mathematics

MDPI

Article

High-Dimensional Regression Adjustment Estimation for Average Treatment Effect with Highly Correlated Covariates

Zeyu Diao [1], Lili Yue [2,*], Fanrong Zhao [3] and Gaorong Li [1]

[1] School of Statistics, Beijing Normal University, Beijing 100875, China
[2] School of Statistics and Data Science, Nanjing Audit University, Nanjing 211815, China
[3] School of Mathematical Science, Shanxi University, Taiyuan 030006, China
* Correspondence: yuell91@nau.edu.cn

Abstract: Regression adjustment is often used to estimate average treatment effect (ATE) in randomized experiments. Recently, some penalty-based regression adjustment methods have been proposed to handle the high-dimensional problem. However, these existing high-dimensional regression adjustment methods may fail to achieve satisfactory performance when the covariates are highly correlated. In this paper, we propose a novel adjustment estimation method for ATE by combining the semi-standard partial covariance (SPAC) and regression adjustment methods. Under some regularity conditions, the asymptotic normality of our proposed SPAC adjustment ATE estimator is shown. Some simulation studies and an analysis of HER2 breast cancer data are carried out to illustrate the advantage of our proposed SPAC adjustment method in addressing the highly correlated problem of the Rubin causal model.

Keywords: average treatment effect; highly correlated covariates; regression adjustment; rubin causal model; semi-standard partial covariance

MSC: 62D99; 62E20; 62J07

Citation: Diao, Z.; Yue, L.; Zhao, F.; Li, G. High-Dimensional Regression Adjustment Estimation for Average Treatment Effect with Highly Correlated Covariates. *Mathematics* **2022**, *10*, 4715. https://doi.org/10.3390/math10244715

Academic Editor: Davide Valenti

Received: 8 November 2022
Accepted: 10 December 2022
Published: 12 December 2022

1. Introduction

Accompanied by the rapid development in information technology, people have opportunities to collect a massive amount of data in many fields, such as genomics, biomedicine, aerography and so on, where the dimension of covariates p often far exceeds the sample size n. Despite the promising application prospects, there are many problems and challenges among statistical inference for high-dimensional data. For instance, the sample covariance matrix is huge and noninvertible under the setting $p > n$, the unimportant covariates are highly correlated with the response variable because they are associated with the important covariates ([1]). To deal with these problems and challenges, many penalty-based approaches have been proposed to select important covariates and estimate the unknown parameters simultaneously, including Lasso ([2]), SCAD ([3]) and Elastic-net ([4]) penalties. The above literatures mainly focus on considering the regression models and traditional correlations between the covariates and response variable.

In some cases, the traditional correlations cannot fully depict the influence mechanism of variables. Researchers have studied the causal relations among the variables and developed the Rubin causal (Neyman-Rubin) models (see [5,6]); details can be found in Refs. [7,8]. For the case of high-dimensional data, Refs. [9,10] suggested that standard high-dimensional penalty-based methods can be used to estimate the average treatment effect (ATE). Ref. [11] developed a risk-consistent regression adjustment approach for ATE using Lasso penalty in [2]. Ref. [12] proposed Lasso-adjusted ATE estimator by combining the Lasso penalty and regression adjustment method, and showed that the proposed method can reduce the variance of unadjusted ATE estimator in [6]. Ref. [13] further considered the multicollinearity problem in high dimensions, and proposed an Elastic-net adjustment method for ATE.

However, the correlations between the important and unimportant covariates are usually higher than those of the important covariates in high-dimensional settings (see [14,15]). Under this status, the irrepresentable condition ([16]) could fail such that the Lasso-based penalty methods may fail to correctly estimate the signs and distinguish the important and unimportant covariates. Then, the corresponding adjustment ATE estimator may perform poorly. So far, many research scholars have considered the highly correlated problem in high dimensions and provided some effective methods to undertake variable selection. For example, Ref. [17] proposed the Peter-Clark-simple (PC-simple) algorithm to select the important covariates using partial correlation, Ref. [18] developed the factor-adjusted regularized model selection (Farm-Select) method. Ref. [19] gave the semi-standard partial covariance (SPAC) method to effectively choose covariates which have a direct effect on the response variable, and showed that the SPAC outperforms the PC-simple and Farm-Select methods when the original irrepresentable condition in Ref. [16] fails. Nevertheless, these variable selection methods have not yet been used on the fields of causal inference.

In this paper, we consider the estimation problem of ATE in the Rubin causal model with highly correlated covariates. The main contributions of this paper are four-fold. Firstly, the SPAC adjustment estimator is developed by a novel combination of the SPAC variable selection and regression adjustment methods. Secondly, the framework is an extension of that in [19] to study the causal inference and [12] to handle the highly correlated problem. Thirdly, the theoretical property is shown under some regularity conditions. Fourthly, the performance of our proposed SPAC adjustment method is satisfactory, which can be observed by the numerical results of a real data analysis and some simulation studies.

The rest of this article is organized as follows. In Section 2, the SPAC adjustment method for ATE is proposed for the Rubin causal model with highly correlated covariates in high dimensions, and the asymptotic property of the proposed SPAC-Lasso adjustment estimator for ATE is also developed under some regularity conditions. In Section 3, some simulation studies are assigned to assess the effectiveness of our proposed SPAC adjustment method. In Section 4, the proposed estimation approach is applied to an HER2 breast cancer dataset. Some concluding discussions are provided in Section 5. The Appendix A is devoted to some Lemmas related to the proof of theorem.

Notation 1. *For the sake of description, some notations are introduced as follows. For any column vector* $\boldsymbol{u} = (u_1, \dots, u_p)^{\mathrm{T}}$ *and a subset* $S \subset \{1, \dots, p\}$, *let* $\|\boldsymbol{u}\|_1 = \sum_{j=1}^{p} |u_i|$, $\|\boldsymbol{u}\|_2^2 = \sum_{j=1}^{p} u_i^2$ *and* $\|\boldsymbol{u}\|_\infty = \max_{i=1,\dots,p} |u_i|$, $\boldsymbol{u}^S = \{u_j : j \in S\}$, S^C *denotes the complement of* S, $|S|$ *denote the cardinality of* S. *For a matrix* $\mathbf{D}$, $\mathbf{D}^{\mathrm{T}}$ *and* $\mathbf{D}^{-1}$ *denote the transpose and inverse of matrix* $\mathbf{D}$, *respectively. The notation "*$\xrightarrow{\mathcal{L}}$*" denotes the convergence in distribution.*

2. Methodology and Theoretical Property

2.1. Spac Adjustment Method for ATE

We frame our analysis in terms of the Rubin causal model. Let i be the units in the population of size n, Y_i be the potential outcome variable, $\boldsymbol{x}_i = (x_{i1}, \dots, x_{ip})^{\mathrm{T}} \in \mathbb{R}^p$ be the p-dimensional covariates with p far exceeding the sample size n, the full design matrix of the experiment be $\mathbf{X} = (\boldsymbol{x}_1, \dots, \boldsymbol{x}_n)^{\mathrm{T}}$, each covariate $\boldsymbol{X}_j = (x_{1j}, \dots, x_{nj})^{\mathrm{T}}$ $(j = 1, \dots, p)$ is standardized with $\boldsymbol{X}_j^{\mathrm{T}} \boldsymbol{X}_j = n$ and $\sum_{i=1}^{n} x_{ij} = 0$. The observed data $\boldsymbol{x}_i$ $(i = 1, \dots, n)$ can be viewed as independent identically distributed (i.i.d.) from a distribution with mean $\mathbf{0}$ and positive definite covariance matrix $\boldsymbol{\Sigma}_{p \times p}$, and all the diagonal elements of $\boldsymbol{\Sigma}$ are equal to 1. Each unit is randomly assigned to the treatment group or control group, and the treatment indicator is denoted by T_i with $T_i = 1$ for a treated individual and $T_i = 0$ otherwise. Then, the observed potential outcome for individual i is

$$Y_i^{\mathrm{obs}} = T_i Y_i(1) + (1 - T_i) Y_i(0),$$

where $Y_i(1)$ and $Y_i(0)$ are the corresponding potential outcomes under treatment and control groups, respectively, that is, $Y_i^{\text{obs}} = Y_i(1)$ for $T_i = 1$, $Y_i^{\text{obs}} = Y_i(0)$ for $T_i = 0$. The numbers of the treated and control units are equal to $n_A = |A|$ and $n_B = |B|$, respectively, with $A = \{i \in \{1, \ldots, n\} : T_i = 1\}$, $B = \{i \in \{1, \ldots, n\} : T_i = 0\}$, and $n_A + n_B = n$.

In randomized experiments, the sample is often not randomly taken from the population (superpopulation) of interest (see [12,13,20]). In this paper, we focus on ATE in the finite sample, which is defined as

$$\tau = \overline{Y}_1 - \overline{Y}_0, \tag{1}$$

where $\overline{Y}_1 = n^{-1}\sum_{i=1}^n Y_i(1)$ and $\overline{Y}_0 = n^{-1}\sum_{i=1}^n Y_i(0)$ are the average responses if all individuals receive treatment or not. Clearly, the averages of potential outcomes over the whole population $\overline{Y}_1$ and $\overline{Y}_0$ are fixed. Based on the idea of replacing the population averages $\overline{Y}_s$ $(s = 0, 1)$ with the sample averages, a nature unadjusted ATE estimator is obtained as follows,

$$\hat{\tau}_{\text{unadj}} = \frac{1}{n_A}\sum_{i \in A} Y_i(1) - \frac{1}{n_B}\sum_{i \in B} Y_i(0). \tag{2}$$

As pointed out by [12,21,22], the information of covariates $\boldsymbol{x}_i$ can often be used to adjust the estimator in (2) in hope of improving estimation precision. For the high-dimensional data, Ref. [12] proposed the following Lasso-adjusted ATE estimator

$$\hat{\tau}_{\text{Lasso}} = \left\{\overline{Y}_A - (\overline{\boldsymbol{x}}_A - \overline{\boldsymbol{x}})^{\mathrm{T}}\hat{\boldsymbol{\beta}}^A_{\text{Lasso}}\right\} - \left\{\overline{Y}_B - (\overline{\boldsymbol{x}}_B - \overline{\boldsymbol{x}})^{\mathrm{T}}\hat{\boldsymbol{\beta}}^B_{\text{Lasso}}\right\}, \tag{3}$$

where $\overline{Y}_A = n_A^{-1}\sum_{i \in A} Y_i(1)$, $\overline{Y}_B = n_B^{-1}\sum_{i \in B} Y_i(0)$, $\overline{\boldsymbol{x}}_A = n_A^{-1}\sum_{i \in A}\boldsymbol{x}_i$, $\overline{\boldsymbol{x}}_B = n_B^{-1}\sum_{i \in B}\boldsymbol{x}_i$, $\overline{\boldsymbol{x}} = n^{-1}\sum_{i=1}^n \boldsymbol{x}_i$, and the terms $\overline{\boldsymbol{x}}_w - \overline{\boldsymbol{x}}$ for $w = A$ and B illustrate the fluctuations between the subsample and full sample of covariates. The adjustment vectors $\hat{\boldsymbol{\beta}}^w_{\text{Lasso}}$ are obtained based on the Lasso penalty,

$$\hat{\boldsymbol{\beta}}^w_{\text{Lasso}} = \underset{\boldsymbol{\beta} \in \mathbb{R}^p}{\operatorname{argmin}}\left[\frac{1}{2n_w}\sum_{i \in w}\left\{Y_i^{\text{obs}} - \overline{Y}_w - (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_w)^{\mathrm{T}}\boldsymbol{\beta}\right\}^2 + \lambda_w\sum_{j=1}^p |\beta_j|\right], \qquad w = A, B, \tag{4}$$

where $\lambda_w > 0$ are regularization parameters for Lasso.

However, traditional penalty-based methods fail to effectively estimate the signs and select the important covariates when the important and unimportant covariates are highly correlated, see the details in [19]. This is especially critical in high-dimensional settings. To solve this problem, the SPAC method is proposed to capture the signal strengths of important covariates while eliminating the effects of covariates that are not directly related with the potential outcome variable Y but highly correlated with important covariates. The SPAC between $\boldsymbol{Y}^{\text{obs}}$ and the j-th covariate $\boldsymbol{X}_j$ is defined as

$$\gamma_j = \beta_j / d_{jj}^{1/2}, \qquad j = 1, \ldots, p, \tag{5}$$

where d_{jj} is the j-th diagonal element of precision matrix $\boldsymbol{\Sigma}^{-1}$, $1/d_{jj}^{1/2} = \{\text{Var}(\boldsymbol{X}_j | \boldsymbol{X}_{-j})\}^{1/2}$ $= (1 - R_j^2)^{1/2}$ (see Refs. [23,24]), where $\boldsymbol{X}_{-j} = \{\boldsymbol{X}_k : k = 1, ..., j-1, j+1, ..., p\}$, R_j denotes the multiple correlations between the j-th covariate $\boldsymbol{X}_j$ and all the other covariates. In particular, γ_j is the same as β_j if $\boldsymbol{X}_j$ is independent of the other covariates. Otherwise, the SPAC γ_j mitigates the effect of correlations among the covariates by using $(1 - R_j^2)^{1/2}$ to multiply β_j. Obviously, $\beta_j = 0$ if and only if $\gamma_j = 0$ for $j = 1, \ldots, p$. Hence, the SPAC estimator of adjustment vector can be obtained by replacing β_j in (4) with γ_j,

$$\hat{\gamma}^w_{\text{SPAC-Lasso}} = \underset{\gamma \in \mathbb{R}^p}{\operatorname{argmin}}\left[\frac{1}{2n_w}\sum_{i \in w}\left\{Y_i^{\text{obs}} - \overline{Y}_w - (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_w)^{\mathrm{T}}\hat{\mathbf{D}}\boldsymbol{\gamma}\right\}^2 + \lambda_w\sum_{j=1}^p \hat{d}_{jj}|\gamma_j|\right], \tag{6}$$

where $w = A, B$, $\widehat{\mathbf{D}} = \mathrm{diag}\{\hat{d}_{11}^{1/2}, \ldots, \hat{d}_{pp}^{1/2}\}$, $\hat{d}_{jj}$ is the consistent estimator of the j-th diagonal element of precision matrix. In detail, $\hat{d}_{jj}$ can be adopted as the constrained L_1-minimization estimation (CLIME, [25]), residual variance estimator ([26]), robust matrix estimator ([27]). Consequently, the adjustment vectors $\hat{\boldsymbol{\beta}}^{w}_{\mathrm{SPAC-Lasso}}$ can be given by using (5),

$$\hat{\boldsymbol{\beta}}^{w}_{\mathrm{SPAC-Lasso}} = \widehat{\mathbf{D}}\hat{\gamma}^{w}_{\mathrm{SPAC-Lasso}}, \qquad w = A, B. \tag{7}$$

Then, the SPAC-Lasso adjustment estimator of ATE is defined as

$$\hat{\tau}_{\mathrm{SPAC-Lasso}} = \left\{\overline{Y}_A - (\overline{\boldsymbol{x}}_A - \overline{\boldsymbol{x}})^{\mathrm{T}}\hat{\boldsymbol{\beta}}^{A}_{\mathrm{SPAC-Lasso}}\right\} - \left\{\overline{Y}_B - (\overline{\boldsymbol{x}}_B - \overline{\boldsymbol{x}})^{\mathrm{T}}\hat{\boldsymbol{\beta}}^{B}_{\mathrm{SPAC-Lasso}}\right\}. \tag{8}$$

Similarly, we can obtain the SPAC-SCAD estimator of ATE by using the SCAD penalty in (6). The performance of our proposed SPAC adjustment methods (SPAC-Lasso and SPAC-SCAD) will be compared with those of the existing ATE estimation methods (unadjusted, Lasso-adjusted, SCAD-adjusted, Elastic-net adjusted) in the following simulation studies, and the theoretical property of the SPAC-Lasso adjustment estimator will be shown in the following subsection.

2.2. Regularity Conditions and Theoretical Property

For Rubin causal model in randomized experiments, there are no assumptions for the relationship between potential outcome variable Y and covariates $\boldsymbol{x}$. To study the theoretical property of the proposed estimator $\hat{\tau}_{\mathrm{SPAC-Lasso}}$, we make the following linear decomposition and define the approximate sparsity, which are similar to that in [12].

Decomposition of the potential outcomes. The potential outcome can be divided into a linear term of covariates and an error term, which is formed as,

$$Y_i(1) = \overline{Y}_1 + (\boldsymbol{x}_i - \overline{\boldsymbol{x}})^{\mathrm{T}}\boldsymbol{\beta}^A + e_i^A, \quad Y_i(0) = \overline{Y}_0 + (\boldsymbol{x}_i - \overline{\boldsymbol{x}})^{\mathrm{T}}\boldsymbol{\beta}^B + e_i^B, \qquad i = 1, \ldots, n, \tag{9}$$

where $\overline{\boldsymbol{x}} = n^{-1}\sum_{i=1}^{n}\boldsymbol{x}_i$, $\boldsymbol{\beta}^A$ and $\boldsymbol{\beta}^B$ are p-dimensional vectors of coefficients. In the above decomposition (9), all the quantities are fixed and deterministic numbers, and $\overline{e^A} = \overline{e^B} = 0$, where $\overline{e^A} = n^{-1}\sum_{i=1}^{n} e_i^A$, $\overline{e^B} = n^{-1}\sum_{i=1}^{n} e_i^B$.

Definition 1. *Similar to [12,13], we define the approximate sparsity measures s_λ^A and s_λ^B for treatment and control groups as*

$$s_\lambda^A = \sum_{j=1}^{p}\min\left\{\left|\beta_j^A\right|\lambda_A^{-1}, 1\right\}, \quad s_\lambda^B = \sum_{j=1}^{p}\min\left\{\left|\beta_j^B\right|\lambda_B^{-1}, 1\right\},$$

which are more flexible than $s^w = |\{j : \beta_j^w \neq 0\}|$ with $w = A, B$. s_λ^A and s_λ^B are allowed to grow with n, $s_\lambda = \max\{s_\lambda^A, s_\lambda^B\}$.

In addition, the following regularity conditions are also needed to obtain the asymptotic normality of the proposed SPAC-Lasso adjustment estimation.

(C1) $\tilde{p}_A = n_A/n \to p_A$ and $\tilde{p}_B = n_B/n \to p_B$ as $n \to \infty$, and $p_A \in (0,1), p_B \in (0,1)$.
(C2) For $j = 1, \ldots, p$, there is a fixed constant $L > 0$ such that $n^{-1}\sum_{i=1}^{n}(x_{ij} - (\overline{\boldsymbol{x}})_j)^4 \leq L$, $n^{-1}\sum_{i=1}^{n}(e_i^A)^4 \leq L$ and $n^{-1}\sum_{i=1}^{n}(e_i^B)^4 \leq L$.
(C3) The eigenvalues of the sample covariance matrix $n^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{X}$ are bounded away from zero and infinity.
(C4) There exists a constant $B > 0$ such that $\|\boldsymbol{\beta}^A\|_1 \leq B$, $\|\boldsymbol{\beta}^B\|_1 \leq B$.
(C5) Let δ_n be the maximum covariance between the error terms and the covariates

$$\delta_n = \max_{\omega=A,B}\left\{\max_j\left|\frac{1}{n}\sum_{i=1}^{n}(x_{ij} - (\overline{\boldsymbol{x}})_j)\left(e_i^\omega - \overline{e^\omega}\right)\right|\right\}.$$

Assume that $\delta_n = o\big(1/(s_\lambda\sqrt{\log p})\big)$ and $(s_\lambda \log p)/\sqrt{n} = o(1)$.

(C6) Let $\boldsymbol{\Sigma}_* = n^{-1}\sum_{i=1}^n \widehat{\mathbf{D}}^{-1}(\boldsymbol{x}_i - \overline{\boldsymbol{x}})(\boldsymbol{x}_i - \overline{\boldsymbol{x}})^{\mathrm{T}}\widehat{\mathbf{D}}^{-1}$. There exist constants $C_0 > 0$ and $\xi > 1$, such that

$$\left\|\boldsymbol{h}_{\gamma_*}^S\right\|_1 \le C_0 s_\lambda \|\boldsymbol{\Sigma}_* \boldsymbol{h}_{\gamma_*}\|_\infty, \qquad \forall \boldsymbol{h}_{\gamma_*} \in \mathcal{C},$$

where $\mathcal{C} = \left\{\boldsymbol{h}_{\gamma_*} : \left\|\boldsymbol{h}_{\gamma_*}^{S^c}\right\|_1 \le \xi \left\|\boldsymbol{h}_{\gamma_*}^S\right\|_1\right\}$ and $S = \left\{j : |\beta_j^A| > \lambda_A \text{ or } |\beta_j^B| > \lambda_B\right\}$.

(C7) Let $\nu = \min\{1/70, (3\tilde{p}_A)^2/70, (3 - 3\tilde{p}_A)^2/70\}$. For some constants $c > 0$, $L_0 > 0$, $0 < \eta < (\xi - 1)/(\xi + 1)$ and $1/\eta < M < \infty$, the regularization parameters of the SPAC-Lasso satisfy that

$$\lambda_A \in (\frac{1}{\eta}, M] \times \Big(\frac{2c(1+\nu)L^{1/2}}{\tilde{p}_A\sqrt{L_0}}\sqrt{\frac{2\log p}{n}} + \frac{\delta_n}{\sqrt{L_0}}\Big),$$
$$\lambda_B \in (\frac{1}{\eta}, M] \times \Big(\frac{2c(1+\nu)L^{1/2}}{\tilde{p}_B\sqrt{L_0}}\sqrt{\frac{2\log p}{n}} + \frac{\delta_n}{\sqrt{L_0}}\Big).$$

Condition (C1) is a basic assumption for the probability of receiving the treatment or control. Condition (C2) is a moment condition for x_{ij} and error terms e_i^w $(w = A,\ B)$, which is similar to the conditions in [12,21,22]. Conditions (C3) and (C4) are some regularity conditions for high-dimensional statistical inference (see [12,13,28,29]). Conditions (C5)–(C7) are needed to show the convergence rate for $\hat{\boldsymbol{\beta}}_{\text{SPAC-Lasso}}$, and assumed based on the definition of approximate sparsity. These assumptions are similar to those in [12,13], and are weaker than the assumptions for strict sparsity.

Theorem 1. *Suppose that regularity conditions (C1)–(C7) hold, as $n \to \infty$, then*

$$\sqrt{n}(\hat{\tau}_{\text{SPAC-Lasso}} - \tau) \xrightarrow{\mathcal{L}} N(0, \sigma^2),$$

where

$$\sigma^2 = \lim_{n\to\infty}\left[\frac{1-p_A}{p_A}\sigma_{e^A}^2 + \frac{p_A}{1-p_A}\sigma_{e^B}^2 + 2\sigma_{e^Ae^B}\right],$$

and $\sigma_{e^A}^2 = n^{-1}\sum_{i=1}^n (e_i^A)^2$, $\sigma_{e^B}^2 = n^{-1}\sum_{i=1}^n (e_i^B)^2$, $\sigma_{e^Ae^B} = n^{-1}\sum_{i=1}^n e_i^A e_i^B$.

Theorem 1 shows that the asymptotic normality of the proposed SPAC-Lasso adjustment estimator $\hat{\tau}_{\text{SPAC-Lasso}}$ for highly correlated covariates based on the approximate sparsity measures and appropriate tuning parameters λ_A and λ_B. Without loss of generality, we assume that $\overline{Y}_1 = 0$, $\overline{Y}_0 = 0$ and $\overline{\boldsymbol{x}} = 0$. The assumptions and the results in Theorem 1 are similar to that in [12,13].

Proof. According to the decomposition of $Y_i(1)$ and $Y_i(0)$ in (9), we have

$$\begin{aligned}
\sqrt{n}(\hat{\tau}_{\text{SPAC-Lasso}} - \tau) =& \sqrt{n}\Big(\overline{Y}_A - \overline{\boldsymbol{x}}_A^{\mathrm{T}}\hat{\boldsymbol{\beta}}_{\text{SPAC-Lasso}}^A\Big) - \sqrt{n}\Big(\overline{Y}_B - \overline{\boldsymbol{x}}_B^{\mathrm{T}}\hat{\boldsymbol{\beta}}_{\text{SPAC-Lasso}}^B\Big) \\
=& \sqrt{n}\Big(\overline{\boldsymbol{x}}_A^{\mathrm{T}}\boldsymbol{\beta}^A + \bar{e}_A - \overline{\boldsymbol{x}}_A^{\mathrm{T}}\hat{\boldsymbol{\beta}}_{\text{SPAC-Lasso}}^A\Big) \\
& - \sqrt{n}\Big(\overline{\boldsymbol{x}}_B^{\mathrm{T}}\boldsymbol{\beta}^B + \bar{e}_B - \overline{\boldsymbol{x}}_B^{\mathrm{T}}\hat{\boldsymbol{\beta}}_{\text{SPAC-Lasso}}^B\Big) \\
=& \underbrace{\sqrt{n}(\bar{e}_A - \bar{e}_B)}_{I_1} - \underbrace{\sqrt{n}\Big(\overline{\boldsymbol{x}}_A^{\mathrm{T}}\boldsymbol{h}^A - \overline{\boldsymbol{x}}_B^{\mathrm{T}}\boldsymbol{h}^B\Big)}_{I_2},
\end{aligned} \tag{10}$$

where $\boldsymbol{h}^A = \hat{\boldsymbol{\beta}}_{\text{SPAC-Lasso}}^A - \boldsymbol{\beta}^A$ and $\boldsymbol{h}^B = \hat{\boldsymbol{\beta}}_{\text{SPAC-Lasso}}^B - \boldsymbol{\beta}^B$, $\bar{e}_A = n_A^{-1}\sum_{i\in A} e_i^A$, $\bar{e}_B = n_B^{-1}\sum_{i\in B} e_i^B$. Combining the Theorem 1 in [21] and replacing a and b with e^A and e^B, we have $I_1 \xrightarrow{\mathcal{L}} N(0, \sigma^2)$, where σ^2 is defined in Theorem 1.

By using the Hölder inequality, we have

$$\left|\bar{\boldsymbol{x}}_A^{\mathrm{T}}\boldsymbol{h}^A\right| \le \|\bar{\boldsymbol{x}}_A\|_\infty \|\boldsymbol{h}^A\|_1.$$

Invoking Lemma 1 in [13] and conditions (C1)–(C2), we have

$$\|\bar{\boldsymbol{x}}_A\|_\infty = O_p\left(\sqrt{\frac{\log p}{n}}\right). \tag{11}$$

According to (5), we obtain that

$$\begin{aligned}
\boldsymbol{h}^A &= \hat{\boldsymbol{\beta}}^A_{\mathrm{SPAC-Lasso}} - \boldsymbol{\beta}^A = \widehat{\mathbf{D}}\hat{\boldsymbol{\gamma}}^A_{\mathrm{SPAC-Lasso}} - \mathbf{D}\boldsymbol{\gamma}^A \\
&= \widehat{\mathbf{D}}\left(\hat{\boldsymbol{\gamma}}^A_{\mathrm{SPAC-Lasso}} - \boldsymbol{\gamma}^A\right) + \left(\widehat{\mathbf{D}} - \mathbf{D}\right)\boldsymbol{\gamma}^A \\
&= \widehat{\mathbf{D}}\boldsymbol{h}^A_\gamma + \left(\widehat{\mathbf{D}} - \mathbf{D}\right)\boldsymbol{\gamma}^A,
\end{aligned} \tag{12}$$

where $\boldsymbol{h}^A_\gamma = \hat{\boldsymbol{\gamma}}^A_{\mathrm{SPAC-Lasso}} - \boldsymbol{\gamma}^A$, $\mathbf{D} = \mathrm{diag}\left\{d_{11}^{1/2}, \ldots, d_{pp}^{1/2}\right\}$.

Using Lemma A3 in the following Appendix A, we have

$$\|\boldsymbol{h}^A_\gamma\|_1 = o_p\left(\frac{1}{\sqrt{\log p}}\right).$$

Together with (12) and conditions (C3)–(C4), we have $\|\boldsymbol{h}^A\|_1 = o_p\left(\frac{1}{\sqrt{\log p}}\right)$. Then,

$$\sqrt{n}\,\bar{\boldsymbol{x}}_A^{\mathrm{T}}\boldsymbol{h}^A = \sqrt{n}\cdot O_p\left(\sqrt{\frac{\log p}{n}}\right)\cdot o_p\left(\frac{1}{\sqrt{\log p}}\right) = o_p(1).$$

Similarly, we can obtain that $\sqrt{n}\,\bar{\boldsymbol{x}}_B^{\mathrm{T}}\boldsymbol{h}^B = o_p(1)$. Hence, we have $I_2 = o_p(1)$. This completes the proof of Theorem 1. □

3. Simulation Studies

In this section, the performance of the proposed SPAC-Lasso, SPAC-SCAD adjustment estimators are evaluated, and compared with those of the unadjusted estimator (unadj) and the penalty-based regression adjustment estimators (Lasso, SCAD, Enet). The R package "glmnet" is used to solve the problems of the Elastic-net and Lasso. To implement the SCAD and SPAC-SCAD methods, $a = 3.7$ is chosen and the R package "ncvreg" is used ([3]). In addition, the estimation of precision matrix is implemented by the R package "fastclime" of [30]. For each regression adjustment method, the tuning parameter is selected by the 10-fold cross-validation. The results are based on 2000 repeated simulations.

The potential outcomes are generated as follows,

$$\begin{aligned}
Y_i(1) &= \sum_{j=1}^{p} x_{ij}\beta_j + z + e_i^A, \quad i = 1, \ldots, n, \\
Y_i(0) &= \sum_{j=1}^{p} x_{ij}\beta_j + e_i^B, \qquad\quad i = 1, \ldots, n,
\end{aligned}$$

where $n = 250$, $p = 500, 1000$ and 2000, $z \sim U(0,2)$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^{\mathrm{T}}$ is the coefficient vector, error terms e_i^A and e_i^B are i.i.d generated from $N(0,1)$. The covariates vector $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ip})^{\mathrm{T}}$ is drawn from a multivariate normal distribution $N(\mathbf{0}_{p\times 1}, \boldsymbol{\Sigma}_{p\times p})$, and the covariance matrix $\boldsymbol{\Sigma}_{p\times p}$ has the following block-exchangeable structure,

$$\mathbf{\Sigma}_{p\times p} = \begin{pmatrix} \mathbf{\Sigma}^{11}_{q\times q} & \mathbf{\Sigma}^{12}_{q\times(p-q)} \\ (\mathbf{\Sigma}^{12}_{q\times(p-q)})^{\mathrm{T}} & \mathbf{\Sigma}^{22}_{(p-q)\times(p-q)} \end{pmatrix},$$

where q is the number of the non-zero elements, and

$$(\mathbf{\Sigma}^{11})_{s,j} = \begin{cases} 1 & s = j \\ \alpha_1 & s \neq j \end{cases}, \ (\mathbf{\Sigma}^{12})_{s,j} = \alpha_2, \ (\mathbf{\Sigma}^{22})_{s,j} = \begin{cases} 1 & s = j \\ \alpha_3 & s \neq j \end{cases}.$$

Here, the parameter vector $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)$ measures the correlations of covariates. To explore the effect of the correlations of covariates, we consider three different choices of $\boldsymbol{\alpha}$ as $\boldsymbol{\alpha} = (0.1, 0.3, 0.8)$, $(0.2, 0.5, 0.9)$ and $(0.5, 0.7, 0.9)$. For the coefficient vector $\boldsymbol{\beta}$, the first q coefficients take the nonzero values while the remaining $p - q$ elements are set to zero. In this simulation, we set $q = 9$ and let

$$\boldsymbol{\beta} = (1, 1, 1, 1.5, 1.5, 1.5, 2, 2, 2, 0, \ldots, 0)^{\mathrm{T}}.$$

From the generated data, we randomly assign $n_A = 125$ units to the treatment group A and the remaining $n_B = n - n_A = 125$ units to the control group B.

To assess the finite-sample performance of our proposed SPAC adjustment method (SPAC-Lasso, SPAC-SCAD), we compute the $|\text{Bias}|$, the standard deviations (SD) and the root-mean square errors (RMSE) of each estimator. In our simulation studies, $|\text{Bias}|$ represents absolute difference between the estimated ATE and the true ATE. The numerical results are shown in Table 1.

Table 1. Finite sample performance of the ATE estimators.

		$\boldsymbol{\alpha} = (0.1, 0.3, 0.8)$			$\boldsymbol{\alpha} = (0.2, 0.5, 0.9)$			$\boldsymbol{\alpha} = (0.5, 0.7, 0.9)$		
p	Methods	|Bias|	SD	RMSE	|Bias|	SD	RMSE	|Bias|	SD	RMSE
$p = 500$	unadj	0.0040	0.7855	0.7855	0.0058	0.8778	0.8778	0.0035	1.1844	1.1844
	Lasso	0.0021	0.1155	0.1155	0.0085	0.1512	0.1515	0.0028	0.1643	0.1643
	SCAD	0.0010	0.0948	0.0948	0.0159	0.2358	0.2364	0.0019	0.2839	0.2839
	Enet	0.0023	0.1098	0.1098	0.0086	0.1533	0.1535	0.0032	0.1645	0.1646
	SPAC-Lasso	0.0014	0.1068	0.1068	0.0020	0.1050	0.1050	0.0008	0.1099	0.1099
	SPAC-SCAD	0.0011	0.0946	0.0946	0.0016	0.0977	0.0978	0.0013	0.1284	0.1284
$p = 1000$	unadj	0.0052	0.7796	0.7796	0.0162	0.9152	0.9153	0.0429	1.1905	1.1913
	Lasso	0.0009	0.1199	0.1199	0.0027	0.1515	0.1515	0.0093	0.1494	0.1497
	SCAD	0.0003	0.0918	0.0918	0.0043	0.2403	0.2404	0.0132	0.2665	0.2669
	Enet	0.0013	0.1136	0.1136	0.0025	0.1524	0.1524	0.0087	0.1523	0.1526
	SPAC-Lasso	0.0001	0.1051	0.1051	0.0007	0.1075	0.1075	0.0036	0.0955	0.0956
	SPAC-SCAD	0.0002	0.0916	0.0916	0.0000	0.0971	0.0971	0.0044	0.1157	0.1158
$p = 2000$	unadj	0.0104	0.7765	0.7766	0.0556	0.9237	0.9254	0.0696	1.2492	1.2512
	Lasso	0.0013	0.1200	0.1201	0.0010	0.1658	0.1658	0.0092	0.1801	0.1804
	SCAD	0.0004	0.0997	0.0997	0.0037	0.2521	0.2521	0.0134	0.2620	0.2623
	Enet	0.0020	0.1158	0.1158	0.0027	0.1659	0.1659	0.0063	0.1879	0.1880
	SPAC-Lasso	0.0005	0.1074	0.1074	0.0004	0.1045	0.1045	0.0040	0.1104	0.1105
	SPAC-SCAD	0.0002	0.0991	0.0991	0.0013	0.0969	0.0969	0.0034	0.1459	0.1459

From the results in Table 1, we observe the following results.

(1) When $\boldsymbol{\alpha} = (0.1, 0.3, 0.8)$, our proposed SPAC methods (SPAC-Lasso, SPAC-SCAD) outperform the unadj method in terms of SDs and RMSEs, and have similar performance with Lasso, SCAD and Enet. Specifically, the SPAC adjustment method reduces the RMSE of unadjusted estimator (unadj) by 86–88%.

(2) As the correlations of covariates $\boldsymbol{\alpha}$ increase, the superiority of the proposed SPAC adjustment method becomes obvious. For example, when $p = 2000$ and $\boldsymbol{\alpha} = (0.5, 0.7, 0.9)$, the RMSEs of SPAC-Lasso and SPAC-SCAD are 39% and 45% smaller than those of the Lasso and SCAD, respectively.

The variable selection performance is assessed by the mean number of selected nonzero coefficients (S), the false negative rate (FNR), and false positive rate (FPR). The FNR and FPR are defined as

$$\text{FNR}: \ \frac{\sum_{j=1}^{p} I(\hat{\beta}_j = 0, \beta_j \neq 0)}{\sum_{j=1}^{p} I(\beta_j \neq 0)}, \qquad \text{FPR}: \ \frac{\sum_{j=1}^{p} I(\hat{\beta}_j \neq 0, \beta_j = 0)}{\sum_{j=1}^{p} I(\beta_j = 0)},$$

where $I(\cdot)$ is the indicator function. The FNR and FPR indicate the proportion of important covariates which are not selected and the proportion of selected unimportant covariates, respectively. The smaller false rates (FNR and FPR) indicate a better performance for variable selection. The variable selection results are listed in the following Table 2.

Table 2. Variable selection results for treatment and control groups.

		$\boldsymbol{\alpha} = (0.1, 0.3, 0.8)$			$\boldsymbol{\alpha} = (0.2, 0.5, 0.9)$			$\boldsymbol{\alpha} = (0.5, 0.7, 0.9)$		
p	**Methods**	**S**	**FNR**	**FPR**	**S**	**FNR**	**FPR**	**S**	**FNR**	**FPR**
$p = 500$	Lasso	17.650	0.000	0.018	34.636	0.140	0.054	43.091	0.170	0.073
	SCAD	9.009	0.000	0.000	12.937	0.552	0.018	32.333	0.819	0.063
	Enet	18.271	0.000	0.019	36.085	0.154	0.058	45.158	0.182	0.077
	SPAC-Lasso	9.007	0.000	0.000	9.000	0.000	0.000	9.000	0.000	0.000
	SPAC-SCAD	9.000	0.000	0.000	9.000	0.000	0.000	8.937	0.007	0.000
$p = 1000$	Lasso	20.116	0.000	0.011	40.922	0.260	0.035	47.313	0.103	0.040
	SCAD	9.325	0.000	0.000	21.004	0.755	0.019	38.149	0.886	0.037
	Enet	20.841	0.000	0.012	42.743	0.266	0.036	49.456	0.114	0.042
	SPAC-Lasso	9.193	0.000	0.000	9.005	0.000	0.000	9.000	0.000	0.000
	SPAC-SCAD	9.000	0.000	0.000	9.000	0.000	0.000	8.984	0.000	0.000
$p = 2000$	Lasso	20.083	0.000	0.006	42.091	0.277	0.018	54.969	0.281	0.024
	SCAD	9.082	0.000	0.000	20.077	0.716	0.009	42.502	0.958	0.021
	Enet	20.748	0.000	0.006	44.270	0.286	0.019	58.556	0.293	0.026
	SPAC-Lasso	9.218	0.000	0.000	9.002	0.000	0.000	9.000	0.000	0.000
	SPAC-SCAD	9.000	0.000	0.000	9.000	0.000	0.000	8.974	0.003	0.000

From Table 2, we obtain the following results.

(1) When the important covariates and unimportant covariates are weakly correlated $\alpha_2 = 0.3$ of $\boldsymbol{\alpha} = (0.1, 0.3, 0.8)$, the SCAD and our proposed SPAC-Lasso, SPAC-SCAD adjustment methods perform well in terms of S, FNRs and FPRs, where the false rates (FNRs and FPRs) are nearly 0. In comparison, the proportions of the selected unimportant variables (FPR) of Lasso and Enet are relatively large, which is also reflected in the number of selected nonzero elements (S).

(2) When the correlations of covariates increase, the proposed SPAC adjustment method (SPAC-Lasso and SPAC-SCAD) has satisfactory performance, while the existing penalty-based regression adjustment methods (Lasso, SCAD, Enet) perform badly. The mean numbers of the selected nonzero coefficients (S) of our proposed method are close to the number of true nonzero elements 9, but the existing adjusted methods fail to correctly select the nonzero and zero coefficients (relatively large FNRs and FPRs). For example, when $\boldsymbol{\alpha} = (0.5, 0.7, 0.9)$, the proportions of important covariates which are not selected (FNR) of SCAD exceed 0.819, while the largest FNR of SPAC-SCAD is only 0.007.

To further assess the performance of our proposed SPAC adjustment method, we calculate the mean of variance estimates (MVE) for σ in Theorem 1, the mean coverage probability (MCP) and mean interval length (MIL) of the 95% confidence intervals $[\hat{\tau} - Z_{0.975} \cdot \hat{\sigma}/\sqrt{n}, \hat{\tau} + Z_{0.975} \cdot \hat{\sigma}/\sqrt{n}]$, where Z_α is the α quantile of the standard normal distribution. We then compare the results of proposed method with those of the existing unadjusted (unadj) and penalty-based regression adjustment methods (Lasso, SCAD, Enet) in Table 3.

For the unadjusted method, the variance estimator is defined as

$$\hat{\sigma}^2_{\text{unadj}} = \frac{n}{n_A} \cdot \frac{1}{n_A - 1} \sum_{i \in A} \left\{ Y_i(1) - \overline{Y}_A \right\}^2 + \frac{n}{n_B} \cdot \frac{1}{n_B - 1} \sum_{i \in B} \left\{ Y_i(1) - \overline{Y}_B \right\}^2.$$

For the adjustment methods (SPAC-Lasso, SPAC-SCAD, Lasso, SCAD, Enet), we give the following Neyman-type conservative estimate of the variance σ^2, which is similar to that in [12,13].

$$\hat{\sigma}^2 = \frac{n}{n_A}\hat{\sigma}^2_{e^A} + \frac{n}{n_B}\hat{\sigma}^2_{e^B},$$

where

$$\hat{\sigma}^2_{e^A} = \frac{1}{n_A - df^A}\sum_{i\in A}\left\{Y_i(1) - \overline{Y}_A - (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_A)^{\mathrm{T}}\hat{\boldsymbol{\beta}}^A\right\}^2,$$
$$\hat{\sigma}^2_{e^B} = \frac{1}{n_B - df^B}\sum_{i\in B}\left\{Y_i(0) - \overline{Y}_B - (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_B)^{\mathrm{T}}\hat{\boldsymbol{\beta}}^B\right\}^2,$$

and $df^A = \|\hat{\boldsymbol{\beta}}^A\|_0 + 1$ and $df^B = \|\hat{\boldsymbol{\beta}}^B\|_0 + 1$ are degrees of freedom for treatment and control groups, respectively. $\hat{\boldsymbol{\beta}}^A$ and $\hat{\boldsymbol{\beta}}^B$ are estimated adjustment vectors and obtained by different penalties (Lasso, SCAD, Enet) and SPAC methods (SPAC-Lasso, SPAC-SCAD) in (4) and (7).

Table 3. The performance of the variance estimates and confidence intervals.

		$\boldsymbol{\alpha} = (0.1, 0.3, 0.8)$			$\boldsymbol{\alpha} = (0.2, 0.5, 0.9)$			$\boldsymbol{\alpha} = (0.5, 0.7, 0.9)$		
p	**Methods**	**MVE**	**MCP**	**MIL**	**MVE**	**MCP**	**MIL**	**MVE**	**MCP**	**MIL**
$p = 500$	unadj	12.638	0.956	3.133	14.293	0.952	3.543	18.832	0.949	4.669
	Lasso	2.257	0.984	0.560	2.415	0.953	0.599	2.464	0.937	0.611
	SCAD	2.066	0.994	0.512	3.645	0.942	0.904	4.322	0.945	1.071
	Enet	2.179	0.985	0.540	2.430	0.951	0.603	2.511	0.938	0.622
	SPAC-Lasso	2.204	0.988	0.546	2.107	0.988	0.522	2.207	0.986	0.547
	SPAC-SCAD	2.065	0.994	0.512	1.993	0.990	0.494	2.427	0.979	0.602
$p = 1000$	unadj	12.217	0.940	3.029	14.500	0.945	3.595	18.441	0.947	4.572
	Lasso	2.297	0.982	0.569	2.433	0.949	0.603	2.283	0.943	0.566
	SCAD	2.092	0.995	0.519	3.801	0.952	0.942	4.157	0.948	1.031
	Enet	2.209	0.982	0.548	2.443	0.950	0.606	2.330	0.944	0.578
	SPAC-Lasso	2.233	0.991	0.554	2.204	0.986	0.546	1.980	0.987	0.491
	SPAC-SCAD	2.094	0.995	0.519	2.076	0.992	0.515	2.243	0.983	0.556
$p = 2000$	unadj	12.646	0.961	3.135	14.984	0.953	3.715	20.061	0.954	4.974
	Lasso	2.224	0.982	0.551	2.542	0.942	0.630	2.534	0.918	0.628
	SCAD	2.051	0.987	0.509	3.856	0.941	0.956	4.161	0.949	1.032
	Enet	2.137	0.980	0.530	2.561	0.946	0.635	2.650	0.916	0.657
	SPAC-Lasso	2.171	0.991	0.538	2.147	0.989	0.532	2.165	0.986	0.537
	SPAC-SCAD	2.046	0.989	0.507	2.042	0.993	0.506	2.591	0.973	0.642

From Table 3, we observe that:

(1) When $\boldsymbol{\alpha} = (0.1, 0.3, 0.8)$, the proposed SPAC adjustment method (SPAC-Lasso, SPAC-SCAD) performs better than the unadjusted (unadj) method, and performs similarly to the penalty-based adjustment methods (Lasso, SCAD, Enet) in terms of MVE, MCP and MIL.

(2) When the important and unimportanr covariates are highly correlated, the coverage probabilities of proposed SPAC adjustment method (SPAC-Lasso, SPAC-SCAD) are higher than those of the unadj, Lasso, SCAD and Enet methods, the MVE-values of SPAC-Lasso and SPAC-SCAD are smaller than those of the other methods. For example, when $\boldsymbol{\alpha} = (0.5, 0.7, 0.9)$, the MCPs of Lasso, SCAD and Enet methods are uniformly below 0.950, while the MCP-values of SPAC-Lasso and SPAC-SCAD are aound 0.980.

(3) The mean interval lengths (MILs) of the SPAC-Lasso and SPAC-SCAD are shorter than those of the unadj, Lasso, SCAD and Enet. Particularly, when $\boldsymbol{\alpha} = (0.5, 0.7, 0.9)$, the MIL-values of SPAC-Lasso and SPAC-SCAD are 10–14% and 38–46% shorter than those of the Lasso and SCAD, respectively.

4. A Real Data Analysis

In the clinic, the human epidermal growth factor receptor type 2 (HER2) is considered as an important indicator in the classification of the breast cancer. Overexpression or amplification of HER2 (HER2+) might account for around 20% of early breast cancers. As a monoclonal antibody, trastuzumab (also known as Herceptin) has been shown to improve the event-free survival rate and the results of chemotherapy in patients with HER2+ breast cancer ([31]).

In this section, we shall consider the estimation problem of the average treatment effect (trastuzumab) and apply the proposed SPAC adjustment method to the dataset based on a NeoAdjuvant Herceptin (NOAH) randomized clinical trial. The dataset was originally demonstrated in [31] and collected in the Gene Expression Omnibus (GSE50948), and further studied by Refs. [13,32,33].

There were $n = 156$ patients in the trail: 63 patients received trastuzumab and neoadjuvant chemotherapy (treatment group, $T_i = 1$) and 93 patients received neoadjuvant chemotherapy alone (control group, $T_i = 0$), $i = 1, \ldots, n$. The pathological complete response (pCR) was measured by the absence of residual invasive breast cancer and viewed as the potential outcome variable Y_i. For each patient, 54,675 gene probes were observed and regarded as the covariates.

The dimension of covariates p = 54,675 is much larger than the sample size $n = 156$, we first apply the sure independence screening (SIS) method proposed in [1] to exclude some insignificant variables and reduce the dimension p to a suitable size. Following the suggestions of [13,34], the genes with little variation in intensity (i.e., for j-th gene satisfies $\max(\boldsymbol{X}_j) - \min(\boldsymbol{X}_j) \leq k$ with a given value k) are also removed. Then, $p_* = 2573$ gene probes are retained. Based on the dataset, we apply six methods (unadj, Lasso, SCAD, Enet and our proposed SPAC-Lasso and SPAC-SCAD) to estimate ATE. The tuning parameters of five regression adjustment methods (Lasso, SCAD, Enet, SPAC-Lasso and SPAC-SCAD) are chosen by 10-fold cross validation. For each method, we calculate the ATE estimator ($\hat{\tau}$), the number of the selected nonzero coefficients (S), asymptotic variance estimator ($\hat{\sigma}$) and 95% confidence interval length (L). The numerical results are presented in Table 4.

Table 4. The performance of different methods for the treatment effect estimation of trastuzumab.

	unadj	Lasso	SCAD	Enet	SPAC-Lasso	SPAC-SCAD
$\hat{\tau}$	0.2555	0.2491	0.2488	0.2473	0.2454	0.2435
S	–	16.500	17.000	20.000	15.000	8.500
$\hat{\sigma}$	0.9670	0.8031	0.8519	0.7566	0.7136	0.7317
L	0.3035	0.2521	0.2674	0.2375	0.2240	0.2296

The results in Table 4 show that all the ATE estimates are around 0.250. Combing this with that in [13,31,32], the trastuzumab indeed alleviates the patient's conditions and improve the prognosis. In addition, we find that the numbers of covariates selected by the SPAC adjustment method (SPAC-Lasso, SPAC-SCAD) are less than those selected by Lasso, SCAD and Enet, which is consistent with the discovery in the simulation studies. The estimated asymptotic variances ($\hat{\sigma}$) and 95% confidence interval lengths (L) of SPAC-Lasso and SPAC-SCAD are smallest. Specially, the values of $\hat{\sigma}$ of SPAC-Lasso and SPAC-SCAD are 11% and 14% smaller than those of the Lasso and SCAD, respectively, which implies that our proposed SPAC adjustment method can improve the asymptotic performance of the existing unadjusted and penalty-based regression adjustment methods.

5. Conclusions

In this paper, we studied the estimation problem of ATE for Rubin causal model when the covariates are highly correlated. We proposed the SPAC adjustment method (SPAC-Lasso, SPAC-SCAD) for ATE by combining SPAC variable selection method, Lasso and SCAD penalty functions, and regression adjustment technique, which is an extension for the SPAC method of high-dimensional regression models. In theory, we showed the asymptotic normality of the proposed SPAC-Lasso adjustment estimator under some regularity conditions. By some simulation studies and a real data analysis, we showed the advantages of our proposed method in terms of estimating average treatment effect and selecting the important covariates. Thus, the proposed SPAC adjustment method can improve the estimation accuracy for the Rubin causal model with highly correlated covariates.

Author Contributions: Conceptualization, G.L.; Methodology, L.Y.; Validation, F.Z.; Formal analysis, F.Z.; Investigation, Z.D.; Data curation, Z.D.; Writing—original draft preparation, Z.D.; Writing—review and editing, L.Y.; Supervision, L.Y.; Project administration, G.L.; Funding acquisition, G.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (Nos. 12001277, 12271046, 12131006 and 11971001), the National Social Science Foundation of China (No. 21BTJ030), the Tianjin Natural Science Foundation (No. 19JCZDJC32300).

Data Availability Statement: Not applicable.

Acknowledgments: The authors are grateful to the editor, the associate editor and the three anonymous referees for the constructive comments and suggestions that led to significant improvement of an early manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Some Lemmas and Their Proofs

This Section will provide three Lemmas that are needed for the proof of Theorem 1. We will drop the superscript on $\boldsymbol{h}_\gamma$, e, γ and $\hat{\gamma}$ and focus on the proof for treatment group A, as the same analysis can be applied to control group B.

Lemma A1. *Let* $\mathcal{M}_1 = \left\{ \left\| \frac{1}{n_A} \sum_{i\in A} \widetilde{\boldsymbol{X}}_i(e_i - \bar{e}_A) \right\|_\infty + \left\| \frac{1}{n_A} \sum_{i\in A} \widetilde{\boldsymbol{X}}_i(\boldsymbol{x}_i - \bar{\boldsymbol{x}}_A)^{\mathrm{T}}(\mathbf{D} - \widehat{\mathbf{D}})\boldsymbol{\gamma} \right\|_\infty \le \eta\lambda_A \right\}$, *where* $\bar{e}_A = n_A^{-1}\sum_{i\in A} e_i^A$. *Suppose that regularity conditions (C1)–(C7) hold, then*

$$P(\mathcal{M}_1) \ge 1 - \frac{2}{p}.$$

Proof. Recalling $\widetilde{\boldsymbol{X}}_i$, we have

$$\begin{aligned} \frac{1}{n_A}\sum_{i\in A} \widetilde{\boldsymbol{X}}_i(e_i - \bar{e}_A) &= \frac{1}{n_A}\sum_{i\in A} \widehat{\mathbf{D}}^{-1}(\boldsymbol{x}_i - \bar{\boldsymbol{x}}_A)(e_i - \bar{e}_A) \\ &= \widehat{\mathbf{D}}^{-1} \cdot \left(\frac{1}{n_A}\sum_{i\in A} \boldsymbol{x}_i e_i - \bar{\boldsymbol{x}}_A \bar{e}_A \right). \end{aligned}$$

By the condition (C3) and the sufficient accuracy of CLIME estimator $\hat{d}_{jj}$, there exists constants L_0 and L_1 such that for a sufficiently large n,

$$L_0 \le d_{11}, \ldots, d_{pp}, \hat{d}_{11}, \ldots, \hat{d}_{pp} \le L_1,$$

then combined with the triangle inequality, we have

$$\begin{aligned} \left\| \frac{1}{n_A}\sum_{i\in A} \widetilde{\boldsymbol{X}}_i(e_i - \bar{e}_A) \right\|_\infty &= \max_{1\le j\le p} \hat{d}_{jj}^{-1/2} \left(\frac{1}{n_A}\sum_{i\in A} \boldsymbol{x}_{ij} e_i - \bar{\boldsymbol{x}}_{A_j}\bar{e}_A \right) \\ &\le L_0^{-1/2} \left\| \frac{1}{n_A}\sum_{i\in A} \boldsymbol{x}_i e_i - \bar{\boldsymbol{x}}_A \bar{e}_A \right\|_\infty \end{aligned}$$

$$\leq L_0^{-1/2} \underbrace{\left\| \frac{1}{n_A} \sum_{i \in A} \boldsymbol{x}_i e_i \right\|_\infty}_{J_1} + L_0^{-1/2} \underbrace{\|\overline{\boldsymbol{x}}_A \overline{e}_A\|_\infty}_{J_2}, \tag{A1}$$

where $\overline{\boldsymbol{x}}_{A_j}$ is the j-th element of $\overline{\boldsymbol{x}}_A$.

For the first term J_1 in (A1), we have

$$\begin{aligned} J_1 &\leq \left\| \frac{1}{n_A} \sum_{i \in A} \boldsymbol{x}_i e_i - \frac{1}{n} \sum_{i=1}^n \boldsymbol{x}_i e_i \right\|_\infty + \left\| \frac{1}{n} \sum_{i=1}^n \boldsymbol{x}_i e_i \right\|_\infty \\ &\leq \left\| \frac{1}{n_A} \sum_{i \in A} \boldsymbol{x}_i e_i - \frac{1}{n} \sum_{i=1}^n \boldsymbol{x}_i e_i \right\|_\infty + \delta_n, \end{aligned}$$

where δ_n is defined in condition (C5). By condition (C2) and Cauchy-Schwarz inequality, we have

$$\frac{1}{n} \sum_{i=1}^n x_{ij}^2 e_i^2 \leq \left(\frac{1}{n} \sum_{i=1}^n x_{ij}^4 \right)^{1/2} \left(\frac{1}{n} \sum_{i=1}^n e_i^4 \right)^{1/2} \leq L.$$

Using Lemma S1 in [12], we can show that

$$P\left(\left\| \frac{1}{n_A} \sum_{i \in A} \boldsymbol{x}_i e_i - \frac{1}{n} \sum_{i=1}^n \boldsymbol{x}_i e_i \right\|_\infty > t_n \right) \leq 2\exp\left\{ \log p - \frac{n_A \tilde{p}_A t_n^2}{(1+\nu)^2 L} \right\} = 2\exp\{-\log p\} = \frac{2}{p},$$

where $t_n = (1+\nu) L^{1/2} \tilde{p}_A^{-1} \sqrt{\frac{2\log p}{n}}$. Hence,

$$P(J_1 \leq t_n + \delta_n) \geq 1 - \frac{2}{p}. \tag{A2}$$

For the second term J_2 in (A1), using the condition (C2) and Lemma 1 in [13], we have

$$P\left(\|\overline{\boldsymbol{x}}_A \overline{e}_A\|_\infty \leq \frac{(1+\nu) L^{1/2}}{\tilde{p}_A} \sqrt{\frac{2\log p}{n}} \right) \geq 1 - \frac{2}{p}. \tag{A3}$$

Together with (A2) and (A3), it is easy to see that

$$P\left(\left\| \frac{1}{n_A} \sum_{i \in A} \widetilde{\boldsymbol{X}}_i (e_i - \overline{e}_A) \right\|_\infty \leq L_0^{-1/2} (2t_n + \delta_n) \right) \geq 1 - \frac{2}{p}. \tag{A4}$$

Due to $\widetilde{\boldsymbol{X}}_i = \widehat{\mathbf{D}}^{-1}(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_A)$, we have

$$\begin{aligned} & \left\| \frac{1}{n_A} \sum_{i \in A} \widehat{\mathbf{D}}^{-1} (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_A)(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_A)^{\mathrm{T}} (\mathbf{D} - \widehat{\mathbf{D}}) \gamma \right\|_\infty \\ \leq & \frac{1}{\sqrt{L_0}} O_p\left(M_1 \sqrt{\frac{\log p}{n}} \right) \cdot \left\| \frac{1}{n_A} \sum_{i \in A} (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_A)(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_A)^{\mathrm{T}} \gamma \right\|_\infty \\ = & \frac{1}{\sqrt{L_0}} O_p\left(M_1 \sqrt{\frac{\log p}{n}} \right) \cdot \left\| \frac{1}{n_A} \sum_{i \in A} \boldsymbol{x}_i \boldsymbol{x}_i^{\mathrm{T}} \gamma - \overline{\boldsymbol{x}}_A \overline{\boldsymbol{x}}_A^{\mathrm{T}} \gamma \right\|_\infty \\ \leq & \frac{1}{\sqrt{L_0}} O_p\left(M_1 \sqrt{\frac{\log p}{n}} \right) \cdot \left(\left\| \frac{1}{n_A} \sum_{i \in A} \boldsymbol{x}_i \boldsymbol{x}_i^{\mathrm{T}} \gamma \right\|_\infty + \left\| \overline{\boldsymbol{x}}_A \overline{\boldsymbol{x}}_A^{\mathrm{T}} \gamma \right\|_\infty \right) \\ \leq & \frac{1}{\sqrt{L_0}} O_p\left(M_1 \sqrt{\frac{\log p}{n}} \right) \end{aligned}$$

$$
\begin{aligned}
&\cdot \left(\left\| \left(\frac{1}{n_A} \sum_{i\in A} \boldsymbol{x}_i \boldsymbol{x}_i^{\mathrm{T}} - \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i^{\mathrm{T}} \right) \gamma \right\|_{\infty} + \left\| \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i^{\mathrm{T}} \gamma \right\|_{\infty} + \left\| \overline{\boldsymbol{x}}_A \overline{\boldsymbol{x}}_A^{\mathrm{T}} \gamma \right\|_{\infty} \right) \\
\leq &\frac{1}{\sqrt{L_0}} O_p \left(M_1 \sqrt{\frac{\log p}{n}} \right) \\
&\cdot \left(\left\| \frac{1}{n_A} \sum_{i\in A} \boldsymbol{x}_i \boldsymbol{x}_i^{\mathrm{T}} - \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i^{\mathrm{T}} \right\|_{\infty} \cdot \|\gamma\|_1 + \|\gamma\|_1 + \left\| \overline{\boldsymbol{x}}_A \overline{\boldsymbol{x}}_A^{\mathrm{T}} \right\|_{\infty} \cdot \|\gamma\|_1 \right),
\end{aligned} \tag{A5}
$$

where $M_1 > 0$. By Cauchy-Schwarz inequality and condition (C2), we have

$$
\frac{1}{n} \sum_{i=1}^{n} x_{ij}^2 x_{ik}^2 \leq \left(\frac{1}{n} \sum_{i=1}^{n} x_{ij}^4 \right)^{\frac{1}{2}} \left(\frac{1}{n} \sum_{i=1}^{n} x_{ik}^4 \right)^{\frac{1}{2}} \leq L.
$$

Combined with Lemma S1 in [12], we have

$$
P\left(\left\| \frac{1}{n_A} \sum_{i\in A} \boldsymbol{x}_i \boldsymbol{x}_i^{\mathrm{T}} - \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i^{\mathrm{T}} \right\|_{\infty} \geq \frac{(1+\nu) L^{1/2}}{\tilde{p}_A} \sqrt{\frac{3\log p}{n}} \right) \leq \frac{2}{p}.
$$

By Lemma 1 in [13], we have

$$
\left\| \overline{\boldsymbol{x}}_A \overline{\boldsymbol{x}}_A^{\mathrm{T}} \right\|_{\infty} \leq (\|\overline{\boldsymbol{x}}_A\|_{\infty})^2 = o_p\left(\sqrt{\frac{\log p}{n}} \right).
$$

Recall the definition for SPAC and condition (C4), we have

$$
\|\gamma\|_1 = \sum_{j=1}^{p} \left| \frac{\beta_j}{\sqrt{d_{jj}}} \right| \leq \frac{1}{\sqrt{L_0}} \sum_{j=1}^{p} |\beta_j| \leq \frac{B}{\sqrt{L_0}}.
$$

Together the above results, we have

$$
I_2 = \left\| \frac{1}{n_A} \sum_{i\in A} \widetilde{\mathbf{X}}_i (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_A)^{\mathrm{T}} (\mathbf{D} - \widehat{\mathbf{D}}) \gamma \right\|_{\infty} = O_p\left(\sqrt{\frac{\log p}{n}} \right). \tag{A6}
$$

By (A4), (A6) and condition (C7), we have

$$
P\left(\left\| \frac{1}{n_A} \sum_{i\in A} \widetilde{\mathbf{X}}_i (e_i - \bar{e}_A) \right\|_{\infty} + \left\| \frac{1}{n_A} \sum_{i\in A} \widetilde{\mathbf{X}}_i (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_A)^{\mathrm{T}} (\mathbf{D} - \widehat{\mathbf{D}}) \gamma \right\|_{\infty} \leq \eta \lambda_A \right) \geq 1 - \frac{2}{p}.
$$

Then the proof is finished. □

Lemma A2. *Let* $\mathcal{M}_2 = \left\{ \left\| \frac{1}{n_A} \sum_{i\in A} \widetilde{\mathbf{X}}_i \widetilde{\mathbf{X}}_i^{\mathrm{T}} - \frac{1}{n} \sum_{i=1}^{n} \widehat{\mathbf{D}}^{-1} \boldsymbol{x}_i \boldsymbol{x}_i^{\mathrm{T}} \widehat{\mathbf{D}}^{-1} \right\|_{\infty} \leq C_1 \sqrt{\frac{\log p}{n}} \right\}$ *and* $C_1 = 2(1+\nu) L^{1/2} (\tilde{p}_A L_0)^{-1}$. *Suppose that regularity conditions (C1)–(C3) hold, then*

$$
P(\mathcal{M}_2) \geq 1 - \frac{2}{p}.
$$

Proof. From the definition of $\widetilde{\mathbf{X}}_i$ in (A9), we have

$$
\left\| \frac{1}{n_A} \sum_{i\in A} \widetilde{\mathbf{X}}_i \widetilde{\mathbf{X}}_i^{\mathrm{T}} - \frac{1}{n} \sum_{i=1}^{n} \widehat{\mathbf{D}}^{-1} \boldsymbol{x}_i \boldsymbol{x}_i^{\mathrm{T}} \widehat{\mathbf{D}}^{-1} \right\|_{\infty} \leq \frac{1}{L_0} \left\| \frac{1}{n_A} \sum_{i\in A} \boldsymbol{x}_i \boldsymbol{x}_i^{\mathrm{T}} - \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i^{\mathrm{T}} - \overline{\boldsymbol{x}}_A \overline{\boldsymbol{x}}_A^{\mathrm{T}} \right\|_{\infty}
$$

$$\leq \frac{1}{L_0} \underbrace{\left\| \frac{1}{n_A} \sum_{i \in A} \boldsymbol{x}_i \boldsymbol{x}_i^{\mathrm{T}} - \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i^{\mathrm{T}} \right\|_\infty}_{(*)} + \frac{1}{L_0} \underbrace{\left\| \overline{\boldsymbol{x}}_A \overline{\boldsymbol{x}}_A^{\mathrm{T}} \right\|_\infty}_{(**)}, \tag{A7}$$

the last inequation is obtained by triangle inequality.

By Cauchy-Schwarz inequality and condition (C2), we have

$$\frac{1}{n} \sum_{i=1}^{n} x_{ij}^2 x_{ik}^2 \leq \left(\frac{1}{n} \sum_{i=1}^{n} x_{ij}^4 \right)^{1/2} \left(\frac{1}{n} \sum_{i=1}^{n} x_{ik}^4 \right)^{1/2} \leq L.$$

Invoking Lemma S1 in [12] and $n_A/n = \tilde{p}_A$ in condition (C1), we can bound the first term $(*)$ in (A7) as follows,

$$P\left(\left\| \frac{1}{n_A} \sum_{i \in A} \boldsymbol{x}_i \boldsymbol{x}_i^{\mathrm{T}} - \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i^{\mathrm{T}} \right\|_\infty \geq \frac{(1+\nu) L^{1/2}}{\tilde{p}_A} \sqrt{\frac{3 \log p}{n}} \right) \leq 2 \exp\{-\log p\} = \frac{2}{p}. \tag{A8}$$

For the second term $(**)$ in (A7), we have

$$\left\| \overline{\boldsymbol{x}}_A \overline{\boldsymbol{x}}_A^{\mathrm{T}} \right\|_\infty \leq \| \overline{\boldsymbol{x}}_A \|_\infty^2 = o_p\left(\sqrt{\frac{\log p}{n}} \right).$$

Together with (A8), we have

$$P\left(\left\| \frac{1}{n_A} \sum_{i \in A} \widetilde{\mathbf{X}}_i \widetilde{\mathbf{X}}_i^{\mathrm{T}} - \frac{1}{n} \sum_{i=1}^{n} \widehat{\mathbf{D}}^{-1} \boldsymbol{x}_i \boldsymbol{x}_i^{\mathrm{T}} \widehat{\mathbf{D}}^{-1} \right\|_\infty \leq C_1 \sqrt{\frac{\log p}{n}} \right) \geq 1 - \frac{2}{p}.$$

Then the proof is finished. □

Lemma A3. *Suppose that regularity conditions (C1)–(C7) hold, then*

$$\|\boldsymbol{h}_\gamma\|_1 = o_p\left(\frac{1}{\sqrt{\log p}} \right),$$

where $\boldsymbol{h}_\gamma = \hat{\gamma}_{\mathrm{SPAC-Lasso}} - \boldsymbol{\gamma}$.

Proof. Note the SPAC estimator $\hat{\gamma}_{\mathrm{SPAC-Lasso}}$ is defined by

$$\hat{\gamma}_{\mathrm{SPAC-Lasso}} = \underset{\gamma}{\operatorname{argmin}} \left[\frac{1}{2n_A} \sum_{i \in A} \left\{ Y_i(1) - \overline{Y}_A - (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_A)^{\mathrm{T}} \widehat{\mathbf{D}} \boldsymbol{\gamma} \right\}^2 + \lambda_A \sum_{j=1}^{p} \hat{d}_{jj} |\gamma_j| \right],$$

which can be rewritten as

$$\hat{\gamma}_* = \underset{\gamma_*}{\operatorname{argmin}} \left[\frac{1}{2n_A} \sum_{i \in A} \left\{ Y_i(1) - \overline{Y}_A - \widetilde{\mathbf{X}}_i^{\mathrm{T}} \boldsymbol{\gamma}_* \right\}^2 + \lambda_A \sum_{j=1}^{p} |\gamma_{*_j}| \right], \tag{A9}$$

where $\widetilde{\mathbf{X}}_i = \widehat{\mathbf{D}}^{-1}(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_A)$, $\boldsymbol{\gamma}_* = \widehat{\mathbf{D}}^2 \boldsymbol{\gamma}$.

Then, the Karush-Kuhn-Tucker (KKT) condition for $\hat{\boldsymbol{\gamma}}^*$ is

$$\frac{1}{n_A} \sum_{i \in A} \widetilde{\mathbf{X}}_i \left\{ Y_i(1) - \overline{Y}_A - \widetilde{\mathbf{X}}_i^{\mathrm{T}} \hat{\boldsymbol{\gamma}}_* \right\} = \lambda_A \boldsymbol{\kappa}, \tag{A10}$$

where $\boldsymbol{\kappa}$ is the subgradient of $\|\boldsymbol{\gamma}_*\|_1$ at $\boldsymbol{\gamma}_* = \hat{\boldsymbol{\gamma}}_*$, that is

$$\boldsymbol{\kappa} \in \partial\|\boldsymbol{\gamma}_*\|_1\Big|_{\boldsymbol{\gamma}_*=\hat{\boldsymbol{\gamma}}_*} \qquad \text{with} \quad \begin{cases} \kappa_j \in [-1,1], & \text{if} \quad \hat{\gamma}_{*_j} = 0, \\ \kappa_j = \operatorname{sign}(\hat{\gamma}_{*_j}), & \text{otherwise.} \end{cases}$$

By the decomposition of $Y_i(1)$ in (9), we have

$$\begin{aligned} Y_i(1) - \overline{Y}_A &= (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_A)^{\mathrm{T}}\boldsymbol{\beta} + e_i - \overline{e}_A \\ &= (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_A)^{\mathrm{T}}\mathbf{D}\boldsymbol{\gamma} + e_i - \overline{e}_A \\ &= \widetilde{\boldsymbol{X}}_i^{\mathrm{T}}\boldsymbol{\gamma}_* + (e_i - \overline{e}_A) + (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_A)^{\mathrm{T}}(\mathbf{D} - \widehat{\mathbf{D}})\boldsymbol{\gamma}. \end{aligned}$$

Hence, (A10) can be expressed as

$$\frac{1}{n_A}\sum_{i\in A}\widetilde{\boldsymbol{X}}_i\widetilde{\boldsymbol{X}}_i^{\mathrm{T}}(\boldsymbol{\gamma}_* - \hat{\boldsymbol{\gamma}}_*) + \frac{1}{n_A}\sum_{i\in A}\widetilde{\boldsymbol{X}}_i(e_i - \overline{e}_A) + \frac{1}{n_A}\sum_{i\in A}\widetilde{\boldsymbol{X}}_i(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_A)^{\mathrm{T}}(\mathbf{D} - \widehat{\mathbf{D}})\boldsymbol{\gamma} = \lambda_A\boldsymbol{\kappa}, \tag{A11}$$

where $\mathbf{D} = \operatorname{diag}\{d_{11}^{1/2}, \ldots, d_{pp}^{1/2}\}$. Premultiplying (A11) by $-\boldsymbol{h}_{\gamma_*}^{\mathrm{T}} = (\boldsymbol{\gamma}_* - \hat{\boldsymbol{\gamma}}_*)^{\mathrm{T}}$, we have

$$\begin{aligned} \lambda_A(\boldsymbol{\gamma}_* - \hat{\boldsymbol{\gamma}}_*)^{\mathrm{T}}\boldsymbol{\kappa} =& \frac{1}{n_A}\sum_{i\in A}(\boldsymbol{\gamma}_* - \hat{\boldsymbol{\gamma}}_*)^{\mathrm{T}}\widetilde{\boldsymbol{X}}_i\widetilde{\boldsymbol{X}}_i^{\mathrm{T}}(\boldsymbol{\gamma}_* - \hat{\boldsymbol{\gamma}}_*) - \frac{1}{n_A}\sum_{i\in A}\boldsymbol{h}_{\gamma_*}^{\mathrm{T}}\widetilde{\boldsymbol{X}}_i(e_i - \overline{e}_A) \\ &- \frac{1}{n_A}\sum_{i\in A}\boldsymbol{h}_{\gamma_*}^{\mathrm{T}}\widetilde{\boldsymbol{X}}_i(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_A)^{\mathrm{T}}(\mathbf{D} - \widehat{\mathbf{D}})\boldsymbol{\gamma}. \end{aligned}$$

Then, we have

$$\begin{aligned} \frac{1}{n_A}\sum_{i\in A}\left(\widetilde{\boldsymbol{X}}_i^{\mathrm{T}}\boldsymbol{h}_{\gamma_*}\right)^2 \leq& \lambda_A(\|\boldsymbol{\gamma}_*\|_1 - \|\hat{\boldsymbol{\gamma}}_*\|_1) + \frac{1}{n_A}\sum_{i\in A}\boldsymbol{h}_{\gamma_*}^{\mathrm{T}}\widetilde{\boldsymbol{X}}_i(e_i - \overline{e}_A) \\ &+ \frac{1}{n_A}\sum_{i\in A}\boldsymbol{h}_{\gamma_*}^{\mathrm{T}}\widetilde{\boldsymbol{X}}_i(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_A)^{\mathrm{T}}(\mathbf{D} - \widehat{\mathbf{D}})\boldsymbol{\gamma}. \end{aligned}$$

Based on Hölder inequality, the above inequation can be written as

$$\begin{aligned} \frac{1}{n_A}\sum_{i\in A}\left(\widetilde{\boldsymbol{X}}_i^{\mathrm{T}}\boldsymbol{h}_{\gamma_*}\right)^2 \leq& \lambda_A(\|\boldsymbol{\gamma}_*\|_1 - \|\hat{\boldsymbol{\gamma}}_*\|_1) + \|\boldsymbol{h}_{\gamma_*}\|_1\left\|\frac{1}{n_A}\sum_{i\in A}\widetilde{\boldsymbol{X}}_i(e_i - \overline{e}_A)\right\|_\infty \\ &+ \|\boldsymbol{h}_{\gamma_*}\|_1\left\|\frac{1}{n_A}\sum_{i\in A}\widetilde{\boldsymbol{X}}_i(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_A)^{\mathrm{T}}(\mathbf{D} - \widehat{\mathbf{D}})\boldsymbol{\gamma}\right\|_\infty. \end{aligned}$$

Using Lemma A1 in the Appendix, we have

$$\frac{1}{n_A}\sum_{i\in A}\left(\widetilde{\boldsymbol{X}}_i^{\mathrm{T}}\boldsymbol{h}_{\gamma_*}\right)^2 \leq \lambda_A(\|\boldsymbol{\gamma}_*\|_1 - \|\hat{\boldsymbol{\gamma}}_*\|_1 + \eta\|\boldsymbol{h}_{\gamma_*}\|_1).$$

According to the triangle inequality and the definition of $\boldsymbol{h}_{\gamma_*} = \hat{\boldsymbol{\gamma}}_* - \boldsymbol{\gamma}_*$, we have

$$\begin{aligned} \|\boldsymbol{\gamma}_*\|_1 - \|\hat{\boldsymbol{\gamma}}_*\|_1 &\leq 2\left\|\boldsymbol{\gamma}_*^{S^c}\right\|_1 + \left\|\hat{\boldsymbol{\gamma}}_*^{S} - \boldsymbol{\gamma}_*^{S}\right\|_1 - \left\|\hat{\boldsymbol{\gamma}}_*^{S^c} - \boldsymbol{\gamma}_*^{S^c}\right\|_1 \\ &= \left\|\boldsymbol{h}_{\gamma_*}^{S}\right\|_1 - \left\|\boldsymbol{h}_{\gamma_*}^{S^c}\right\|_1 + 2\left\|\boldsymbol{\gamma}_*^{S^c}\right\|_1. \end{aligned}$$

Hence,

$$\begin{aligned} \frac{1}{n_A}\sum_{i\in A}\left(\widetilde{\boldsymbol{X}}_i^{\mathrm{T}}\boldsymbol{h}_{\gamma_*}\right)^2 \leq& \lambda_A\left(\left\|\boldsymbol{h}_{\gamma_*}^{S}\right\|_1 - \left\|\boldsymbol{h}_{\gamma_*}^{S^c}\right\|_1 + 2\left\|\boldsymbol{\gamma}_*^{S^c}\right\|_1 + \eta\|\boldsymbol{h}_{\gamma_*}\|_1\right) \\ \leq& \lambda_A\left[(\eta - 1)\left\|\boldsymbol{h}_{\gamma_*}^{S^c}\right\|_1 + (1+\eta)\left\|\boldsymbol{h}_{\gamma_*}^{S}\right\|_1 + 2\left\|\boldsymbol{\gamma}_*^{S^c}\right\|_1\right]. \end{aligned}$$

Noting that $n_A^{-1}\sum_{i\in A}\left(\widetilde{\mathbf{X}}_i^{\mathrm{T}}\boldsymbol{h}_{\gamma_*}\right)^2 \geq 0$, and by the definition of s_λ in Definition 1, we have

$$(1-\eta)\left\|\boldsymbol{h}_{\gamma_*}^{S^c}\right\|_1 \leq (1+\eta)\left\|\boldsymbol{h}_{\gamma_*}^{S}\right\|_1 + 2\left\|\gamma_*^{S^c}\right\|_1 \leq (1+\eta)\left\|\boldsymbol{h}_{\gamma_*}^{S}\right\|_1 + \frac{2L_1}{\sqrt{L_0}}s_\lambda\lambda_A. \tag{A12}$$

Next we will consider two cases for $(1+\eta)\left\|\boldsymbol{h}_{\gamma_*}^{S}\right\|_1 + 2L_1L_0^{-1/2}s_\lambda\lambda_A$, respectively.
(i) $(1+\eta)\left\|\boldsymbol{h}_{\gamma_*}^{S}\right\|_1 + 2L_1L_0^{-1/2}s_\lambda\lambda_A \geq (1-\eta)\xi\left\|\boldsymbol{h}_{\gamma_*}^{S}\right\|_1$. By (A12), we have

$$\begin{aligned}
\|\boldsymbol{h}_{\gamma_*}\|_1 = \left\|\boldsymbol{h}_{\gamma_*}^{S}\right\|_1 + \left\|\boldsymbol{h}_{\gamma_*}^{S^c}\right\|_1 &\leq \left\|\boldsymbol{h}_{\gamma_*}^{S}\right\|_1 + \frac{(1+\eta)}{1-\eta}\left\|\boldsymbol{h}_{\gamma_*}^{S}\right\|_1 + \frac{2L_1s_\lambda\lambda_A}{\sqrt{L_0}(1-\eta)} \\
&= \left(\frac{1+\eta}{1-\eta}+1\right)\left\|\boldsymbol{h}_{\gamma_*}^{S}\right\|_1 + \frac{2L_1s_\lambda\lambda_A}{\sqrt{L_0}(1-\eta)} \\
&\leq \frac{2L_1s_\lambda\lambda_A}{\sqrt{L_0}(1-\eta)}\left(\frac{2}{(1-\eta)\xi-(1+\eta)}+1\right).
\end{aligned}$$

Combining the above results with conditions (C5) and (C7), we can show that $s_\lambda\lambda_A = o\left(\frac{1}{\sqrt{\log p}}\right)$.
(ii) $(1+\eta)\left\|\boldsymbol{h}_{\gamma_*}^{S}\right\|_1 + 2L_1L_0^{-1/2}s_\lambda\lambda_A < (1-\eta)\xi\left\|\boldsymbol{h}_{\gamma_*}^{S}\right\|_1$. By (A12), we can obtain that

$$\left\|\boldsymbol{h}_{\gamma_*}^{S^c}\right\|_1 \leq \xi\left\|\boldsymbol{h}_{\gamma_*}^{S}\right\|_1.$$

By condition (C6), we have

$$\|\boldsymbol{h}_{\gamma_*}\|_1 = \left\|\boldsymbol{h}_{\gamma_*}^{S}\right\|_1 + \left\|\boldsymbol{h}_{\gamma_*}^{S^c}\right\|_1 \leq (1+\xi)\left\|\boldsymbol{h}_{\gamma_*}^{S}\right\|_1 \leq (1+\xi)C_0s_\lambda\|\boldsymbol{\Sigma}_*\boldsymbol{h}_{\gamma_*}\|_\infty. \tag{A13}$$

Using (A11) and Lemma A1, and combining with the triangle inequality, we can show that

$$\begin{aligned}
&\left\|\frac{1}{n_A}\sum_{i\in A}\widetilde{\mathbf{X}}_i\widetilde{\mathbf{X}}_i^{\mathrm{T}}\boldsymbol{h}_{\gamma_*}\right\|_\infty \\
\leq&\left\|\frac{1}{n_A}\sum_{i\in A}\widetilde{\mathbf{X}}_i\widetilde{\mathbf{X}}_i^{\mathrm{T}}\boldsymbol{h}_{\gamma_*} - \frac{1}{n_A}\sum_{i\in A}\widetilde{\mathbf{X}}_i(e_i-\bar{e}_A) - \frac{1}{n_A}\sum_{i\in A}\widetilde{\mathbf{X}}_i(\boldsymbol{x}_i-\bar{\boldsymbol{x}}_A)^{\mathrm{T}}(\mathbf{D}-\widehat{\mathbf{D}})\gamma\right\|_\infty \\
&+\left\|\frac{1}{n_A}\sum_{i\in A}\widetilde{\mathbf{X}}_i(e_i-\bar{e}_A)\right\|_\infty + \left\|\frac{1}{n_A}\sum_{i\in A}\widetilde{\mathbf{X}}_i(\boldsymbol{x}_i-\bar{\boldsymbol{x}}_A)^{\mathrm{T}}(\mathbf{D}-\widehat{\mathbf{D}})\gamma\right\|_\infty \\
\leq&(1+\eta)\lambda_A,
\end{aligned} \tag{A14}$$

where the last inequality holds on the set $\mathcal{M}_1$ of Lemma A1. When the events $\mathcal{M}_1$ and $\mathcal{M}_2$ of Lemma A2 hold, we have

$$\begin{aligned}
&\left\|\frac{1}{n}\sum_{i=1}^{n}\widehat{\mathbf{D}}^{-1}\boldsymbol{x}_i\boldsymbol{x}_i^{\mathrm{T}}\widehat{\mathbf{D}}^{-1}\boldsymbol{h}_{\gamma_*}\right\|_\infty \\
\leq&\left\|\frac{1}{n}\sum_{i=1}^{n}\widehat{\mathbf{D}}^{-1}\boldsymbol{x}_i\boldsymbol{x}_i^{\mathrm{T}}\widehat{\mathbf{D}}^{-1}\boldsymbol{h}_{\gamma_*} - \frac{1}{n_A}\sum_{i\in A}\widetilde{\mathbf{X}}_i\widetilde{\mathbf{X}}_i^{\mathrm{T}}\boldsymbol{h}_{\gamma_*}\right\|_\infty + \left\|\frac{1}{n_A}\sum_{i\in A}\widetilde{\mathbf{X}}_i\widetilde{\mathbf{X}}_i^{\mathrm{T}}\boldsymbol{h}_{\gamma_*}\right\|_\infty \\
\leq&C_1\sqrt{\frac{\log p}{n}}\|\boldsymbol{h}_{\gamma_*}\|_1 + \left\|\frac{1}{n_A}\sum_{i\in A}\widetilde{\mathbf{X}}_i\widetilde{\mathbf{X}}_i^{\mathrm{T}}\boldsymbol{h}_{\gamma_*}\right\|_\infty.
\end{aligned}$$

By condition (C5) and (A14), we can show that

$$\begin{aligned}\|\boldsymbol{h}_{\gamma_*}\|_1 &\le (1+\xi)C_0\left(s_\lambda\sqrt{\frac{\log p}{n}}\|\boldsymbol{h}_{\gamma_*}\|_1 + (1+\eta)s_\lambda\lambda_A\right)\\ &\le (1+\xi)C_0\{o(1)\|\boldsymbol{h}_{\gamma_*}\|_1 + (1+\eta)s_\lambda\lambda_A\}.\end{aligned}$$

Hence, we obtain that $\|\boldsymbol{h}_{\gamma_*}\|_1 = o_p\left(\frac{1}{\sqrt{\log p}}\right)$ by using the conditions (C5) and (C7).

Combining the cases (i) and (ii), we know that $\|\boldsymbol{h}_{\gamma_*}\|_1 = o_p\left(\frac{1}{\sqrt{\log p}}\right)$ holds. According to the definitions of $\boldsymbol{h}_\gamma$, $\boldsymbol{h}_{\gamma_*}$ and $\hat{\gamma}_*$, we have

$$\|\boldsymbol{h}_\gamma\|_1 = \|\hat{\gamma}_{\text{SPAC-Lasso}} - \gamma\|_1 = o_p\left(\frac{1}{\sqrt{\log p}}\right).$$

Then the proof is finished. □

References

1. Fan, J.; Lv, J. Sure independence screening for ultrahigh dimensional feature space. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2008**, *70*, 849–911. [CrossRef] [PubMed]
2. Tibshirani, R. Regression shrinkage and selection via the Lasso. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **1996**, *58*, 267–288. [CrossRef]
3. Fan, J.; Li, R. Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Am. Stat. Assoc.* **2001**, *96*, 1348–1360. [CrossRef]
4. Zou, H.; Hastie, T. Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2005**, *67*, 301–320. [CrossRef]
5. Rubin, D.B. Estimating causal effects of treatments in randomized and nonrandomized Studies. *J. Educ. Psychol.* **1974**, *66*, 688–701. [CrossRef]
6. Neyman, J. On the application of probability theory to agricultural experiments. Essay on principles, section 9. Translation of original 1923 paper, which appeared in roczniki nauk rolniczych. *Stat. Sci.* **1990**, *5*, 465–472.
7. Rubin, D.B. *Matched Sampling for Causal Effects*; Cambridge University Press: New York, NY, USA, 2006.
8. Imbens, G.W.; Rubin, D.B. *Causal Inference in Statistics, Social, and Biomedical Sciences: An Introduction*; Cambridge University Press: New York, NY, USA, 2015.
9. Belloni, A.; Chernozhukov, V.; Hansen, C. Inference on treatment effects after selection among high-dimensional controls. *Rev. Econ. Stud.* **2014**, *81*, 608–650. [CrossRef]
10. Belloni, A.; Chernozhukov, V.; Fernández-Val, I.; Hansen, C. Program evaluation and causal inference with high-dimensional data. *Econometrica* **2017**, *85*, 233–298. [CrossRef]
11. Wager, S.; Du, W.F.; Taylor, J.; Tibshirani, R. High-dimensional regression adjustments in randomized experiments. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 12673–12678. [CrossRef]
12. Bloniarz, A.; Liu, H.Z.; Zhang, C.H.; Sekhon, J.S.; Yu, B. Lasso adjustments of treatment effect estimates in randomized experiments. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 7383–7390. [CrossRef]
13. Yue, L.L.; Li, G.R.; Lian, H.; Wan, X. Regression adjustment for treatment effect with multicollinearity in high dimensions. *Comput. Stat. Data Anal.* **2019**, *134*, 17–35. [CrossRef]
14. Wang, H.; Lengerich, B.J.; Aragam, B.; Xing, E.P. Precision Lasso: Accounting for correlations and linear dependencies in high-dimensional genomic data. *Bioinformatics* **2019**, *35*, 1181–1187. [CrossRef] [PubMed]
15. Zhu, W.; Lévy-Leduc, C.; Ternès, N. A variable selection approach for highly correlated predictors in high-dimensional genomic data. *Bioinformatics* **2021**, *37*, 2238–2244. [CrossRef] [PubMed]
16. Zhao, P.; Yu, B. On model selection consistency of Lasso. *J. Mach. Learn. Res.* **2006**, *7*, 2541–2563.
17. Bühlmann, P.; Kalisch, M.; Maathuis, M.H. Variable selection in high-dimensional linear models: Partially faithful distributions and the PC-simple algorithm. *Biometrika* **2010**, *97*, 261–278. [CrossRef]
18. Fan, J.; Shao, Q.M.; Zhou, W.X. Are discoveries spurious Distributions of maximum spurious correlations and their applications. *Ann. Stat.* **2018**, *46*, 989–1017. [CrossRef]
19. Xue, F.; Qu, A. Semi-Standard partial covariance variable selection when irrepresentable conditions fail. *Stat. Sin.* **2022**, *32*, 1881–1909. [CrossRef]
20. Imbens, G.W. Nonparametric estimation of average treatment effects under exogeneity: A review. *Rev. Econ. Stat.* **2004**, *86*, 4–29. [CrossRef]
21. Freedman, D.A. On regression adjustments in experiments with several treatments. *Ann. Appl. Stat.* **2008**, *2*, 176–196. [CrossRef]

22. Lin, W. Agnostic notes on regression adjustments to experimental data: reexamining Freedman's critique. *Ann. Appl. Stat.* **2013**, *7*, 295–318. [CrossRef]
23. Lauritzen, S.L. *Graphical Models*; Clarendon Press: Oxford, UK, 1996.
24. Raveh, A. On the use of the inverse of the correlation matrix in multivariate data analysis. *Am. Stat.* **1985**, *39*, 39–42.
25. Cai, T.; Liu, W.; Luo, X. A constrained L_1 minimization approach to sparse precision matrix estimation. *J. Am. Stat. Assoc.* **2011**, *106*, 594–607. [CrossRef]
26. Balmand, S.; Dalalyan, A.S. On estimation of the diagonal elements of a sparse precision matrix. *Electron. J. Stat.* **2016**, *10*, 1551–1579. [CrossRef]
27. Avella-Medina, M.; Battey, H.S.; Fan, J.; Li, Q. Robust estimation of high-dimensional covariance and precision matrices. *Biometrika* **2018**, *105*, 271–284. [CrossRef] [PubMed]
28. Fan, J.; Peng, H. Nonconcave penalized likelihood with a diverging number of parameters. *Ann. Stat.* **2004**, *32*, 928–961. [CrossRef]
29. Blazère, M.; Loubes, J.M.; Gamboa, F. Oracle inequalities for a group Lasso procedure applied to generalized linear models in high dimension. *IEEE Trans. Inf. Theory* **2014**, *60*, 2303–2318. [CrossRef]
30. Pang, H.; Liu, H.; Vanderbei, R.J. The fastclime package for linear programming and large-scale precision matrix estimation in R. *J. Mach. Learn. Res.* **2014**, *15*, 489–493.
31. Gianni, L.; Eiermann, W.; Semiglazov, V.; Manikhas, A.; Lluch, A.; Tjulandin, S.; Zambetti, M.; Vazquez, F.; Byakhow, M.; Lichinitser, M.; et al. Neoadjuvant chemotherapy with trastuzumab followed by adjuvant trastuzumab versus neoadjuvant chemotherapy alone, in patients with HER2-positive locally advanced breast cancer (the NOAH trial): A randomised controlled superiority trial with a parallel HER2-negative cohort. *Lancet* **2010**, *375*, 377–384.
32. Prat, A.; Bianchini, G.; Thomas, M.; Belousov, A.; Cheang, M.C.; Koehler, A.; Gómez, P.; Semiglazov, V.; Eiermann, W.; Tjulandin, S.; et al. Research-based PAM50 subtype predictor identifies higher responses and improved survival outcomes in HER2-positive breast cancer in the NOAH study. *Clin. Cancer Res.* **2014**, *20*, 511–521. [CrossRef]
33. Roth, J.; Simon, N. A framework for estimating and testing qualitative interactions with applications to predictive biomarkers. *Biostatistics* **2018**, *19*, 263–280. [CrossRef]
34. Dudoit, S.; Fridlyand, J.; Speed, T.P. Comparison of discrimination methods for the classification of tumors using gene expression data. *J. Am. Stat. Assoc.* **2002**, *97*, 77–87. [CrossRef]