda \sum_{i=1}^q \sum_{j=1}^s \min \left\{ \frac{|\beta_{ij}|}{\tau}, 1 \right\}. \tag{8}$$
over the new parameter space $C$:

$$C = \{ \beta : \beta_{ij} \geq 0, \sum_{j=1}^{q} \beta_{ij} = \alpha_i, \ i \in \{1, \ldots, q\}, \ j \in \{1, \ldots, g_i\} \}.$$  

It is straightforward to verify that the new parameter space $C$ is convex and the objective function $f$ in (8) is a convex loss function plus a piecewise linear regularization function. Using the results in [21,29], our optimization problem is equivalent to

$$\min_{\beta} \ f(\beta) + I_C(\alpha)$$

subject to: $\beta_{ij} = \alpha_{ij}$, $i \in \{1, \ldots, q\}, j \in \{1, \ldots, g_i\}$, (9)

where $I_C(\alpha)$ is the indicator function of space $C$ (i.e., $I_C(\alpha) = 0$ for $\alpha \in C$, and $I_C(\alpha) = \infty$ for $\alpha \not\in C$). This is a standard starting form of the ADMM algorithm. Before following three steps in the ADMM algorithm, we write the augmented Lagrangian for (9),

$$h(\beta) = n^{-1}\|Y - X\beta\|_2^2 + \lambda \sum_{i=1}^{q} \sum_{j=1}^{g_i} \min\{ |\beta_{ij}|/\tau, 1\} + \rho/2 \|\beta - \alpha + \mu\|_2^2$$  

where $\mu$ is a scaled dual variable associated with the constraint $\beta = \alpha$. Here $\rho > 0$ is a penalty parameter.

In each iteration of ADMM, we perform alternating minimization of the augmented Lagrangian over $\beta$ and $\alpha$. At the $k$-th iteration we update variables $\beta$, $\alpha$ and $\mu$ by the following steps:

$$\beta^{k+1} = \arg \min_{\beta} n^{-1}\|Y - X\beta\|_2^2 + \lambda \sum_{i=1}^{q} \sum_{j=1}^{g_i} \min\{ |\beta_{ij}|/\tau, 1\} + \rho/2 \|\beta - \alpha + \mu^k\|_2^2$$  

(11)

$$\alpha^{k+1} = \prod_{C}(\beta^{k+1} + u^k)$$  

(12)

$$\mu^{k+1} = \mu^k + (\beta^{k+1} - \alpha^{k+1})$$  

(13)

where $\prod_{C}$ denotes the Euclidean projection onto space $C$. In the first step of ADMM, we fix $\alpha$ and $\mu$ and update the value of $\beta$ by minimization of the augmented Lagrangian; in the second step, we fix $\beta$ and $\mu$ and update the value of $\alpha$ by projection onto space $C$; and finally, we fix $\alpha$ and $\beta$ and update the dual variable $\mu$. Algorithm 1 summarizes the framework of our algorithm and the details are referred to Appendix A.2.

**Algorithm 1** ADMM algorithm for the minimization of (10).

(Initialization) Let $\beta^0, \alpha^0, \mu^0 \in \mathbb{R}^p$ be the initial parameter vectors;  
(Iteration) At each iteration $k$, Update $\beta$ using Algorithm A1 (see Appendix A.2), $\alpha$ by the projection (12), and $\mu$ by direct calculation (13);  
(Termination) Stop when $\beta$, $\alpha$, $\mu$ converge.

6. Simulation Results

In this section, we show the performance of our estimation procedure via simulations and discuss the selection of appropriate tuning parameters $\lambda$, $\tau$ and $\rho$ in the augmented Lagrangian (10).

The detailed information about the simulation study, including the model and its variations, the methods to conduct the experiments and the choice of tuning parameters are given in Appendix A.3.

The performance of various methods is evaluated by both variable selection performance and estimation accuracy. In terms of variable selection, we use four criteria:
the mean true positive(TP), the mean false positive(FP), positive selection rate(PSR), and negative selection rate(NSR) [30], respectively. The true positive(TP) is defined as 
\[
\text{#TP} = \sum_{i=1}^{n} \sum_{j=1}^{q} I(\hat{\beta}_{ij}^* = 0, \hat{\beta}_{ij} \neq 0),
\]
which counts the variables with true non-zero coefficients and estimated non-zero coefficients. The false positive(FP) is defined as 
\[
\text{#FP} = \sum_{i=1}^{n} \sum_{j=1}^{q} I(\hat{\beta}_{ij}^* = 0, \hat{\beta}_{ij} \neq 0),
\]
which counts the variables with true zero coefficients but estimated as non-zero coefficients. Additionally, PSR is the ratio of TP and the total number of the true non-zero coefficients. Similar to PSR, NSR is the ratio of FP and the total number of the true zero coefficients. Regarding estimation accuracy, we adopt the mean and the standard deviation of model errors (ME) as criteria, where the model error is defined as 
\[
ME(\hat{\beta}) = (\hat{\beta} - \beta^*)^\top \mathbb{E}(XX^\top)(\hat{\beta} - \beta^*),
\]
and the expectation is taken only with respect to new observation \((X, y)\). The running time of each method is evaluated using the machine with Intel Core i5 CPU, 2.4 GHz, 8 GB RAM. All methods were implemented in R.

6.1. Case 1

In this case \(n > p\), five methods are compared. The first one is our method, named as constrained TLP method (CTLP). The second method uses the Lasso penalty for sparsity constraint instead of TLP, and we name it the constrained \(L_1\) method (CL\(_1\)). This method will fail to give a sparse portfolio because of the neutrality constraint, as is pointed out in Section 3; we present the results of CL\(_1\) here for confirmation. The third method uses TLP for the sparsity constraint, but ignores the sector neutrality, and we refer to this method as TLP. The fourth and fifth methods are the index tracking procedures in the following form,

\[
\min_\beta \quad \text{TE}(\beta) + \lambda \|\beta\|_0
\]

subject to \(\beta^\top 1 = 1\) and \(\beta > 0\),

with \(\text{TE}(\beta)\) being empirical tracking error (ETE) and Huber empirical tracking error (HETE), which are denoted by ETE and HETE, respectively. In ETE, \(\text{TE}(\beta)\) is defined in (3), while in HETE, \(\text{TE}(\beta) = \frac{1}{n} 1^\top \phi(y - X\beta)\) where \(\phi(x) = (\phi(x_1), \ldots, \phi(x_n))^\top\), and

\[
\phi(x) = \begin{cases} 
  x^2 & \text{if } |x| \leq M \\
  M(2|x| - M) & \text{if } |x| > M,
\end{cases}
\]

with \(M\) being the Huber parameter. Note that ETE and HETE methods ignore the sector-neutral constraints. These two methods can be carried out using R-package sparseIndexTracking [31].

We first investigate the case when the covariates \(X\) have an independent structure. Table 2 shows the simulation results for both \(\sigma = 1\) and \(\sigma = 3\) settings. For the \(\sigma = 1\) setting, methods CTLP and CL\(_1\) have smaller model errors than ETE and HETE, and the model errors of ETE and HETE are much smaller than that of TLP. When it comes to the variable selection ability, all methods correctly select non-zero variables. CTLP, ETE and HETE perform similarly, and better than CL\(_1\) in that CL\(_1\) tends to select too many trivial variables and thus fails to produce sparse portfolios, which confirms our earlier assertion. On the other hand, TLP makes the fewest mistakes in selecting both trivial and non-trivial variables. When the variance of error terms becomes larger, the simulated data are noisier. For the \(\sigma = 3\) setting, CTLP and CL\(_1\) still have the smallest model error while TLP performs quite poorly in terms of model errors. Every method selects almost all nontrivial variables except for TLP, which selects slightly fewer nontrivial and trivial variables.

Now we investigate the simulation settings when the covariates \(X\) have AR(1) correlation, without and with group structures. Similar conclusions can be reached from Tables 3 and 4. In the large signal-to-noise ratio setting (\(\sigma = 1\)), the model errors of CTLP and CL\(_1\) are similar and smaller than that of their competitors. Besides, CTLP, ETE and HETE maintain a good balance between selections of trivial and nontrivial variables whereas CL\(_1\) selects too many trivial variables. TLP gives good variable selection results but
has poor estimation accuracy. In the small signal-to-noise ratio setting (\( \sigma = 3 \)), the model errors of all five methods become larger than those in the case of the large signal-to-noise ratio setting above. However, the relative performances of the five methods are similar in terms of model errors and variable selection ability as well.

### Table 2. The comparison of five methods in the case of \( n > p \), independent correlation matrix.

<table>
<thead>
<tr>
<th>Method</th>
<th>ME</th>
<th>SE</th>
<th>TP</th>
<th>FP</th>
<th>PSR</th>
<th>NSR</th>
<th>Runtime (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \sigma = 1 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CTLP</td>
<td>0.0742</td>
<td>0.0025</td>
<td>14</td>
<td>18.85</td>
<td>100.00%</td>
<td>21.92%</td>
<td>2.78</td>
</tr>
<tr>
<td>CL(_1)</td>
<td>0.0741</td>
<td>0.0025</td>
<td>14</td>
<td>52.36</td>
<td>100.00%</td>
<td>60.88%</td>
<td>1.10</td>
</tr>
<tr>
<td>TLP</td>
<td>0.1650</td>
<td>0.0042</td>
<td>14</td>
<td>6.77</td>
<td>100.00%</td>
<td>7.87%</td>
<td>0.12</td>
</tr>
<tr>
<td>ETE</td>
<td>0.0858</td>
<td>0.0026</td>
<td>14</td>
<td>18.76</td>
<td>100.00%</td>
<td>21.81%</td>
<td>0.02</td>
</tr>
<tr>
<td>HETE</td>
<td>0.0924</td>
<td>0.0026</td>
<td>14</td>
<td>18.8</td>
<td>100.00%</td>
<td>21.86%</td>
<td>1.45</td>
</tr>
<tr>
<td>( \sigma = 3 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CTLP</td>
<td>0.6415</td>
<td>0.0202</td>
<td>13.98</td>
<td>15.64</td>
<td>99.86%</td>
<td>18.19%</td>
<td>1.66</td>
</tr>
<tr>
<td>CL(_1)</td>
<td>0.6712</td>
<td>0.0198</td>
<td>13.98</td>
<td>31.55</td>
<td>99.86%</td>
<td>36.69%</td>
<td>1.00</td>
</tr>
<tr>
<td>TLP</td>
<td>1.2989</td>
<td>0.0338</td>
<td>13.55</td>
<td>7.52</td>
<td>96.79%</td>
<td>8.74%</td>
<td>0.11</td>
</tr>
<tr>
<td>ETE</td>
<td>0.7673</td>
<td>0.0215</td>
<td>13.95</td>
<td>18.74</td>
<td>99.64%</td>
<td>21.79%</td>
<td>0.02</td>
</tr>
<tr>
<td>HETE</td>
<td>0.9818</td>
<td>0.0278</td>
<td>13.9</td>
<td>19.33</td>
<td>99.29%</td>
<td>22.48%</td>
<td>1.47</td>
</tr>
</tbody>
</table>

### Table 3. The comparison of five methods in the case of \( n > p \), AR(1) correlation matrix without group structure.

<table>
<thead>
<tr>
<th>Method</th>
<th>ME</th>
<th>SE</th>
<th>TP</th>
<th>FP</th>
<th>PSR</th>
<th>NSR</th>
<th>Runtime (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \sigma = 1 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CTLP</td>
<td>0.0697</td>
<td>0.0024</td>
<td>14</td>
<td>16.65</td>
<td>100.00%</td>
<td>19.36%</td>
<td>2.81</td>
</tr>
<tr>
<td>CL(_1)</td>
<td>0.0696</td>
<td>0.0024</td>
<td>14</td>
<td>51.04</td>
<td>100.00%</td>
<td>59.35%</td>
<td>1.10</td>
</tr>
<tr>
<td>TLP</td>
<td>0.1382</td>
<td>0.0043</td>
<td>14</td>
<td>6.74</td>
<td>100.00%</td>
<td>7.84%</td>
<td>0.14</td>
</tr>
<tr>
<td>ETE</td>
<td>0.0796</td>
<td>0.0025</td>
<td>14</td>
<td>16.22</td>
<td>100.00%</td>
<td>18.86%</td>
<td>0.02</td>
</tr>
<tr>
<td>HETE</td>
<td>0.0884</td>
<td>0.0026</td>
<td>14</td>
<td>16.21</td>
<td>100.00%</td>
<td>18.85%</td>
<td>1.49</td>
</tr>
<tr>
<td>( \sigma = 3 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CTLP</td>
<td>0.5956</td>
<td>0.0171</td>
<td>13.98</td>
<td>12.92</td>
<td>99.86%</td>
<td>15.02%</td>
<td>1.95</td>
</tr>
<tr>
<td>CL(_1)</td>
<td>0.6161</td>
<td>0.0179</td>
<td>13.98</td>
<td>29.21</td>
<td>99.79%</td>
<td>33.97%</td>
<td>1.12</td>
</tr>
<tr>
<td>TLP</td>
<td>1.0480</td>
<td>0.0268</td>
<td>13.81</td>
<td>7.26</td>
<td>98.64%</td>
<td>8.44%</td>
<td>0.13</td>
</tr>
<tr>
<td>ETE</td>
<td>0.7136</td>
<td>0.0196</td>
<td>13.95</td>
<td>16.2</td>
<td>99.64%</td>
<td>18.84%</td>
<td>0.02</td>
</tr>
<tr>
<td>HETE</td>
<td>0.9427</td>
<td>0.0285</td>
<td>13.91</td>
<td>17.73</td>
<td>99.36%</td>
<td>20.62%</td>
<td>1.43</td>
</tr>
</tbody>
</table>

### Table 4. The comparison of five methods in the case of \( n > p \), AR(1) correlation matrix with group structure.

<table>
<thead>
<tr>
<th>Method</th>
<th>ME</th>
<th>SE</th>
<th>TP</th>
<th>FP</th>
<th>PSR</th>
<th>NSR</th>
<th>Runtime (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \sigma = 1 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CTLP</td>
<td>0.0651</td>
<td>0.0020</td>
<td>14</td>
<td>15.26</td>
<td>100.00%</td>
<td>17.74%</td>
<td>2.75</td>
</tr>
<tr>
<td>CL(_1)</td>
<td>0.0651</td>
<td>0.0020</td>
<td>14</td>
<td>50.32</td>
<td>100.00%</td>
<td>58.51%</td>
<td>1.04</td>
</tr>
<tr>
<td>TLP</td>
<td>0.1337</td>
<td>0.0039</td>
<td>14</td>
<td>6.28</td>
<td>100.00%</td>
<td>7.30%</td>
<td>0.14</td>
</tr>
<tr>
<td>ETE</td>
<td>0.0765</td>
<td>0.0022</td>
<td>14</td>
<td>15.32</td>
<td>100.00%</td>
<td>17.81%</td>
<td>0.02</td>
</tr>
<tr>
<td>HETE</td>
<td>0.0841</td>
<td>0.0025</td>
<td>14</td>
<td>15.4</td>
<td>100.00%</td>
<td>17.91%</td>
<td>1.42</td>
</tr>
<tr>
<td>( \sigma = 3 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CTLP</td>
<td>0.5740</td>
<td>0.0171</td>
<td>13.99</td>
<td>13.36</td>
<td>99.93%</td>
<td>15.53%</td>
<td>1.96</td>
</tr>
<tr>
<td>CL(_1)</td>
<td>0.5836</td>
<td>0.0185</td>
<td>13.99</td>
<td>44.79</td>
<td>99.93%</td>
<td>52.08%</td>
<td>1.20</td>
</tr>
<tr>
<td>TLP</td>
<td>0.9061</td>
<td>0.0276</td>
<td>13.89</td>
<td>10.65</td>
<td>99.21%</td>
<td>12.38%</td>
<td>0.13</td>
</tr>
<tr>
<td>ETE</td>
<td>0.6742</td>
<td>0.0203</td>
<td>13.96</td>
<td>17.07</td>
<td>99.71%</td>
<td>19.85%</td>
<td>0.02</td>
</tr>
<tr>
<td>HETE</td>
<td>0.9259</td>
<td>0.0253</td>
<td>13.9</td>
<td>17.05</td>
<td>99.29%</td>
<td>19.83%</td>
<td>1.49</td>
</tr>
</tbody>
</table>

From the last columns in Tables 2–4 we see that CTLP takes more time than other methods in training because of additional loops required to satisfy the sector neutrality constraint. In contrast, TLP and ETE consume the least time among all methods.
6.2. Case 2

When \( p > n \), the variable selection becomes a high-dimensional problem and necessitates sparsity. Theoretically, the CL1 method with the Lasso penalty fails to yield sparse models due to sector neutrality and nonnegativity constraints. However, we can solve the convex optimization problem using some common convex optimizers like CVXR [32] to obtain an approximate solution \( CL_1^* \). Tables 5–7 present the results of this case.

The results are similar to that of the low-dimensional case, but all four methods have larger model errors and select fewer nontrivial variables. In the large signal-to-noise ratio setting (\( \sigma = 1 \)), CTLP has smaller model errors than other methods. CTLP, ETE and HETE have a similar ability in variable selection. TLP has larger model errors and it tends to select fewer nontrivial variables. The average model error of \( CL_1^* \) is between CTLP and other methods. It usually selects about 20% trivial variables as in tables if we consider the numbers less than \( 10^{-4} \) as 0. As the accuracy of numbers increases, more trivial variables are selected. No estimated coefficients of variables are exactly 0 with moderate accuracy, which proves that \( CL_1^* \) is just an approximate solution and the \( L_1 \) term does not work. In the small signal-to-noise ratio setting (\( \sigma = 3 \)), there is no significant difference between CTLP and ETE in terms of model errors, but CTLP selects fewer trivial variables. As to the other two methods, HETE gives less accurate estimation, and TLP performs less well in terms of both variable selection and estimation. Although \( CL_1^* \) yields the smallest model errors, it cannot provide the exact solution with the sparsity that we need. All five methods run faster than in the \( p < n \) case due to the smaller sample size. CTLP is still the slowest one while ETE is the fastest one.

In conclusion, although CTLP takes more time to run, it performs similarly or better than the competitors in terms of both model errors and variable selection ability.

### Table 5. The comparison of four methods in the case of \( n < p \), independent correlation matrix.

<table>
<thead>
<tr>
<th>Method</th>
<th>ME</th>
<th>SE</th>
<th>TP</th>
<th>FP</th>
<th>PSR</th>
<th>NSR</th>
<th>Runtime (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTLP</td>
<td>0.1352</td>
<td>0.0114</td>
<td>13.61</td>
<td>18.27</td>
<td>97.21%</td>
<td>21.24%</td>
<td>2.35</td>
</tr>
<tr>
<td>( CL_1^* )</td>
<td>0.1758</td>
<td>0.0101</td>
<td>13.65</td>
<td>19.10</td>
<td>97.50%</td>
<td>22.21%</td>
<td>0.64</td>
</tr>
<tr>
<td>TLP</td>
<td>0.3433</td>
<td>0.0299</td>
<td>11.09</td>
<td>12.77</td>
<td>79.21%</td>
<td>14.85%</td>
<td>0.10</td>
</tr>
<tr>
<td>ETE</td>
<td>0.1724</td>
<td>0.0134</td>
<td>13.34</td>
<td>18.39</td>
<td>95.29%</td>
<td>21.38%</td>
<td>0.07</td>
</tr>
<tr>
<td>HETE</td>
<td>0.1772</td>
<td>0.0134</td>
<td>13.32</td>
<td>18.2</td>
<td>95.14%</td>
<td>21.16%</td>
<td>0.33</td>
</tr>
</tbody>
</table>

### Table 6. The comparison of four methods in the case of \( n < p \), AR(1) correlation matrix without group structure.

<table>
<thead>
<tr>
<th>Method</th>
<th>ME</th>
<th>SE</th>
<th>TP</th>
<th>FP</th>
<th>PSR</th>
<th>NSR</th>
<th>Runtime (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTLP</td>
<td>0.1268</td>
<td>0.0088</td>
<td>13.69</td>
<td>14.68</td>
<td>97.79%</td>
<td>17.07%</td>
<td>2.17</td>
</tr>
<tr>
<td>( CL_1^* )</td>
<td>0.1328</td>
<td>0.0062</td>
<td>13.79</td>
<td>16.86</td>
<td>98.50%</td>
<td>19.60%</td>
<td>0.65</td>
</tr>
<tr>
<td>TLP</td>
<td>0.3060</td>
<td>0.0261</td>
<td>11.87</td>
<td>11.02</td>
<td>84.79%</td>
<td>12.81%</td>
<td>0.10</td>
</tr>
<tr>
<td>ETE</td>
<td>0.1643</td>
<td>0.0111</td>
<td>13.36</td>
<td>15.02</td>
<td>95.43%</td>
<td>17.47%</td>
<td>0.06</td>
</tr>
<tr>
<td>HETE</td>
<td>0.1628</td>
<td>0.0110</td>
<td>13.33</td>
<td>14.71</td>
<td>95.21%</td>
<td>17.10%</td>
<td>0.31</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Method</th>
<th>ME</th>
<th>SE</th>
<th>TP</th>
<th>FP</th>
<th>PSR</th>
<th>NSR</th>
<th>Runtime (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTLP</td>
<td>1.0728</td>
<td>0.0772</td>
<td>10.79</td>
<td>11.67</td>
<td>77.07%</td>
<td>13.57%</td>
<td>2.67</td>
</tr>
<tr>
<td>( CL_1^* )</td>
<td>0.8185</td>
<td>0.0304</td>
<td>11.24</td>
<td>15.54</td>
<td>80.29%</td>
<td>18.07%</td>
<td>0.64</td>
</tr>
<tr>
<td>TLP</td>
<td>1.8885</td>
<td>0.0644</td>
<td>7.03</td>
<td>4.53</td>
<td>50.21%</td>
<td>5.27%</td>
<td>0.05</td>
</tr>
<tr>
<td>ETE</td>
<td>1.0795</td>
<td>0.0439</td>
<td>10.86</td>
<td>15.4</td>
<td>77.57%</td>
<td>17.91%</td>
<td>0.03</td>
</tr>
<tr>
<td>HETE</td>
<td>1.2447</td>
<td>0.0507</td>
<td>10.39</td>
<td>15.27</td>
<td>74.21%</td>
<td>17.76%</td>
<td>0.37</td>
</tr>
</tbody>
</table>
Table 7. The comparison of four methods in the case of $n < p$, AR(1) correlation matrix with group structure.

<table>
<thead>
<tr>
<th>Method</th>
<th>ME</th>
<th>SE</th>
<th>TP</th>
<th>FP</th>
<th>PSR</th>
<th>NSR</th>
<th>Runtime (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTLP</td>
<td>0.1189</td>
<td>0.0067</td>
<td>13.71</td>
<td>16.23</td>
<td>97.93%</td>
<td>18.87%</td>
<td>2.03</td>
</tr>
<tr>
<td>$CL^*_1$</td>
<td>0.1328</td>
<td>0.0082</td>
<td>13.66</td>
<td>16.66</td>
<td>97.57%</td>
<td>19.37%</td>
<td>0.63</td>
</tr>
<tr>
<td>TLP</td>
<td>0.2726</td>
<td>0.0229</td>
<td>12.94</td>
<td>15.04</td>
<td>92.43%</td>
<td>17.49%</td>
<td>0.10</td>
</tr>
<tr>
<td>ETE</td>
<td>0.1600</td>
<td>0.0100</td>
<td>13.62</td>
<td>16.03</td>
<td>97.29%</td>
<td>18.64%</td>
<td>0.05</td>
</tr>
<tr>
<td>HETE</td>
<td>0.1625</td>
<td>0.0104</td>
<td>13.57</td>
<td>16</td>
<td>96.93%</td>
<td>18.60%</td>
<td>0.32</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Method</th>
<th>ME</th>
<th>SE</th>
<th>TP</th>
<th>FP</th>
<th>PSR</th>
<th>NSR</th>
<th>Runtime (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTLP</td>
<td>1.0246</td>
<td>0.0526</td>
<td>10.7</td>
<td>11.09</td>
<td>76.43%</td>
<td>12.90%</td>
<td>2.24</td>
</tr>
<tr>
<td>$CL^*_1$</td>
<td>0.8340</td>
<td>0.0472</td>
<td>11.36</td>
<td>15.41</td>
<td>81.14%</td>
<td>17.92%</td>
<td>0.62</td>
</tr>
<tr>
<td>TLP</td>
<td>1.9927</td>
<td>0.0993</td>
<td>6.72</td>
<td>4.12</td>
<td>48.00%</td>
<td>4.79%</td>
<td>0.05</td>
</tr>
<tr>
<td>ETE</td>
<td>1.0223</td>
<td>0.0412</td>
<td>10.88</td>
<td>14.78</td>
<td>77.71%</td>
<td>17.19%</td>
<td>0.03</td>
</tr>
<tr>
<td>HETE</td>
<td>1.0964</td>
<td>0.0404</td>
<td>10.53</td>
<td>13.94</td>
<td>75.21%</td>
<td>16.21%</td>
<td>0.39</td>
</tr>
</tbody>
</table>

7. Real Data Results

Now we are to apply the proposed methodology to sparse index tracking for the CSI 300 Index. In this application, we use the daily return series of the CSI 300 Index and all stocks in this index from 2014 to 2018. The CSI 300 is a capitalization-weighted stock market index designed to replicate the performance of the top 300 stocks traded in the Shanghai and Shenzhen stock exchanges. The index is compiled by the China Securities Index and is considered a blue chip index for Mainland China stock exchanges. The return series, the methodology of the CSI 300 Index, the names and weights of the index constituents, and the corresponding sectors are available from China Securities Index (http://www.csindex.com.cn/) (accessed on 31 September 2018).

According to the guidelines for the industry classification by the Global Industry Classification Standard (GICS) and the weights of constituents making up the CSI 300 Index, the stocks can be divided into 11 major sectors, including Materials, Communication services, Real estate, Industrials, Utilities, Financials, Consumer Discretionary, Energy, Consumer Staples, Information Technology and Health Care. In the composition of the CSI 300 index, the constituents, as well as their weights, are reviewed every six months. As a consequence, every six months, the stock numbers and the weights of sectors will change. Due to its dynamic properties, we must update our model periodically. We train and tune the model on the first day after constituent adjustment is implemented with the daily return series of the index and its renewed constituents before half a year. The total weights within sectors are calculated based on the renewed weights as well. Then the model will be tested with daily return series before the next adjustment day. These steps are in accordance with the practical procedure.

We treat the estimated coefficients as the weights of stocks in building an index-tracking portfolio. The daily tracking error or prediction error is adopted to evaluate the model performance. The daily tracking error measures the deviation of the index daily return from the portfolio daily return, defined as $error = \hat{r}_t - r_t$, where $r_t$ is the daily return of the index, $\hat{r}_t$ is the daily return of the constructed portfolio. We will compare our method CTLP with TLP, ETE and HETE as described in Section 6.1 in terms of tracking errors. Since the training sample size is smaller than the number of index constituents, there is no exact solution without sparsity, but an approximate solution by a common convex optimizer. Thus, we show the $CL^*_1$ solution as in Section 6.2. For an additional comparison in the Lasso family, we give the performance of standard sparse group lasso [33] by SGL. To give a more intuitive view of the advantages of the proposed method, we also present the performance of portfolios created by traditional methods, including equal-weighted portfolio (EW) and inverse volatility weighted portfolio (IVW). Both portfolios consist of 300 stocks. The weights in the EW portfolio are equal while the weights in the IVW portfolio are inversely proportional to their historical volatility.
The CSI 300 index has been adjusted eight times from December 2014 to November 2018. Each time the constituents of the index were adjusted, we update the model and test it with the return series for the next six months. Table 8 presents the mean and standard deviation of the daily tracking errors in each test set. It also shows the number of stocks building the portfolio. We highlight the method with the smallest mean daily tracking error in each period. It is clear that in a majority of periods, CTLP has the smallest mean tracking error among its competitors. Even in the period, such as 2017H1 and 2017H2, when CTLP is not the best, it is still very close to the best one. In terms of the standard deviation of tracking errors, CTLP has a slightly larger standard deviation than ETE and HETE, but significantly smaller than TLP. However, TLP and HETE tend to yield sparser portfolios than CTLP and ETE. As to tracking errors, $CL_1^*$ performs not bad. But it has difficulties to give a sparse portfolio in several periods which violates our motivation to propose this research. The standard sparse group lasso (SGL) may provide negative coefficients but satisfy sector-neutral constraints. Although it has sparse enough solutions, its tracking errors are the largest. The traditional EW and IVW portfolios are not sparse and yield unstable performance. In some periods they track the index tightly, while in other periods they produce large tracking errors. We also present the average runtime of the first four methods in Table 9, indicating that CTLP and HETE run more slowly than the other four methods. Even though the computation of CTLP requires longer computation time due to additional sector neutral constraints, it is still worthwhile for its competitive performance and the fact that this computation is not a frequent task.

As illustration, we display analysis results for the most recent two periods – the first and second half of 2018, in Figures 1 and 2, Tables 10 and 11. Figures 1 and 2 draw the cumulative profit and loss of CTLP, TLP, ETE, HETE and the index. The closer of the cumulative profit and loss line to the red line, which represents that of the index, the better the method. It is clear that the CTLP replicates the index the best among all four methods. Tables 10 and 11 present the number of stocks and the total weights of different sectors. Compared with the index, all portfolios built by five methods are sparse. However, the portfolio given by TLP consists of such a small number of stocks that it cannot track the index well. In the TLP portfolio, some sectors even have zero weights. CTLP and $CL_1^*$ selected a portfolio with sector weights strictly equal to that of the index because of the sector neutrality constraint, whereas ETE and HETE produced sparse portfolios with sector weights only roughly equal to that of the index.

![Figure 1](image1.png)

Figure 1. The cumulative profit and loss of portfolios built by different methods and the index in 2018H1.
In summary, the proposed CTLP method demonstrates its advantages over the competitive methods in sparse index tracking. The sector neutrality constraint of CTLP guarantees the resulted sparse portfolio has the same sector risk exposure as the index. Most of the time, the CTLP method also gives smaller tracking errors. In addition, the non-negativity and sparsity of the CTLP portfolios are often desired properties in practical applications.

Table 8. The performance of the selected portfolio. By different methods in eight periods. Mean, Std. Dev and Num stand for the mean, the standard deviation of tracking error and the number of stocks in the portfolio. In the columns of period, for example, 2018H1 stands for the first half year of 2018.

<table>
<thead>
<tr>
<th>Period</th>
<th>Method</th>
<th>Mean</th>
<th>Std. Dev</th>
<th>Num</th>
<th>Period</th>
<th>Method</th>
<th>Mean</th>
<th>Std. Dev</th>
<th>Num</th>
</tr>
</thead>
<tbody>
<tr>
<td>2018H2</td>
<td>CTLP</td>
<td>0.018%</td>
<td>0.0016</td>
<td>124</td>
<td>2016H2</td>
<td>CTLP</td>
<td>0.015%</td>
<td>0.0013</td>
<td>107</td>
</tr>
<tr>
<td></td>
<td>CL₁⁺</td>
<td>0.028%</td>
<td>0.0016</td>
<td>297</td>
<td></td>
<td>CL₁⁺</td>
<td>0.027%</td>
<td>0.0007</td>
<td>300</td>
</tr>
<tr>
<td></td>
<td>SGL</td>
<td>0.123%</td>
<td>0.0126</td>
<td>57</td>
<td></td>
<td>SGL</td>
<td>−0.073%</td>
<td>0.0056</td>
<td>65</td>
</tr>
<tr>
<td></td>
<td>TLP</td>
<td>0.081%</td>
<td>0.0046</td>
<td>42</td>
<td></td>
<td>TLP</td>
<td>0.020%</td>
<td>0.0020</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td>ETE</td>
<td>0.031%</td>
<td>0.0013</td>
<td>188</td>
<td></td>
<td>ETE</td>
<td>0.023%</td>
<td>0.0009</td>
<td>219</td>
</tr>
<tr>
<td></td>
<td>HETE</td>
<td>0.026%</td>
<td>0.0013</td>
<td>98</td>
<td></td>
<td>HETE</td>
<td>0.016%</td>
<td>0.0011</td>
<td>128</td>
</tr>
<tr>
<td></td>
<td>EW</td>
<td>0.091%</td>
<td>0.0064</td>
<td>300</td>
<td></td>
<td>EW</td>
<td>−0.016%</td>
<td>0.0022</td>
<td>300</td>
</tr>
<tr>
<td></td>
<td>IVW</td>
<td>0.087%</td>
<td>0.0066</td>
<td>300</td>
<td></td>
<td>IVW</td>
<td>0.055%</td>
<td>0.0060</td>
<td>300</td>
</tr>
<tr>
<td>2018H1</td>
<td>CTLP</td>
<td>0.008%</td>
<td>0.0031</td>
<td>116</td>
<td></td>
<td>CTLP</td>
<td>0.002%</td>
<td>0.0022</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td>CL₁⁺</td>
<td>0.014%</td>
<td>0.0017</td>
<td>126</td>
<td></td>
<td>CL₁⁺</td>
<td>0.025%</td>
<td>0.0021</td>
<td>120</td>
</tr>
<tr>
<td></td>
<td>SGL</td>
<td>0.043%</td>
<td>0.0093</td>
<td>57</td>
<td></td>
<td>SGL</td>
<td>0.062%</td>
<td>0.0109</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>TLP</td>
<td>0.024%</td>
<td>0.0051</td>
<td>34</td>
<td></td>
<td>TLP</td>
<td>0.031%</td>
<td>0.0033</td>
<td>49</td>
</tr>
<tr>
<td></td>
<td>ETE</td>
<td>0.022%</td>
<td>0.0013</td>
<td>129</td>
<td></td>
<td>ETE</td>
<td>0.031%</td>
<td>0.0018</td>
<td>140</td>
</tr>
<tr>
<td></td>
<td>HETE</td>
<td>0.024%</td>
<td>0.0013</td>
<td>79</td>
<td></td>
<td>HETE</td>
<td>0.033%</td>
<td>0.0017</td>
<td>121</td>
</tr>
<tr>
<td></td>
<td>EW</td>
<td>0.018%</td>
<td>0.0086</td>
<td>300</td>
<td></td>
<td>EW</td>
<td>0.052%</td>
<td>0.0040</td>
<td>300</td>
</tr>
<tr>
<td></td>
<td>IVW</td>
<td>0.027%</td>
<td>0.0079</td>
<td>300</td>
<td></td>
<td>IVW</td>
<td>0.052%</td>
<td>0.0035</td>
<td>300</td>
</tr>
<tr>
<td>2017H2</td>
<td>CTLP</td>
<td>−0.008%</td>
<td>0.0019</td>
<td>123</td>
<td></td>
<td>CTLP</td>
<td>0.055%</td>
<td>0.0034</td>
<td>117</td>
</tr>
<tr>
<td></td>
<td>CL₁⁺</td>
<td>−0.005%</td>
<td>0.0015</td>
<td>296</td>
<td></td>
<td>CL₁⁺</td>
<td>0.064%</td>
<td>0.0021</td>
<td>130</td>
</tr>
<tr>
<td></td>
<td>SGL</td>
<td>−0.095%</td>
<td>0.0060</td>
<td>35</td>
<td></td>
<td>SGL</td>
<td>0.195%</td>
<td>0.0208</td>
<td>84</td>
</tr>
<tr>
<td></td>
<td>TLP</td>
<td>−0.040%</td>
<td>0.0035</td>
<td>25</td>
<td></td>
<td>TLP</td>
<td>0.135%</td>
<td>0.0087</td>
<td>51</td>
</tr>
<tr>
<td></td>
<td>ETE</td>
<td>−0.012%</td>
<td>0.0013</td>
<td>220</td>
<td></td>
<td>ETE</td>
<td>0.064%</td>
<td>0.0017</td>
<td>179</td>
</tr>
<tr>
<td></td>
<td>HETE</td>
<td>−0.017%</td>
<td>0.0016</td>
<td>84</td>
<td></td>
<td>HETE</td>
<td>0.062%</td>
<td>0.0019</td>
<td>121</td>
</tr>
<tr>
<td></td>
<td>EW</td>
<td>−0.015%</td>
<td>0.0048</td>
<td>300</td>
<td></td>
<td>EW</td>
<td>0.076%</td>
<td>0.0049</td>
<td>300</td>
</tr>
<tr>
<td></td>
<td>IVW</td>
<td>0.020%</td>
<td>0.0040</td>
<td>300</td>
<td></td>
<td>IVW</td>
<td>−0.935%</td>
<td>0.0479</td>
<td>300</td>
</tr>
<tr>
<td>2017H1</td>
<td>CTLP</td>
<td>−0.014%</td>
<td>0.0012</td>
<td>129</td>
<td></td>
<td>CTLP</td>
<td>0.060%</td>
<td>0.0032</td>
<td>120</td>
</tr>
<tr>
<td></td>
<td>CL₁⁺</td>
<td>−0.016%</td>
<td>0.0009</td>
<td>299</td>
<td></td>
<td>CL₁⁺</td>
<td>−0.011%</td>
<td>0.0020</td>
<td>147</td>
</tr>
<tr>
<td></td>
<td>SGL</td>
<td>0.050%</td>
<td>0.0058</td>
<td>11</td>
<td></td>
<td>SGL</td>
<td>−0.358%</td>
<td>0.0153</td>
<td>58</td>
</tr>
<tr>
<td></td>
<td>TLP</td>
<td>−0.024%</td>
<td>0.0032</td>
<td>33</td>
<td></td>
<td>TLP</td>
<td>−0.166%</td>
<td>0.0075</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>ETE</td>
<td>−0.010%</td>
<td>0.0009</td>
<td>177</td>
<td></td>
<td>ETE</td>
<td>0.004%</td>
<td>0.0018</td>
<td>166</td>
</tr>
<tr>
<td></td>
<td>HETE</td>
<td>−0.011%</td>
<td>0.0012</td>
<td>83</td>
<td></td>
<td>HETE</td>
<td>−0.009%</td>
<td>0.0024</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>EW</td>
<td>0.017%</td>
<td>0.0023</td>
<td>300</td>
<td></td>
<td>EW</td>
<td>0.009%</td>
<td>0.0041</td>
<td>300</td>
</tr>
<tr>
<td></td>
<td>IVW</td>
<td>0.015%</td>
<td>0.0017</td>
<td>300</td>
<td></td>
<td>IVW</td>
<td>0.045%</td>
<td>0.0176</td>
<td>300</td>
</tr>
</tbody>
</table>

Bold means the method has the smallest mean daily tracking error in this period.

Table 9. The average runtime of different methods.

<table>
<thead>
<tr>
<th></th>
<th>CTLP</th>
<th>CL₁⁺</th>
<th>SGL</th>
<th>TLP</th>
<th>ETE</th>
<th>HETE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Runtime (s)</td>
<td>47.60</td>
<td>0.75</td>
<td>0.14</td>
<td>18.48</td>
<td>1.67</td>
<td>46.78</td>
</tr>
</tbody>
</table>
Figure 2. The cumulative profit and loss of portfolios built by different methods and the index in 2018H2.

Table 10. The number of stocks and the sum of weights within sectors in selected portfolio by different methods and the index in 2018H1.

<table>
<thead>
<tr>
<th>Sector</th>
<th>Number of Stocks</th>
<th>Sum of Weights</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Index</td>
<td>CTLP</td>
</tr>
<tr>
<td>Materials</td>
<td>25</td>
<td>11</td>
</tr>
<tr>
<td>Communication</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Real Estate</td>
<td>20</td>
<td>6</td>
</tr>
<tr>
<td>Industrials</td>
<td>60</td>
<td>14</td>
</tr>
<tr>
<td>Utilities</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td>Financials</td>
<td>58</td>
<td>22</td>
</tr>
<tr>
<td>Consumer dis.</td>
<td>42</td>
<td>12</td>
</tr>
<tr>
<td>Energy</td>
<td>11</td>
<td>3</td>
</tr>
<tr>
<td>Consumer sta.</td>
<td>14</td>
<td>5</td>
</tr>
<tr>
<td>Information</td>
<td>42</td>
<td>16</td>
</tr>
<tr>
<td>Health Care</td>
<td>18</td>
<td>8</td>
</tr>
</tbody>
</table>

Table 11. The number of stocks and the sum of weights within sectors in selected portfolio by different methods and the index in 2018H2.

<table>
<thead>
<tr>
<th>Sector</th>
<th>Number of Stocks</th>
<th>Sum of Weights</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Index</td>
<td>CTLP</td>
</tr>
<tr>
<td>Materials</td>
<td>36</td>
<td>9</td>
</tr>
<tr>
<td>Communications</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Real Estate</td>
<td>17</td>
<td>5</td>
</tr>
<tr>
<td>Industrials</td>
<td>57</td>
<td>14</td>
</tr>
<tr>
<td>Utilities</td>
<td>10</td>
<td>4</td>
</tr>
<tr>
<td>Financials</td>
<td>59</td>
<td>24</td>
</tr>
<tr>
<td>Consumer dis.</td>
<td>37</td>
<td>14</td>
</tr>
<tr>
<td>Energy</td>
<td>13</td>
<td>2</td>
</tr>
<tr>
<td>Consumer sta.</td>
<td>13</td>
<td>8</td>
</tr>
<tr>
<td>Information</td>
<td>34</td>
<td>10</td>
</tr>
<tr>
<td>Health Care</td>
<td>22</td>
<td>10</td>
</tr>
</tbody>
</table>
8. Summary

Motivated by a sparse index tracking problem, we propose a new method to do sparse variable selection under constraints. Our methodology extends the traditional variable selection with added constraints. Constraints either represent the lower dimensional structure of the data or special characteristics of practical applications.

In the sparse index tracking problem, sparsity, sector neutrality, and nonnegativity constraints are necessary to build an efficient, sector-risk neutral portfolio with lower transaction costs to track the performance of the index. We proved the consistency and asymptotic distribution for the constrained high-dimensional variable selection using our method. We also developed an efficient algorithm for the estimation of the stock weights of the sparse portfolio. Both simulations and the real application confirmed the validity and advantages of the new methodology.

In portfolio management applications, additional constraints may be incorporated into index tracking, for example, low volatility or size neutrality constraints. We leave these problems for future investigations.

Author Contributions: Conceptualization, Y.C.; Data curation, Y.C.; Investigation, Y.C.; Methodology, S.C.; Software, S.C.; Supervision, X.L.; Validation, S.C.; Writing—original draft, Y.C.; Writing—review & editing, X.L. All authors have read and agreed to the published version of the manuscript.

Funding: Liu’s research is partially sponsored by the Shanghai Pujiang Program (21PJC056) and Shanghai Institute of International Finance and Economics.

Data Availability Statement: Publicly available datasets were analyzed in this study. This data can be found here: http://www.csindex.com.cn/.

Acknowledgments: The authors would like to thank the editor and anonymous referees for their helpful comments and suggestions. Our appreciation goes to Mathematics for consideration of publishing our work.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Appendix A

Appendix A.1. Proof of Theorems

Proof of Theorem 1. For \((i, j) \in \mathcal{A}^*\), by the definition of \(\hat{\beta}_{i,j}^{\text{ols}}\), the variance of \(\hat{\beta}_{i,j}^{\text{ols}}\) is a diagonal entry of matrix \(\sigma^2(\mathbf{X}_A^T \mathbf{X}_A)^{-1}\). Therefore,

\[
\Pr[\hat{\beta}_{i_j}^{\text{ols}} \leq \tau] = \Pr[n^{1/2}(\hat{\beta}_{i,j}^{\text{ols}} - \beta_{i,j}^*) \leq -n^{1/2}(\beta_{i,j}^* - \tau)] \\
\leq \Phi \left(-\frac{n^{1/2}(\gamma_{\min} - \tau)}{\sigma \eta_{\min}^{-1/2} \left(1 \mathbf{X}_A^T \mathbf{X}_A\right)}\right).
\]

The variance of \(\hat{\beta}_{i_l}^{\text{ols}}\) is \(\sigma^2 v_i^T (\mathbf{X}_A^T \mathbf{X}_A)^{-1} v_i\), where \(v_i\) is the coefficient vector for the linear combination of \(\hat{\beta}_{i,l}^*\) on the right hand side of Equation (5) using only nonzero values of \(\beta^*\). Note that \(v_i\) is a \(p_0\)-dimensional vector consisting of 0’s and –1’s only, The variance of \(\hat{\beta}_{i,l}^{\text{ols}}\) is bounded above by \(p_0 \sigma^2 \eta_{\min}^{-1} (\mathbf{X}_A^T \mathbf{X}_A)\). Thus,

\[
\Pr[\hat{\beta}_{i_l}^{\text{ols}} \leq \tau] = \Pr[n^{1/2}(\hat{\beta}_{i_l}^{\text{ols}} - \beta_{i,l}^*) \leq -n^{1/2}(\beta_{i,l}^* - \tau)] \\
\leq \Phi \left(-\frac{n^{1/2}(\gamma_{\min} - \tau)}{p_0 \sigma \eta_{\min}^{-1/2} \left(1 \mathbf{X}_A^T \mathbf{X}_A\right)}\right).
\]

It follows that
which completes the proof. □

**Proof of Theorem 2.** We firstly show that \( P[\hat{\beta}^{ols} \neq \hat{\beta}^{lip}] \to 0 \) as \( n \to \infty \). From Mill’s ratio [35] it follows that

\[
\Phi(-x) \leq \frac{1}{\sqrt{2\pi}} \frac{1}{x} e^{-\frac{x^2}{2}}, \quad x > 0.
\]

Hence,

\[
\Phi\left( -\frac{n^{1/2}(\gamma_{\text{min}} - \tau)}{\sigma \eta^{-1/2}_{\text{min}} \left( \frac{1}{n} X_{A}^{T}X_{A} \right)} \right) \leq \frac{1}{\sqrt{2\pi} n^{1/2} d_{1}^{1/2}(\gamma_{\text{min}} - \tau)} \exp\left( -\frac{nd_{1}(\gamma_{\text{min}} - \tau)^{2}}{2\sigma^{2}} \right),
\]

and

\[
\Phi\left( -\frac{n^{1/2}(\gamma_{\text{min}} - \tau)}{p_{0}^{1/2} \sigma \eta^{-1/2}_{\text{min}} \left( \frac{1}{n} X_{A}^{T}X_{A} \right)} \right) \leq \frac{1}{\sqrt{2\pi} n^{1/2} d_{1}^{1/2}(\gamma_{\text{min}} - \tau)} \exp\left( -\frac{nd_{1}(\gamma_{\text{min}} - \tau)^{2}}{2p_{0}\sigma^{2}} \right).
\]

Since \( \gamma_{\text{min}} \approx n^{-1/2} \), \( \lambda \approx n^{3/2} \), \( p_{0} \approx n^{p_{0}} \), and \( a_{\gamma} > a_{\lambda} > a_{p_{0}} \geq 0 \), it follows that
Mathematics 2022, 10, 2645

[36x779]Mathematics
[36x779]2022
[95x779]10
[95x779],
[166x288]TLP function. This is a non-convex minimization which can be solved by the difference
[166x438]and
[166x238]minimization subproblem (11) is equivalent to minimize (A5) with regard to
[166x238]convex optimization can be applied. By DC decomposition and some calculations, the
[166x263]function into the difference of convex functions so that the algorithms and properties of
[166x316]Appendix A.2. The Details of Algorithm
[166x341]consistency in part (A) we obtain the asymptotic distribution (7).

\[
\max \left\{ \frac{\sqrt{2}p_0n^{1/2}}{\sigma \sqrt{\eta_{\min}^{-1}}} \exp \left( - \frac{n(\gamma_{\min} - \tau)^2}{2\sigma^2 \eta_{\min}^{-1}} \right), \right. \\
p_0 \Phi \left( \frac{-n^{1/2}(\gamma_{\min} - \tau)}{\sigma \eta_{\min}^{-1/2}} \left( \frac{1}{\eta_{\min}^{-1}} \right)^{1/2} \right) \right\} \\
\leq \max \left\{ \frac{\sqrt{2}p_0n^{1/2}d_2^{1/2}}{\sigma \sqrt{\pi}}, \frac{p_0^{3/2} \sigma}{\sqrt{2\pi n^{1/2}d_1^{1/2}(\gamma_{\min} - \tau)}} \right\} \exp \left( - \frac{nd_1(\gamma_{\min} - \tau)^2}{2p_0\sigma^2} \right) \rightarrow 0.
\]

Similarly,

\[
q \Phi \left( \frac{-n^{1/2}(\gamma_{\min} - \tau)}{p_0^{1/2} \sigma \eta_{\min}^{-1} \left( \frac{1}{\eta_{\min}^{-1}} \right)^{1/2}} \right) \\
\leq \frac{p_0}{\sqrt{2\pi} n^{1/2}d_1^{1/2}(\gamma_{\min} - \tau)} \exp \left( - \frac{nd_1(\gamma_{\min} - \tau)^2}{2p_0\sigma^2} \right) \rightarrow 0.
\]

Because \( a_p - 1 < a_A < a_C \) and \( a_{ip} \geq 0 \), it follows that

\[
\left( \frac{nC_{\min}}{20\sigma^2} - (a + 1) \log(p + 1) - \frac{n\lambda}{4\sigma^2} \right) \rightarrow \infty,
\]  
(A3)

and

\[
\left( \frac{(a - 1)n\lambda}{6a\sigma^2} - \left( 1 + \frac{1}{a} \right) \left( \log(p + 1) - \frac{5}{3} \right) \right) \rightarrow \infty.
\]  
(A4)

Combining the limits (A1)–(A4), we obtain the selection consistency from Theorem 1.

**Asymptotic properties:** From Equation (6) we know that \( \hat{\beta}_{ols}^{\beta_{ols}} \) has a multivariate normal distribution with mean \( \beta_{ols}^* \) and covariance matrix \( \frac{\sigma^2}{n} \left( \frac{1}{n} X_0 X_0^\top \right)^{-1} \). By the selection consistency in part (A) we obtain the asymptotic distribution (7).

**Appendix A.2. The Details of Algorithm**

The objective function in the first minimization step (11) is a quadratic function plus a TLP function. This is a non-convex minimization which can be solved by the difference of the convex (DC) method [17,23]. The DC method is used to decompose a non-convex function into the difference of convex functions so that the algorithms and properties of convex optimization can be applied. By DC decomposition and some calculations, the minimization subproblem (11) is equivalent to minimize (A5) with regard to \( \beta \),

\[
h^{(m)}(\beta) = L(\beta) + \frac{\lambda}{\tau} \sum_{i=1}^{q} \sum_{j=1}^{g_i} |\beta_{ij}| I(|\beta_{ij}|^{(m-1)} \leq \tau).
\]  
(A5)

Let us begin with the DC decomposition of the objective function \( h(\beta) = h_1(\beta) - h_2(\beta) \), where

\[
h_1(\beta) = L(\beta) + \lambda \sum_{j=1}^{p} I_1(|\beta_{ij}|), \quad h_2(\beta) = \lambda \sum_{i=1}^{q} \sum_{j=1}^{g_i} I_2(|\beta_{ij}|)
\]

\[
L(\beta) = n^{-1} \| Y - X\beta \|_2^2 + \frac{\rho}{2} \| \beta - \alpha^k + \kappa^k \|_2^2
\]

\[
I_1(|\beta_{ij}|) = \frac{|\beta_{ij}|}{\tau}, \quad I_2(|\beta_{ij}|) = \max \{ \frac{|\beta_{ij}|}{\tau} - 1, 0 \}.
\]
It is clear that \( L(\beta) \) is a convex function of \( \beta \). Using the DC decomposition, a sequence of linear approximations \( h^{(m)}(\beta) \) of \( h(\beta) \) is constructed. Let \( \nabla h_2 \) be the subgradient of \( h_2 \). At the \( m \)-th iteration, replacing \( h_2(\beta) \) by its majorization, we obtain

\[
h^{(m)}(\beta) = h_1(\beta) - (h_2(\beta^{(m-1)}) + (|\beta| - |\beta^{(m-1)}|)^\top \nabla h_2(|\beta^{(m-1)}|)),
\]

where \( \beta \) denotes the vector obtained by replacing each component of \( \beta \) with its absolute value. After ignoring the terms that are independent of \( \beta \), the \( h^{(m)}(\beta) \) becomes

\[
h^{(m)}(\beta) = L(\beta) + \frac{\lambda}{\tau} \sum_{i=1}^{q} \sum_{j=1}^{n} \beta_{ij} |\beta_{ij}^{(m-1)}| \leq \tau.
\]

Minimizing (A6) gives the updated value \( \hat{\beta}^{(m)} \). Repeat the process procedure until convergence.

Then, for the minimizing of (A6), we can use the adaptive weights \( \lambda_{ij} = \frac{1}{\tau} \mathbb{1}(|\hat{\beta}_{ij}^{(m-1)}| \leq \tau) \), (A6) can be rewritten as

\[
h^{(m)}(\beta) = L(\beta) + \sum_{i=1}^{q} \sum_{j=1}^{n} \lambda_{ij} |\beta_{ij}|.
\]

Note that the second term in (A7) is a separable function in variable \( \beta \). This property enables us to use coordinate descent algorithm [24,25] for minimization. Similar to the results of regular Lasso [36], the updating formula for \( \beta \) is

\[
\beta_{ij} \leftarrow \left( \frac{2}{n} X^T_{ij} X_{ij} + \rho \right)^{-1} S(Z_{ij}, \lambda_{ij}),
\]

where \( Z_{ij} = \frac{2}{n} X^T_{ij} r_{ij} - \rho (\omega_{ij} + \mu_{ij}) \), \( r_{ij} = Y - X_{-(ij)} \hat{\beta}_{-(ij)} \), \( X_{ij} \) is the \( j \)-th column of \( X \), \( X_{-(ij)} \) is the matrix \( X \) with the \( j \)-th column of \( X \) deleted, and \( \beta_{-(ij)} \) is defined similarly. \( S(z, \lambda) = \text{Sign}(z)(|z| - \lambda) \) is the soft-thresholding operator. Updating Formula (A8) is cycled through all variables in turn. Repeated iteration of (A8) until convergence gives the estimate \( \hat{\beta}^{(m)} \).

Finally, with the inner iteration (A8) and outer iteration Algorithm A1, we solve the minimization subproblem (11).

We summarize the DC algorithm for minimization (11) below.

**Algorithm A1** DC algorithm for the minimization of (11).

- (Initialization) Use \( \hat{\beta}^k \) in \( k \)-th ADMM iteration as the initial estimate \( \hat{\beta}^{(0)} \).
- (Iteration) At \( m \)-th DC iteration, compute \( \hat{\beta}^{(m)} \) by minimizing (A5).
- (Termination) Stop when \( |h(\hat{\beta}^{(m-1)}) - h(\hat{\beta}^{(m)})| < \epsilon \) and no components of \( \hat{\beta}^{(m)} \) is at \( \pm \tau \).

Next, we note that computing the projection (12) directly is not easy. Let \( \mathcal{C}_1 = \{ \beta : \sum_{i=1}^{q} \beta_i = \omega, i = 1, \ldots, q \} \). We divide the projection into two easier sequential projections: the first one is the projection from \( \mathbb{R}^p \) onto space \( \mathcal{C}_1 \), and the second one is the projection from \( \mathcal{C}_1 \) to \( \mathcal{C} \). The Theorem A1 below will guarantee the equivalence between the direct projection on \( \mathcal{C} \) and the two sequential projections.

**Theorem A1.** In Euclidean space, projection onto space \( \mathcal{C} \) \( \beta \overset{\Pi_{\mathcal{C}}}{\longrightarrow} \beta_2 \) is equivalent to the sequential projections onto space \( \mathcal{C}_1 \) and space \( \mathcal{C} \):

\[
\beta \overset{\Pi_{\mathcal{C}_1}}{\longrightarrow} \beta_1 \overset{\Pi_{\mathcal{C}}}{\longrightarrow} \beta_2.
\]

There exists a closed-form solution for \( \beta_1 = \Pi_{\mathcal{C}_1} \beta \).
Let \( \beta_1 = (\beta_{11}, \ldots, \beta_{1p})^\top \). Assume there are \( m_i \) elements in the \( i \)-th group \( g_i \). We have

\[
\beta_{ij} = \beta_j + \omega_i - \sum_{j=1}^{g_i} \beta_j, \quad i \in \{1, \ldots, q\}, \quad j \in \{1, \ldots, g_i\}. \tag{A10}
\]

The computation of the second projection is given in the proof of Theorem A1. The composition of these two sequential projections updates the value of \( \alpha \). The third step (13) is a direct calculation.

**Proof of Theorem A1.** Without loss of generality, we assume that all variables belong to the same group. The proof is similar to the case when the number of groups for variables is more than 1. The parameter space \( C_1 \) becomes \( C_1 = \{ \beta : \sum_j \beta_j = 1 \} \).

Denote \( \beta_1 = \prod_{g} \beta \), and \( \beta_2 = \prod_{g} \beta_1 \), where \( \beta_1 = (\beta_{11}, \ldots, \beta_{1p})^\top \), and \( \beta_2 = (\beta_{21}, \ldots, \beta_{2g})^\top \). The closed-form formula for \( \prod_{C_1} \) is given by (A10).

Now we are to calculate the projection \( \prod_{C_0} \). We assume that \( \beta_1 \) has \( u \) positive elements, \( v \) zero elements, and \( p - u - v \) negative elements:

\[
\begin{align*}
\beta_{1j} &> 0 \quad 1 \leq j \leq u \\
\beta_{1j} &= 0 \quad u + 1 \leq j \leq u + v \\
\beta_{1j} &< 0 \quad u + v + 1 \leq j \leq p.
\end{align*}
\tag{A11}
\]

Let \( c_u = \frac{1 - \sum_{j=1}^{u} \beta_{1j}}{u} \). The projection \( \beta_2 = \prod_{C} \beta_1 \) is given by

\[
\begin{align*}
\beta_{2j} &= \begin{cases} 
\beta_{1j} + c_u & 1 \leq j \leq u \\
0 & u + v + 1 \leq j \leq p.
\end{cases} \tag{A12}
\end{align*}
\]

It is possible that some of \( \beta_{2j}, 1 \leq j \leq u \) are negative. If this is the case, we will repeat the above procedure (A11) and (A12) until there are no negative elements in \( \beta_2 \). This completes the proof. \( \Box \)

**Appendix A.3. The Details of Simulation**

In the simulation study, we consider the model \( y = X\beta + \sigma \epsilon \). The covariates \( X \) are generated from a multivariate normal distribution with all marginal distributions being standard normal distribution \( N(0, 1) \). The correlation matrix of \( X \) is denoted by \( R \).

We assume three settings for the correlation matrix \( R \):

1. Independence structure: \( R \) is the identity matrix;
2. AR(1) correlation without group structure: \( R = (r_{jl})_{j,l=1}^p \), where \( r_{jl} = 0.5^{|j-l|} \) for \( j, l = 1, \ldots, p \);
3. AR(1) correlation with group structure: \( R = (r_{jl})_{j,l=1}^p \), where \( r_{jl} = 0.5^{|j-l|} \) for \( j, l = 1, \ldots, g_i \) and \( r_{jl} = 0 \) for \( j = 1, \ldots, g_i \) and \( l = 1, \ldots, g_i', i \neq i' \).

As to the dimension of the covariates \( X \), two cases \( n < p \) and \( n > p \) will be considered. We use \( p = 100 \). The parameter vector \( \beta \) is divided into 5 groups, with 20 elements in each group. The numbers of non-zero elements in each group are 3, 3, 1, 2, and 5, respectively. The true values of \( \beta \) are

\[
\beta^0 = (0.12, 0.04, 0.05, 0.09, 0.02, 0.03, 0.17, 0.0, 0.0, 0.2, 0.12, 0.02, 0.04, 0.03, 0.02, 0.05, 0.0, 0.0, 0.18, 0.02, 0.04, 0.03, 0.02, 0.05, 0.0, 0.0, 0.15). \]

The sum of coefficients for each group are $\gamma_1 = 0.21$, $\gamma_2 = 0.14$, $\gamma_3 = 0.17$, $\gamma_4 = 0.32$, and $\gamma_5 = 0.16$, respectively. Besides, the numbers of zero elements in each group are 17, 17, 19, 18, and 15, respectively. The sample size $n$ is set to be 50 and 2000, for low- and high-dimensional cases. The distribution of error term $\epsilon$ is the normal distribution $N(0, 0.05^2)$, and the scaling parameter $\sigma$ takes two values: $\sigma = 1$ and $\sigma = 3$.

In each experiment, we randomly divide a dataset into training, tuning, and testing sets of 60%, 20%, and 20% of original sample sizes, respectively. We repeat the experiment 100 times and report the means of ME, TP, FP, PSR, NSR and the standard error of ME for each of the simulation settings.

As to tuning parameter selection, we will use five-fold cross-validation. The optimal choice of tuning parameters $\lambda$ and $\tau$ can be found by grid search [17]. Finding an optimal value for $\rho$ is not a straightforward problem [22], however, ref. [21] provides many heuristics that work well in practice. Here, we find out that a good choice of $\rho$ is around 1.

References


31. Benidis, K.; Palomar, D.P. `sparseIndexTracking: Design of Portfolio of Stocks to Tracks an Index`; R Package Version 0.1.0. Available online: https://CRAN.R-project.org/package=sparseIndexTracking (accessed on 1 September 2019).


