Journal of Machine Learning Research 27 (2026) 1-75 Submitted 3/25; Revised 11/25; Published 5/26

Statistical Inference for High-dimensional Partially Linear Models via Debiased Rank Lasso

Songshan Yang YANGSS@RUC.EDU.CN
Center for Applied Statistics, BRAIN and Institute of Statistics and Big Data
Renmin University of China
Beijing, 100872, China

Delin Zhao* DLZHAO@FZU.EDU.CN
School of Mathematics and Statistics
Fuzhou University
Fuzhou, 350108, China

Runze Li RZLI@PSU.EDU
Department of Statistics
Pennsylvania State University
University Park, PA 16802-2111, USA

Editor: Ji Zhu

Abstract

This paper aims to develop tuning-free and robust regularized methods for partially linear models based on partial residual methods. In order to preserve the near-oracle rate of rank Lasso, we construct a new estimator via a data-splitting procedure and name the resulting estimator as data-splitting (DS) rank Lasso, which is based on partial **prediction** residuals. Thus, the proposed estimation procedure is distinguished from the traditional partial residual based methods, which are based on the model-fitting residuals. We show that the DS rank Lasso enjoys several appealing features including tuning-free property and robust property to heavy-tailed random errors, and possesses near-oracle rate. We further propose statistical inference procedures for individual coefficients using the debiased Lasso techniques. The resulting estimator is termed the debiased DS rank Lasso estimator. This paper reveals an interesting theoretical finding: the DS rank Lasso estimate achieves near-oracle rate and the proposed debiased DS rank Lasso estimator does not suffer asymptotic efficiency loss due to the use of data-splitting. We further establish a simultaneous inference procedure based on multiplier bootstrap. We conduct Monte Carlo studies to assess finite sample performance of the proposed procedures, and illustrate the proposed methodology via an empirical analysis of a real-world data set.

Keywords: Data splitting, heavy-tailed error, tuning-free regularization, variable selection

1. Introduction

Statistical inference for high-dimensional linear and generalized linear models receives increasing attention in the recent literature. See, for example, Chapter 7 of Fan et al. (2020a) and references therein for a comprehensive account of this topic. Zhang and Zhang (2014)

*: Corresponding author.

and van de Geer et al. (2014) proposed debiased and desparsified Lasso to construct inference procedures for low-dimensional regression coefficients based on the Lasso estimator. A decorrelated score method was proposed by Ning and Liu (2017) to derive an inference procedure for low-dimensional regression coefficients based on penalized M-estimators. To have better finite sample performance, bootstrap-assisted procedures were used in Zhang and Cheng (2017) and Dezeure et al. (2017) to conduct simultaneous inference based on the debiased Lasso estimator. Inference on low-dimensional regression coefficients in high-dimensional linear models has been considered by Belloni et al. (2014), Cai and Guo (2017), Zhu and Bradic (2018a), Zhu and Bradic (2018b), Neykov et al. (2018), Javanmard and Montanari (2018), Chang et al. (2021) and among many others. To address the limitation that existing inference methods are not robust against heavy-tailed errors, several important works have proposed distinct approaches: Feng et al. (2013) developed a rank-based score test, Fan et al. (2020b) introduced a debiased rank Lasso estimator, and Cai et al. (2025) proposed a debiased convoluted rank Lasso estimator. This paper aims to develop robust statistical inference procedures for high-dimensional partially linear models.

Let Y be a response along with a p-dimensional covariate vector $\mathbf{x}$ and a univariate covariate Z. The partially linear model was first introduced by Engle et al. (1986) and is defined by

$$Y = \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0 + g_0(Z) + \epsilon, \tag{1}$$

where $g_0(\cdot)$ is a nonparametric baseline function, and ϵ is a random error. Compared with linear models and generalized linear models, there is much less work on inference for partially linear models. Zhu (2017) conducted a nonasymptotic analysis, for a two-step projection based Lasso estimator under the setting of $p \gg n$, and Zhu et al. (2019) extended the debiased Lasso for linear models to partially linear models by using partial residual techniques. Both works are based on penalized least squares with the Lasso penalty. It is well-known that least squares is not robust to outliers and heavy-tailed random errors. This motivates us to consider penalized rank regression (Wang et al., 2020) to achieve robustness. Furthermore, rank Lasso enjoys the tuning-free property for high-dimensional linear models.

In this paper, we first aim to develop a tuning-free and robust regularized regression method for partially linear models based on partial residual methods. In order to preserve the near-oracle rate of rank Lasso under mild conditions on nonparametric estimators, we construct a new estimator via a data-splitting procedure. Specifically, we randomly partition the entire sample into two subsamples with equal sizes, and employ the first-half subsample to estimate the nonparametric functions, and then we obtain a rank Lasso estimate based on the partial **prediction** residuals of the second-half subsample. See Section 2 for the definition of partial prediction residuals. We further switch the roles of the two subsamples and obtain another rank Lasso estimate. Thus, a natural estimator of $\boldsymbol{\beta}_0$ is the average of the two estimators from the data-splitting procedure. This estimator is termed data-splitting (DS) rank Lasso. We show that the DS rank Lasso enjoys several appealing properties including the tuning-free property and robustness to heavy-tailed random errors. We systematically study its asymptotic properties and establish its near-oracle rate.

Another goal of this paper is to develop robust statistical inference procedures for $\boldsymbol{\beta}_0$. We apply the debiased Lasso technique to the proposed DS rank Lasso and propose a new

estimator, termed the debiased DS rank Lasso estimator. We first establish the asymptotic normality of the debiased DS rank Lasso estimator. The asymptotic normality reveals an interesting theoretical finding: the proposed debiased DS rank Lasso estimator does not suffer asymptotic efficiency loss despite the use of data-splitting. Based on the asymptotic normality of the estimator, we can construct robust confidence intervals for either individual coefficients or any finite-dimensional subvector of $\boldsymbol{\beta}_0$. We further develop a simultaneous inference procedure based on the multiplier bootstrap. We show that the two proposed inference procedures are robust to heavy-tailed random errors. Furthermore, a comprehensive numerical study demonstrates the effectiveness of the proposed estimation and inference procedures.

This paper is organized as follows. In Section 2, we propose the DS rank Lasso estimator for high-dimensional partially linear models and establish its theoretical properties. In Section 3, we propose a debiased DS rank Lasso estimator for statistical inference for individual coefficients in Section 3.1 and then develop a simultaneous inference procedure in Section 3.2. Numerical comparisons are presented in Section 4. Some discussion and concluding remarks are given in Section 5. All technical proofs are included in the Appendix.

2. Rank Lasso for Partially Linear Models

It is assumed throughout this paper that $\mathbf{m}_{\mathbf{x}}(z) = E(\mathbf{x} \mid Z = z)$ is well-defined and finite. Furthermore, we focus on settings where $m_0(z) = E(Y \mid Z = z)$ is also well-defined and finite. Based on model (1), it follows that

$$E(Y \mid Z) = E(\mathbf{x} \mid Z)^{\mathrm{T}}\boldsymbol{\beta}_0 + g_0(Z), \tag{2}$$

since it is assumed that $E(\epsilon \mid Z) = 0$. It follows from (2) that

$$Y - m_0(Z) = \{\mathbf{x} - \mathbf{m}_{\mathbf{x}}(Z)\}^{\mathrm{T}}\boldsymbol{\beta}_0 + \epsilon.$$

This leads to the partial residual method for estimating $\boldsymbol{\beta}_0$. Specifically, denote $\widetilde{Y} = Y - m_0(Z)$ and $\widetilde{\mathbf{x}} = \mathbf{x} - \mathbf{m}_{\mathbf{x}}(Z)$. If both $m_0(Z)$ and $\mathbf{m}_{\mathbf{x}}(Z)$ were known, we could directly apply rank Lasso proposed in Wang et al. (2020) for

$$\widetilde{Y} = \widetilde{\mathbf{x}}^{\mathrm{T}}\boldsymbol{\beta}_0 + \epsilon.$$

In practice, both $E(Y \mid Z)$ and $E(\mathbf{x} \mid Z)$ are unknown. Intuitively, we may obtain estimates of $m_0(\cdot)$ and $\mathbf{m}_{\mathbf{x}}(\cdot)$ by regressing Y and $\mathbf{x}$ on Z, respectively. In particular, one may use nonparametric regression techniques to estimate $m_0(\cdot)$ and $\mathbf{m}_{\mathbf{x}}(\cdot)$ so that the forms of the regression functions are flexible. Once we obtain the estimated partial residuals from the nonparametric regression, we can apply the rank Lasso to these residuals. To allow flexibility in choosing the nonparametric estimators, we only assume the convergence rates of these estimators without specifying their exact forms. However, overfitting in the nonparametric regression may lead to poor performance of the rank Lasso if it uses the same data as the nonparametric regression. We propose using data-splitting to deal with this challenge.

2.1 Data-Splitting Rank Lasso

We randomly split the data into two equal halves, $\mathcal{D}_1$ and $\mathcal{D}_2$. For simplicity, assume the sample size N is even, let $n = N/2$, and denote $\mathcal{D}_1 = \left\{\mathbf{x}_i^{(1)}, Y_i^{(1)}, Z_i^{(1)}, i = 1, \ldots, n\right\}$ and

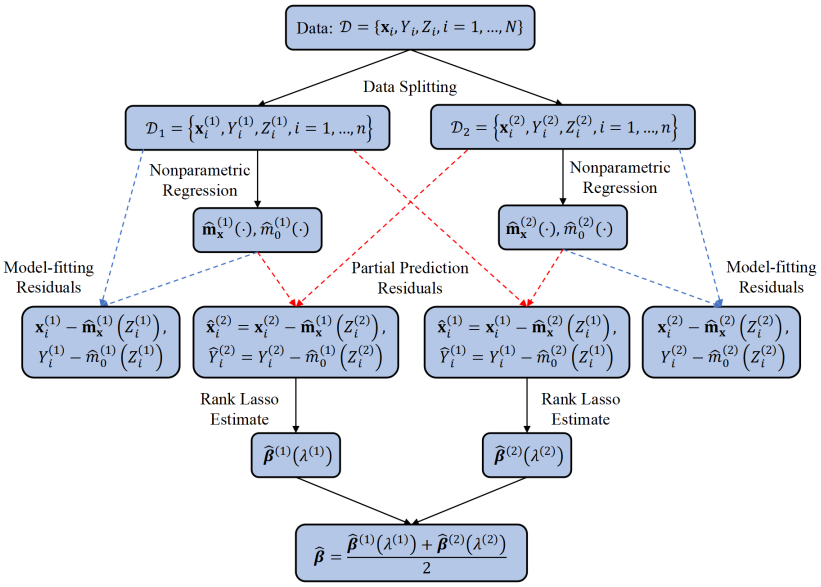


Figure 1: Flowchart of the procedure for constructing the DS rank Lasso estimator.

$\mathcal{D}_2 = \{\mathbf{x}_i^{(2)}, Y_i^{(2)}, Z_i^{(2)}, i = 1, \ldots, n\}$. The strategy of the data-splitting rank Lasso can be described as follows. We use the first half of the data to estimate $m_0(\cdot)$ and $\mathbf{m}_{\mathbf{x}}(\cdot)$, obtaining the nonparametric estimators $\widehat{m}_0^{(1)}(\cdot)$ and $\widehat{\mathbf{m}}_{\mathbf{x}}^{(1)}(\cdot)$. Next, we use the second half of the data to calculate the partial prediction residuals $\widehat{Y}_i^{(2)} = Y_i^{(2)} - \widehat{m}_0^{(1)}(Z_i^{(2)})$ and $\widehat{\mathbf{x}}_i^{(2)} = \mathbf{x}_i^{(2)} - \widehat{\mathbf{m}}_{\mathbf{x}}^{(1)}(Z_i^{(2)})$, $i = 1, \ldots, n$. Here, we introduce the term "**partial prediction residuals**" to distinguish these residuals $\{\widehat{Y}_i^{(2)}, \widehat{\mathbf{x}}_i^{(2)}\}_{i=1}^n$ from the model-fitting residuals $\{Y_i^{(1)} - \widehat{m}_0^{(1)}(Z_i^{(1)}), \mathbf{x}_i^{(1)} - \widehat{\mathbf{m}}_{\mathbf{x}}^{(1)}(Z_i^{(1)})\}_{i=1}^n$. These partial prediction residuals are then used to construct the rank Lasso estimator $\widehat{\boldsymbol{\beta}}^{(1)}(\lambda^{(1)})$, defined as

$$\widehat{\boldsymbol{\beta}}^{(1)}(\lambda^{(1)}) = \underset{\boldsymbol{\beta} \in \mathbb{R}^p}{\arg\min} \Big[\{n(n-1)\}^{-1} \sum_{i \neq j} |(\widehat{Y}_i^{(2)} - \boldsymbol{\beta}^{\mathrm{T}} \widehat{\mathbf{x}}_i^{(2)}) - (\widehat{Y}_j^{(2)} - \boldsymbol{\beta}^{\mathrm{T}} \widehat{\mathbf{x}}_j^{(2)})| + \lambda^{(1)} \|\boldsymbol{\beta}\|_1 \Big],$$

where $\lambda^{(1)}$ is a tuning parameter to be determined by a data-driven method. Similarly, we may switch the roles of $\mathcal{D}_1$ and $\mathcal{D}_2$ and obtain the second rank Lasso estimator $\widehat{\boldsymbol{\beta}}^{(2)}(\lambda^{(2)})$. The data-splitting (DS) rank Lasso estimator is then defined as

$$\widehat{\boldsymbol{\beta}} = \frac{\widehat{\boldsymbol{\beta}}^{(1)}(\lambda^{(1)}) + \widehat{\boldsymbol{\beta}}^{(2)}(\lambda^{(2)})}{2},$$

where we use $\lambda^{(1)}$ and $\lambda^{(2)}$ to allow flexibility in choosing the tuning parameters. Figure 1 illustrates the overall procedure for constructing the DS rank Lasso estimator.

We next show that the proposed DS rank Lasso enjoys the tuning-free property (Wang et al., 2020). Denote $g(z, \boldsymbol{\beta}_0) = E(Y \mid Z = z) - \boldsymbol{\beta}_0^{\mathrm{T}} E(\mathbf{x} \mid Z = z) = m_0(z) - \boldsymbol{\beta}_0^{\mathrm{T}} \mathbf{m}_{\mathbf{x}}(z) = g_0(z)$

and $\widehat{g}^{(1)}(z,\boldsymbol{\beta}_0)=\widehat{m}_0^{(1)}(z)-\boldsymbol{\beta}_0^{\mathrm{T}}\widehat{\mathbf{m}}_{\mathbf{x}}^{(1)}(z)$. Further define $e_i^{(2)}=\epsilon_i^{(2)}-\{\widehat{g}^{(1)}(Z_i^{(2)},\boldsymbol{\beta}_0)-g_0(Z_i^{(2)})\}$. Thus, $\widehat{m}_0^{(1)}(\cdot)$, $\widehat{m}_{\mathbf{x}}^{(1)}(\cdot)$ and $\widehat{g}^{(1)}(\cdot,\boldsymbol{\beta}_0)$ are independent of $\{\epsilon_i^{(2)},i=1,\ldots,n\}$. This is critical for DS rank Lasso to retain the tuning-free property of the rank Lasso for linear regression models. Denote $\boldsymbol{\gamma}=\boldsymbol{\beta}-\boldsymbol{\beta}_0$. By the definition of $\widehat{Y}_i^{(2)}$ and $\widehat{\mathbf{x}}_i^{(2)}$,

$$\begin{aligned}&\{n(n-1)\}^{-1}\sum_{i\neq j}|(\widehat{Y}_i^{(2)}-\boldsymbol{\beta}^{\mathrm{T}}\widehat{\mathbf{x}}_i^{(2)})-(\widehat{Y}_j^{(2)}-\boldsymbol{\beta}^{\mathrm{T}}\widehat{\mathbf{x}}_j^{(2)})|\\ =\;&\{n(n-1)\}^{-1}\sum_{i\neq j}|(e_i^{(2)}-e_j^{(2)})-(\widehat{\mathbf{x}}_i^{(2)}-\widehat{\mathbf{x}}_j^{(2)})^{\mathrm{T}}\boldsymbol{\gamma}|=:Q^{(2)}(\boldsymbol{\gamma}).\end{aligned}$$

Because $Q^{(2)}(\boldsymbol{\gamma})$ is piecewise linear in $\boldsymbol{\gamma}$, its gradient is defined everywhere excluding the linear boundaries. The gradient of $Q^{(2)}(\boldsymbol{\gamma})$ with respect to $\boldsymbol{\gamma}$ is

$$\begin{aligned}\mathbf{S}^{(2)}(\boldsymbol{\gamma})&=-\{n(n-1)\}^{-1}\sum_{i\neq j}(\widehat{\mathbf{x}}_i^{(2)}-\widehat{\mathbf{x}}_j^{(2)})\mathrm{sign}\{(e_i^{(2)}-e_j^{(2)})-(\widehat{\mathbf{x}}_i^{(2)}-\widehat{\mathbf{x}}_j^{(2)})^{\mathrm{T}}\boldsymbol{\gamma}\}\\&=-2\{n(n-1)\}^{-1}\sum_{i=1}^{n}\widehat{\mathbf{x}}_i^{(2)}\sum_{j=1,j\neq i}^{n}\mathrm{sign}\{(e_i^{(2)}-e_j^{(2)})-(\widehat{\mathbf{x}}_i^{(2)}-\widehat{\mathbf{x}}_j^{(2)})^{\mathrm{T}}\boldsymbol{\gamma}\}.\end{aligned}$$

Note that $\sum_{j=1,j\neq i}^{n}\mathrm{sign}\{(e_i^{(2)}-e_j^{(2)})-(\widehat{\mathbf{x}}_i^{(2)}-\widehat{\mathbf{x}}_j^{(2)})^{\mathrm{T}}\boldsymbol{\gamma}\}$ depends on the rank of $e_i^{(2)}-\boldsymbol{\gamma}^{\mathrm{T}}\widehat{\mathbf{x}}_i^{(2)}$ over $\{e_j^{(2)}-\boldsymbol{\gamma}^{\mathrm{T}}\widehat{\mathbf{x}}_j^{(2)},j=1,\cdots,n\}$. Motivated by the tuning parameter selection framework in Wang et al. (2020), we select $\lambda^{(1)}$ to satisfy

$$\mathrm{pr}\{\lambda^{(1)}\geq c\|\mathbf{S}^{(2)}(\mathbf{0})\|_\infty\}\geq 1-\alpha_0$$

for constants $c>1$ and $\alpha_0>0$. Under mild conditions on nonparametric estimators, $\mathbf{S}^{(2)}(\mathbf{0})$ and $\mathbf{S}_0^{(2)}$ (defined below) are asymptotically equivalent with a difference of $o_p(\sqrt{\log p/n})$:

$$\mathbf{S}_0^{(2)}=-2\{n(n-1)\}^{-1}\sum_{i=1}^{n}\widehat{\mathbf{x}}_i^{(2)}\sum_{j=1,j\neq i}^{n}\mathrm{sign}(\epsilon_i^{(2)}-\epsilon_j^{(2)}),$$

where $\{\sum_{j=1,j\neq i}^{n}\mathrm{sign}(\epsilon_i^{(2)}-\epsilon_j^{(2)})\}_{i=1}^{n}$ follows a uniform distribution over permutations of $2\{1,\ldots,n\}-(n+1)=\{-n+1,-n+3,\ldots,n-3,n-1\}$. Following Wang et al. (2020), let $G_{(1)}^{-1}(1-\alpha_0)$ denote the $(1-\alpha_0)$-quantile of the conditional distribution of $\|\mathbf{S}_0^{(2)}\|_\infty$ given $\widehat{\mathbf{x}}_1^{(2)},\ldots,\widehat{\mathbf{x}}_n^{(2)}$, where $\|\cdot\|_\infty$ is the infinity norm. The quantile $G_{(1)}^{-1}(1-\alpha_0)$ can be easily simulated. We then set the tuning parameter $\lambda^{(1)}$ for $\widehat{\boldsymbol{\beta}}^{(1)}(\lambda^{(1)})$ as

$$\widehat{\lambda}^{(1)}=c\Big\{G_{(1)}^{-1}(1-\alpha_0)+c_0\sqrt{\log p/n}\Big\},\tag{3}$$

where $c>1$, and $c_0,\alpha_0>0$ are pre-specified constants. By swapping the roles of $\mathcal{D}_1$ and $\mathcal{D}_2$, we analogously obtain $\widehat{\lambda}^{(2)}$ for $\widehat{\boldsymbol{\beta}}^{(2)}(\lambda^{(2)})$. The final DS rank Lasso estimator is

$$\widehat{\boldsymbol{\beta}}=\frac{\widehat{\boldsymbol{\beta}}^{(1)}(\widehat{\lambda}^{(1)})+\widehat{\boldsymbol{\beta}}^{(2)}(\widehat{\lambda}^{(2)})}{2}.$$

2.2 Rates of Convergence

Let $q = \|\boldsymbol{\beta}_0\|_0$, where $\|\cdot\|_0$ denotes the ℓ_0 norm. Define $\mathcal{B}$ as the index set of the non-zero components of $\boldsymbol{\beta}_0$, and introduce the following cone set:

$$\Gamma = \{\boldsymbol{\gamma} \in \mathbb{R}^p : \|\boldsymbol{\gamma}_{\mathcal{B}^c}\|_1 \leq \bar{c}\|\boldsymbol{\gamma}_{\mathcal{B}}\|_1\},$$

where $\bar{c} = (c+1)/(c-1)$ and c is the constant defined in (3). We denote the cumulative distribution function and probability density function of $\epsilon_i - \epsilon_j$ (for $i \neq j$) as $F^*(\cdot)$ and $f^*(\cdot)$, respectively.

To establish the theoretical properties of the DS Rank Lasso estimator, we impose the following regularity conditions.

(C1) There exist sequences $r_n > 0$ and $\rho_n = o(1)$ such that for $l \in \{1, 2\}$,

$$\text{pr}\left\{\max_{k\in\{0,1,\ldots,p\}} \sup_{z\in\mathcal{Z}} |\widehat{m}_k^{(l)}(z) - m_k(z)| \geq r_n\right\} \leq \rho_n.$$

Here, $\widehat{m}_k^{(l)}(z)$ and $m_k(z)$ are the k-th components of $\widehat{\mathbf{m}}_{\mathbf{x}}^{(l)}(z)$ and $\mathbf{m}_{\mathbf{x}}(z)$, respectively, where $k \in \{1, \ldots, p\}$. The sequence r_n satisfies $(\|\boldsymbol{\beta}_0\|_1 \vee 1)r_n = o(1)$ and $(\|\boldsymbol{\beta}_0\|_1 \vee 1)r_n^2 = o\big(\sqrt{\log p/n}\big)$, where $\|\cdot\|_1$ is the ℓ_1 norm of a vector, and $\mathcal{Z}$ is the support of Z_i.

(C2) We assume that $\widetilde{\mathbf{x}}_i = \mathbf{x}_i - \mathbf{m}_{\mathbf{x}}(Z_i)$ is sub-Gaussian with parameter σ. That is, for any unit vector $\boldsymbol{\alpha} \in \mathbb{R}^p$, the random variable $\boldsymbol{\alpha}^{\mathrm{T}}\widetilde{\mathbf{x}}_i$ is sub-Gaussian with parameter at most σ.

(C3) The population covariance matrix $\boldsymbol{\Sigma} = E(\widetilde{\mathbf{x}}_i\widetilde{\mathbf{x}}_i^{\mathrm{T}})$ is positive definite.

(C4) There exist constants $\delta_0 > 0$ and $\kappa_l, \kappa_u > 0$ such that $\inf_{t\in[-\delta_0,\delta_0]} f^*(t) \geq \kappa_l$ and $\sup_{t\in\mathbb{R}} f^*(t) \leq \kappa_u$. There exists some $L_1 > 0$ such that $|f^*(t_1) - f^*(t_2)| \leq L_1|t_1 - t_2|$ for any $t_1, t_2 \in \mathbb{R}$.

Condition (C1) plays a similar role to Assumption 5 in Zhu et al. (2019). It imposes requirements on the model parameters $(n, p, q, \|\boldsymbol{\beta}_0\|_1)$ and the uniform convergence rate r_n for the nonparametric estimators $\widehat{m}_k(\cdot)$. Although we focus on univariate Z_i in this paper, our results apply to $\mathbf{z}_i \in \mathbb{R}^d$ provided that Condition (C1) holds for r_n. For standard nonparametric regression estimators on a compact domain, r_n is typically of order $O[\{\log(p\vee n)/n\}^r]$ under Condition (C2), after applying a union bound to the $(p+1)$ estimators. For example, under certain regularity conditions (Masry, 1996; Li and Racine, 2007), local constant and local linear kernel estimators (Fan and Gijbels, 1996) can achieve a rate of $r_n = O[\{\log(p \vee n)/n\}^{2/5}]$ when $m_0(\cdot)$ and $\mathbf{m}_{\mathbf{x}}(\cdot)$ are twice continuously differentiable and $E(|\widetilde{Y}_i|^s) < +\infty$ for some $s > 2$. Consequently, Condition (C1) is satisfied for a wide range of $(n, p, q, \|\boldsymbol{\beta}_0\|_1)$. Notably, this moment condition on $\widetilde{Y}_i$ is much weaker than the sub-Gaussian assumption imposed in Zhu (2017) and Zhu et al. (2019). This condition is only required in the first-step for estimating the nonparametric functions and is not needed in the second-step for estimating $\boldsymbol{\beta}_0$. The uniform bound required in Condition (C1) can be relaxed to an ℓ_2 bound, that is, $\text{pr}\big(\max_{k\in\{0,1,\ldots,p\}} E[\{\widehat{m}_k^{(1)}(Z_i^{(2)}) - m_k(Z_i^{(2)})\}^2 \mid \mathcal{D}_1] \geq r_n^2\big) \leq \rho_n$. We assume the uniform bound to simplify the presentation and avoid technical complications.

Condition (C2) assumes that the design matrix is sub-Gaussian, a common assumption in high-dimensional regression (Fan et al., 2020a).

Condition (C3) is a standard assumption for high-dimensional design matrices and is relatively mild as it concerns the population covariance (van de Geer et al., 2014; Javanmard and Montanari, 2014; Zhang and Zhang, 2014; Ning and Liu, 2017; Zhang and Cheng, 2017). Moreover, we only require $\boldsymbol{\Sigma}$ to be positive definite, without imposing uniform boundedness on $\lambda_{\min}^{-1}(\boldsymbol{\Sigma})$ or $\|\boldsymbol{\Theta}\|_1$. Here, $\boldsymbol{\Theta} = \boldsymbol{\Sigma}^{-1}$, and $\|\boldsymbol{\Theta}\|_1 = \max_k \|\boldsymbol{\theta}_k\|_1$ is the ℓ_1 norm of $\boldsymbol{\Theta}$, where $\boldsymbol{\theta}_k$ is the k-th column of $\boldsymbol{\Theta}$. Note that since $\boldsymbol{\Sigma}$ is symmetric, we have $\|\boldsymbol{\Theta}\|_1 \geq \lambda_{\min}^{-1}(\boldsymbol{\Sigma})$. In our theoretical analysis, we allow both $\|\boldsymbol{\Theta}\|_1$ and $\lambda_{\min}^{-1}(\boldsymbol{\Sigma})$ to diverge at a certain rate as n increases.

The assumptions on random errors in Condition (C4) are substantially weaker than those required in existing literature for high-dimensional partially linear models. For example, Zhu (2017) and Zhu et al. (2019) impose sub-Gaussianity on random errors, thereby excluding heavy-tailed distributions such as t-distributions. In general, Condition (C4) is not restrictive and can be satisfied by a wide range of density functions. Two examples are provided in Section A.6 of the Appendix.

To establish the theoretical properties of $\widehat{\boldsymbol{\beta}}^{(l)}(\widehat{\lambda}^{(l)})$ for $l \in \{1, 2\}$, we first introduce the following lemma:

Lemma 1 *Under Conditions (C1), (C2) and (C4), and on the event* $\Big\{ \max_{k\in\{0,1,\ldots,p\}} \sup_{z\in\mathcal{Z}} |\widehat{m}_k^{(1)}(z) - m_k(z)| \leq r_n \Big\}$, *we have the following results:*

(i) There exists some $C_n > 0$ *such that* $pr\Big\{\widehat{\boldsymbol{\beta}}^{(1)}(\widehat{\lambda}^{(1)}) - \boldsymbol{\beta}_0 \in \Gamma \;\Big|\; \mathcal{D}_1\Big\} \geq 1 - \alpha_0 - 4p^{-(C_n^2-1)}$, *where* $C_n \to \infty$ *as* $n \to \infty$.

(ii) For any $C > 1$, *if* $\alpha_0 > 2p^{-(C^2-1)}$, *then* $pr\Big(\widehat{\lambda}^{(1)} \leq c_1\sqrt{\log p/n} \;\Big|\; \mathcal{D}_1\Big) \geq 1 - p\exp(-n)$, *where* $c_1 = c\{c_0 + 4\sqrt{2}C(2\sqrt{3}\sigma + r_n)\}$.

These results also hold when the roles of $\mathcal{D}_1$ *and* $\mathcal{D}_2$ *are switched.*

Part (i) of Lemma 1 indicates that $\widehat{\boldsymbol{\beta}}^{(l)}(\widehat{\lambda}^{(l)}) - \boldsymbol{\beta}_0$ belongs to the cone set Γ with high probability. Part (ii) implies that $\widehat{\lambda}^{(l)}$ has an upper bound of $c_1\sqrt{\log p/n}$ with probability tending to one. Then, the following theorem establishes the ℓ_1 and ℓ_2 convergence rates of $\widehat{\boldsymbol{\beta}}^{(l)}(\widehat{\lambda}^{(l)})$.

Theorem 2 *Under Conditions (C1)–(C4), for any* $C > 1$, *if* $\alpha_0 \geq 2p^{-(C^2-1)}$, *then* $\widehat{\boldsymbol{\beta}}^{(l)}(\widehat{\lambda}^{(l)})$ *satisfies*

$$\begin{aligned} \|\widehat{\boldsymbol{\beta}}^{(l)}(\widehat{\lambda}^{(l)}) - \boldsymbol{\beta}_0\|_2 &\leq \delta\sqrt{q\log p/n}, \\ \|\widehat{\boldsymbol{\beta}}^{(l)}(\widehat{\lambda}^{(l)}) - \boldsymbol{\beta}_0\|_1 &\leq (1+\bar{c})\delta q\sqrt{\log p/n}, \end{aligned}$$

with probability at least $1 - 8p^{-(C^2-1)} - c_2\exp(-c_3 n[\{\lambda_{\min}(\boldsymbol{\Sigma})/q\}^2/\sigma^4 \wedge 1] + \log p) - \rho_n - \alpha_0 - \alpha_n$ *for some universal constants* $c_2, c_3 > 0$, *provided that* $2\{2(C\sigma L_n + r_n)(1+\bar{c})q\delta\sqrt{\log p/n} + 2(\|\boldsymbol{\beta}_0\|_1 \vee 1)r_n\} \leq \delta_0$ *and* $16\kappa_u(\|\boldsymbol{\beta}_0\|_1 \vee 1)r_n^2 \leq c_0\sqrt{\log p/n}$, *where* $\delta = 2(1+\bar{c})[c_1 + 32C\{4\sigma(\kappa_u\delta_0 \vee 1) + r_n\}]/\{\kappa_l\lambda_{\min}(\boldsymbol{\Sigma})\}$, $L_n = \sqrt{\log(p \vee n)}$, $\alpha_n = 4p^{-(C_n^2-1)}$, *with* c_1 *and* C_n *defined in Lemma 1. Here,* $\|\cdot\|_2$ *denotes the* ℓ_2 *norm.*

Theorem 2 demonstrates that under mild conditions on nonparametric estimators and error distributions, the DS rank Lasso estimator achieves near-oracle rates comparable to those of the rank Lasso estimator in linear models (Wang et al., 2020), as if the unknown functions $m_0(\cdot)$ and $\mathbf{m}_{\mathbf{x}}(\cdot)$ were known. Although Zhu (2017) obtained similar convergence rates for a two-step projection based Lasso estimator, their analysis requires the sub-Gaussian error assumption, which is substantially more restrictive.

Remark 3 *Data-splitting allows the rank Lasso to achieve near-oracle rates under mild conditions on nonparametric estimators. Suppose we use the full sample to construct both the nonparametric estimators and the rank Lasso estimator. We define* $\mathbf{S}(\mathbf{0})$ *as the full-sample counterpart of* $\mathbf{S}^{(l)}(\mathbf{0})$*. Following the tuning parameter selection framework in Wang et al. (2020), we choose* λ *such that* $\lambda \geq \|\mathbf{S}(\mathbf{0})\|_\infty$ *with high probability. However, if overfitting occurs in the nonparametric estimators, the quantity* $\|\mathbf{S}(\mathbf{0})\|_\infty$ *may no longer scale as* $O_p(\sqrt{\log p/N})$*. Consequently, without data splitting, the* ℓ_2 *convergence rate of the rank Lasso estimator could fail to achieve the near-oracle rate of* $O_p(\sqrt{q \log p/N})$*. A toy example illustrating this phenomenon is provided in Section A.7 of the Appendix.*

3. Robust Statistical Inference

In this section, we introduce a one-step update to the DS rank Lasso estimator for statistical inference in high-dimensional partially linear models. We define the debiased estimator as

$$\begin{aligned}\widehat{\boldsymbol{\beta}}^{\mathrm{d},(l)} &= \widehat{\boldsymbol{\beta}}^{(l)} + \widehat{\boldsymbol{\Theta}}^{\mathrm{T}}\{n(n-1)\}^{-1}\sum_{i\neq j}\Big[(\widehat{\mathbf{x}}_i^{(l)} - \widehat{\mathbf{x}}_j^{(l)}) \\ &\qquad \times \mathrm{sign}\Big\{(\widehat{Y}_i^{(l)} - \widehat{Y}_j^{(l)}) - (\widehat{\mathbf{x}}_i^{(l)} - \widehat{\mathbf{x}}_j^{(l)})^{\mathrm{T}}\widehat{\boldsymbol{\beta}}^{(l)}\Big\}\Big]\Big/\{4\widehat{f}^*(0)\} \\ &= \widehat{\boldsymbol{\beta}}^{(l)} - \widehat{\boldsymbol{\Theta}}^{\mathrm{T}}\mathbf{S}^{(l)}(\widehat{\boldsymbol{\beta}}^{(l)} - \boldsymbol{\beta}_0)/\{4\widehat{f}^*(0)\}, \end{aligned} \tag{4}$$

for each $l \in \{1, 2\}$, where $\widehat{\boldsymbol{\beta}}^{(l)} = \widehat{\boldsymbol{\beta}}^{(l)}(\widehat{\lambda}^{(l)})$, and $\widehat{\boldsymbol{\Theta}}$ and $\widehat{f}^*(0)$ are estimators of $\boldsymbol{\Theta}$ and $f^*(0)$, respectively.

Estimator of $\boldsymbol{\Theta}$. For simplicity, we directly apply a modified version of the nodewise regression method proposed by van de Geer et al. (2014), as suggested in Zhu et al. (2019), to estimate $\boldsymbol{\Theta}$. As shown in the proofs of Zhu et al. (2019, Theorems 1 and A.1), this method effectively controls the error bound $\|\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}\|_1 = \max_k \|\widehat{\boldsymbol{\theta}}_k - \boldsymbol{\theta}_k\|_1$ under mild conditions, where $\widehat{\boldsymbol{\theta}}_k$ and $\boldsymbol{\theta}_k$ are the k-th columns of $\widehat{\boldsymbol{\Theta}}$ and $\boldsymbol{\Theta}$, respectively. However, it is important to note that any estimator capable of providing reasonable control over $\|\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}\|_1$ can be employed for this task.

Estimator of $f^*(0)$. Based on the observation that $f^*(0) = \int_{-\infty}^{\infty} f^2(t)\,\mathrm{d}t = E\{f(\epsilon_i)\}$, we estimate $f^*(0)$ by

$$\widehat{f}^{*,(l)}(0) = n^{-1}\sum_{i=1}^{n}\widehat{f}_{-i}^{(l)}(\widehat{\epsilon}_i^{(l)}), \tag{5}$$

where $\widehat{f}_{-i}^{(l)}(\widehat{\epsilon}_i^{(l)}) = (n-1)^{-1}\sum_{j=1,j\neq i}^{n} K((\widehat{\epsilon}_i^{(l)} - \widehat{\epsilon}_j^{(l)})/h)/h$ and $\widehat{\epsilon}_i^{(l)} = \widehat{Y}_i^{(l)} - (\widehat{\mathbf{x}}_i^{(l)})^{\mathrm{T}}\widehat{\boldsymbol{\beta}}^{(l)}$. Here, $K(\cdot)$ is a kernel function, and $h > 0$ is the bandwidth. The final estimator for $f^*(0)$ is defined as $\widehat{f}^*(0) = \big(\widehat{f}^{*,(1)}(0) + \widehat{f}^{*,(2)}(0)\big)/2$. For this estimator, we impose the following assumption:

(C5) The kernel function $K : \mathbb{R} \to [0, \infty)$ satisfies (i) $K(t) = K(-t)$ for any $t \in \mathbb{R}$, (ii) $\int_{-\infty}^{\infty} K(t)\,\mathrm{d}t = 1$, (iii) $\int_{-\infty}^{\infty} t^2 K(t)\,\mathrm{d}t < \infty$, (iv) $\int_{-\infty}^{\infty} K^2(t)\,\mathrm{d}t < \infty$, and (v) there exists some $L_2 > 0$ such that $|K(t_1) - K(t_2)| \le L_2|t_1 - t_2|$ for any $t_1, t_2 \in \mathbb{R}$. Furthermore, there exists some $L_3 > 0$ such that the first derivative of f^*, denoted $(f^*)'$, satisfies $|(f^*)'(t_1) - (f^*)'(t_2)| \le L_3|t_1 - t_2|$ for any $t_1, t_2 \in \mathbb{R}$.

Condition (C5) is a standard assumption in kernel density estimation.

The rationale for the proposed debiased estimators stems from the following linear approximation

$$\sqrt{n}\widehat{\boldsymbol{\Theta}}^{\mathrm{T}}\mathbf{S}^{(l)}(\widehat{\boldsymbol{\beta}}^{(l)} - \boldsymbol{\beta}_0)/\{4\widehat{f}^*(0)\} \approx \sqrt{n}\boldsymbol{\Theta}^{\mathrm{T}}\mathbf{S}^{(l)}(\mathbf{0})/\{4f^*(0)\} + \sqrt{n}(\widehat{\boldsymbol{\beta}}^{(l)} - \boldsymbol{\beta}_0).$$

While Hettmansperger and McKean (2010) and Fan et al. (2020b) derived similar expressions for fixed-dimensional linear models, we extend their results to partially linear models and allow the dimensionality p to diverge. Let $F(\cdot)$ denote the distribution function of ϵ_i. It can further be shown that the two terms $-\sqrt{n}\boldsymbol{\theta}_k^{\mathrm{T}}\mathbf{S}^{(l)}(\mathbf{0})/\{4f^*(0)\}$ and $n^{-1/2}\sum_{i=1}^n \boldsymbol{\theta}_k^{\mathrm{T}}\widetilde{\mathbf{x}}_i^{(l)}\{F(\epsilon_i^{(l)}) - 1/2\}/f^*(0)$ are asymptotically equivalent. The following lemma formally states these results.

Lemma 4 *Under Conditions (C1)–(C5),* $\|\boldsymbol{\Theta}\|_1 L_n q\sqrt{\log p/n} = o(1)$, $\|\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}\|_1 = o_p(1)$, $\|\widehat{\boldsymbol{\beta}}^{(l)}\|_0 = O_p(q)$, *and* $h \asymp \{\|\boldsymbol{\Theta}\|_1 q\log(p)/n\}^{1/4} \vee n^{-1/5}$, $\widehat{\boldsymbol{\beta}}^{\mathrm{d},(l)}$ *satisfies*

$$\widehat{\boldsymbol{\beta}}^{\mathrm{d},(l)} - \boldsymbol{\beta}_0 = -\boldsymbol{\Theta}^{T}\mathbf{S}^{(l)}(\mathbf{0})/\{4f^*(0)\} + \mathbf{R}_{n,1}^{(l)},$$

with

$$\begin{aligned}\|\mathbf{R}_{n,1}^{(l)}\|_\infty \;=\; O_p\Big[\|\boldsymbol{\Theta}\|_1\sqrt{q\log p/n}\Big\{\sqrt{q}\|\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}\|_1 \vee \|\boldsymbol{\Theta}\|_1^2 L_n\sqrt{q\log p/n} \\ \vee\|\boldsymbol{\Theta}\|_1(\|\boldsymbol{\beta}_0\|_1 \vee 1)r_n\Big\} \bigvee \|\boldsymbol{\Theta}\|_1^{3/2} L_n^{3/2}\{\log(p)\}^{1/4}(q/n)^{3/4}\Big].\end{aligned}$$

Furthermore, we have

$$-\boldsymbol{\Theta}^{T}\mathbf{S}^{(l)}(\mathbf{0})/\{4f^*(0)\} = n^{-1}\sum_{i=1}^n \boldsymbol{\Theta}^{T}\widetilde{\mathbf{x}}_i^{(l)}\{F(\epsilon_i^{(l)}) - 1/2\}/f^*(0) + \mathbf{R}_{n,2}^{(l)},$$

with

$$\|\mathbf{R}_{n,2}^{(l)}\|_\infty = O_p\Big[\|\boldsymbol{\Theta}\|_1\Big\{(\|\boldsymbol{\beta}_0\|_1 \vee 1)r_n^2 \vee \sqrt{(\|\boldsymbol{\beta}_0\|_1 \vee 1)r_n\log p/n} \vee L_n\log p/n\Big\}\Big].$$

If we were to use only one of the estimators, $\widehat{\boldsymbol{\beta}}^{\mathrm{d},(1)}$ or $\widehat{\boldsymbol{\beta}}^{\mathrm{d},(2)}$, for statistical inference, this would result in a loss of efficiency, since each estimator asymptotically uses only half of the data. However, full efficiency can be recovered by averaging $\widehat{\boldsymbol{\beta}}^{\mathrm{d},(1)}$ and $\widehat{\boldsymbol{\beta}}^{\mathrm{d},(2)}$. The intuition is that, by Lemma 4, $\widehat{\boldsymbol{\beta}}^{\mathrm{d},(l)} - \boldsymbol{\beta}_0$ for $l \in \{1, 2\}$ decomposes into a leading term, $n^{-1}\sum_{i=1}^n \boldsymbol{\Theta}^{\mathrm{T}}\widetilde{\mathbf{x}}_i^{(l)}\{F(\epsilon_i^{(l)}) - 1/2\}/f^*(0)$, plus a remainder term that is asymptotically negligible. The two leading terms are independent and have mean zero, so $\widehat{\boldsymbol{\beta}}^{\mathrm{d},(1)}$ and $\widehat{\boldsymbol{\beta}}^{\mathrm{d},(2)}$ are asymptotically independent and unbiased. Therefore, we define the final debiased DS rank Lasso estimator as

$$\widehat{\boldsymbol{\beta}}^{\mathrm{d}} = \frac{\widehat{\boldsymbol{\beta}}^{\mathrm{d},(1)} + \widehat{\boldsymbol{\beta}}^{\mathrm{d},(2)}}{2}, \tag{6}$$

and in the following section, we show that this debiased estimator achieves the same asymptotic variance as a full-sample oracle estimator. This means that data-splitting does not lead to any asymptotic efficiency loss in our setting.

3.1 Inference for Individual Coefficients

In Theorem 5, we present the theoretical properties of $\widehat{\beta}_k^{\mathrm{d}}$, the k-th component of the debiased estimator $\widehat{\boldsymbol{\beta}}^{\mathrm{d}}$.

Theorem 5 *Under the conditions of Lemma 4, if in addition* $\|\boldsymbol{\Theta}\|_1^3 L_n^3 \{\log(p)\}^{1/2} q^{3/2} n^{-1/2} = o(1)$, $\|\boldsymbol{\Theta}\|_1^2(\|\boldsymbol{\beta}_0\|_1 \vee 1) r_n(\sqrt{q \log p} \vee \log p) = o(1)$, $\sqrt{n}\|\boldsymbol{\Theta}\|_1(\|\boldsymbol{\beta}_0\|_1 \vee 1) r_n^2 = o(1)$, *and* $\|\boldsymbol{\Theta}\|_1 q\sqrt{\log p}\|\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}\|_1 = o_p(1)$, *the debiased DS rank Lasso estimator* $\widehat{\beta}_k^{\mathrm{d}}$ *satisfies*

$$\sqrt{N}(\widehat{\beta}_k^{\mathrm{d}} - \beta_{0,k})/\widehat{\omega}_k \xrightarrow{d} \mathcal{N}(0,1),$$

for each $k \in \{1, \ldots, p\}$, *where*

$$\widehat{\omega}_k^2 = \widehat{\boldsymbol{\theta}}_k^T \Big(\frac{\widehat{\boldsymbol{\Sigma}}^{(1)} + \widehat{\boldsymbol{\Sigma}}^{(2)}}{2} \Big) \widehat{\boldsymbol{\theta}}_k \Big/ [12\{\widehat{f}^*(0)\}^2],$$

with $\widehat{\boldsymbol{\Sigma}}^{(l)} = n^{-1} \sum_{i=1}^n (\widehat{\mathbf{x}}_i^{(l)} - \overline{\widehat{\mathbf{x}}}^{(l)})(\widehat{\mathbf{x}}_i^{(l)} - \overline{\widehat{\mathbf{x}}}^{(l)})^T$ *and* $\overline{\widehat{\mathbf{x}}}^{(l)} = n^{-1} \sum_{i=1}^n \widehat{\mathbf{x}}_i^{(l)}$.

Remark 6 *Without data-splitting, some remainder terms in* $\sqrt{N}(\widehat{\beta}_k^{\mathrm{d}} - \beta_{0,k})$ *may not vanish if overfitting occurs in the nonparametric estimators. This can lead to poor performance of* $\widehat{\beta}_k^{\mathrm{d}}$. *A toy example, as used in Remark 3, is provided in Appendix A.7 to illustrate this issue.*

Based on Theorem 5, we construct the $(1-\alpha)$-confidence interval for $\beta_{0,k}$ as follows:

$$\Big[\widehat{\beta}_k^{\mathrm{d}} - \Phi^{-1}(1-\alpha/2)\widehat{\omega}_k/\sqrt{N}, \quad \widehat{\beta}_k^{\mathrm{d}} + \Phi^{-1}(1-\alpha/2)\widehat{\omega}_k/\sqrt{N} \Big],$$

where $\Phi^{-1}(1-\alpha/2)$ denotes the $(1-\alpha/2)$-quantile of the standard normal distribution.

Theorem 5 implies that under mild conditions, the asymptotic distribution of $\widehat{\beta}_k^{\mathrm{d}}$ behaves as if the unknown functions $m_0(\cdot)$ and $\mathbf{m}_{\mathbf{x}}(\cdot)$ were known. To elaborate, if $m_0(\cdot)$ and $\mathbf{m}_{\mathbf{x}}(\cdot)$ were known, we could define the debiased full-sample (FS) oracle rank Lasso estimator as:

$$\widetilde{\boldsymbol{\beta}}^{\mathrm{d}} = \widetilde{\boldsymbol{\beta}} + \widetilde{\boldsymbol{\Theta}}^{\mathrm{T}} \{N(N-1)\}^{-1} \sum_{i \neq j} (\widetilde{\mathbf{x}}_i - \widetilde{\mathbf{x}}_j) \mathrm{sign}\{(\widetilde{Y}_i - \widetilde{Y}_j) - (\widetilde{\mathbf{x}}_i - \widetilde{\mathbf{x}}_j)^{\mathrm{T}} \widetilde{\boldsymbol{\beta}}\} / \{4\widetilde{f}^*(0)\},$$

where $\widetilde{\boldsymbol{\beta}}$ and $\widetilde{\boldsymbol{\Theta}}$ are, respectively, the ordinary rank Lasso estimator (Wang et al., 2020) and the nodewise Lasso estimator (van de Geer et al., 2014) for $\boldsymbol{\beta}_0$ and $\boldsymbol{\Theta}$, based on $\{\widetilde{\mathbf{x}}_i, \widetilde{Y}_i, i = 1, \ldots, N\}$, and $\widetilde{f}^*(0) = \{N(N-1)\}^{-1} \sum_{i \neq j}^N K[\{(\widetilde{Y}_i - \widetilde{Y}_j) - (\widetilde{\mathbf{x}}_i - \widetilde{\mathbf{x}}_j)^{\mathrm{T}} \widetilde{\boldsymbol{\beta}}\}/h]/h$. Following arguments similar to those in the proofs of Lemma 4 and Theorem 5, it can be shown that

$$\sqrt{N}(\widetilde{\beta}_k^{\mathrm{d}} - \beta_{0,k})/\widetilde{\omega}_k \xrightarrow{d} \mathcal{N}(0,1),$$

where $\widetilde{\omega}_k^2 = \widetilde{\boldsymbol{\theta}}_k^{\mathrm{T}}\widetilde{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{\theta}}_k/[12\{\widetilde{f}^*(0)\}^2]$, $\widetilde{\boldsymbol{\theta}}_k$ is the k-th column of $\widetilde{\boldsymbol{\Theta}}$, $\widetilde{\boldsymbol{\Sigma}} = N^{-1}\sum_{i=1}^N(\widetilde{\mathbf{x}}_i - \overline{\widetilde{\mathbf{x}}})(\widetilde{\mathbf{x}}_i - \overline{\widetilde{\mathbf{x}}})^{\mathrm{T}}$ and $\overline{\widetilde{\mathbf{x}}} = N^{-1}\sum_{i=1}^N \widetilde{\mathbf{x}}_i$. Both $\widehat{\omega}_k^2/\omega_k^2$ and $\widetilde{\omega}_k^2/\omega_k^2$ converge in probability to 1, where $\omega_k^2 = \Theta_{k,k}/[12\{f^*(0)\}^2]$. This implies that the proposed debiased DS rank Lasso estimator achieves the same asymptotic variance as the debiased FS oracle rank Lasso estimator. The intuition is that, by Lemma 4, $\sqrt{N}(\widetilde{\boldsymbol{\beta}}^{\mathrm{d}} - \boldsymbol{\beta}_0)$ decomposes into a leading term, $N^{-1/2}\sum_{i=1}^N \boldsymbol{\Theta}^{\mathrm{T}}\widetilde{\mathbf{x}}_i\{F(\epsilon_i) - 1/2\}/f^*(0)$, plus remainder terms; similarly, $\sqrt{N}(\widehat{\boldsymbol{\beta}}^{\mathrm{d}} - \boldsymbol{\beta}_0)$ decomposes into the same leading term plus its own remainder terms. These remainder terms can be shown to vanish asymptotically under the conditions of Theorem 5. In Section B.1 of the Appendix, we provide simulation studies comparing inference based on the debiased DS rank Lasso estimator with that based on the debiased FS oracle rank Lasso estimator. The results demonstrate that data splitting introduces only minor efficiency loss in finite samples when the sample size is moderately large.

As shown in Zhu et al. (2019, Theorem 1), the debiased two-step projection-based Lasso estimator $\widehat{\boldsymbol{\beta}}^{\mathrm{d}}_{\mathrm{Lasso}}$ satisfies

$$\sqrt{n}(\widehat{\beta}^{\mathrm{d}}_{k,\mathrm{Lasso}} - \beta_{0,k})/\widehat{\omega}_{k,\mathrm{Lasso}} \xrightarrow{d} \mathcal{N}(0,1),$$

under the sub-Gaussian assumption for ϵ_i, where $\widehat{\omega}^2_{k,\mathrm{Lasso}}/\omega^2_{k,\mathrm{Lasso}} \xrightarrow{p} 1$ for $\omega^2_{k,\mathrm{Lasso}} = \Theta_{k,k}E(\epsilon_i^2)$. Consequently, the asymptotic relative efficiency (ARE) of our estimator $\widehat{\beta}^{\mathrm{d}}_k$ relative to $\widehat{\beta}^{\mathrm{d}}_{k,\mathrm{Lasso}}$ is $12\{f^*(0)\}^2E(\epsilon_i^2)$. This ARE can be substantially greater than one for heavy-tailed error distributions (Hettmansperger and McKean, 2010; Wang et al., 2020). For example, when ϵ_i follows a mixture-normal distribution with density $f_\epsilon(x) = (1-\varepsilon)\phi(x) + \varepsilon\phi(x/3)/3$ and ε varies in $\{0.00, 0.01, 0.03, 0.05, 0.10, 0.15\}$, where $\phi(\cdot)$ is the density of the standard normal distribution, the corresponding ARE values are $\{0.955, 1.009, 1.108, 1.196, 1.373, 1.497\}$ (Hettmansperger and McKean, 2010, Table 1.7.1). This indicates that the proposed debiased DS rank Lasso estimator offers greater robustness against heavy-tailed errors compared to the debiased two-step projection-based Lasso estimator.

It is worth mentioning that the initial DS rank Lasso estimator can be replaced by its ANOPE subsampled version, as suggested by Fan et al. (2020b). Specifically, the loss function $Q^{(l)}(\boldsymbol{\gamma})$ can be replaced by

$$Q^{(l),\mathrm{sub}}(\boldsymbol{\gamma}) = (M_n|\mathcal{S}|)^{-1}\sum_{\pi\in\mathcal{S}}\sum_{i=1}^{M_n}|(e^{(2)}_{\pi(i)} - e^{(2)}_{\pi(M_n+i)}) - (\widehat{\mathbf{x}}^{(2)}_{\pi(i)} - \widehat{\mathbf{x}}^{(2)}_{\pi(M_n+i)})^{\mathrm{T}}\boldsymbol{\gamma}|,$$

where $\mathcal{S}$ is sampled with replacement from permutations of $\{1,\ldots,n\}$, $|\mathcal{S}|$ denotes its cardinality, and $M_n = \lfloor n/2 \rfloor$ is the largest integer smaller than or equal to $n/2$. This subsampling method can further reduce computational complexity. Our simulation studies in Section 4.1 demonstrate that ANOPE only results in a negligible loss of statistical efficiency.

In this paper, we randomly split the data into two equal halves to construct a two-split debiased estimator. It is also possible to partition the data into $K \geq 3$ equally sized subsamples. For $l \in \{1,\ldots,K\}$, we compute the nonparametric estimators using all subsamples except the l-th one and then obtain the corresponding debiased estimator similar to (4). The final debiased estimator is then obtained by averaging the K debiased estimators. Similar to the two-split case, the error of the K-split debiased estimator can be decomposed into an $O_p(N^{-1/2})$ asymptotically normal term and a remainder term. For

fixed K, the remainder term is asymptotically negligible, and the K-split debiased estimator shares the same asymptotic normality as $\widetilde{\boldsymbol{\beta}}^{\mathrm{d}}$. Although theoretically valid, this approach is less computationally attractive, as it requires performing K nonparametric regressions without improving asymptotic efficiency compared to the two-split estimator.

Alternatively, one may consider unbalanced splitting instead of balanced splitting. In such cases, the final debiased estimator is a weighted average of the subsample-specific estimators, with weights proportional to their respective sample sizes. Theoretical analysis for unbalanced splitting follows similar reasoning as in the balanced case, and the resulting estimator exhibits analogous asymptotic properties. Since unbalanced splitting likewise does not improve asymptotic efficiency, we adopt the two-split debiased estimator with balanced splitting in this article for simplicity.

3.2 Simultaneous Inference

In high-dimensional settings, researchers may be interested in conducting a simultaneous hypothesis test for a set of coefficients indexed by $\mathcal{G} \subseteq \{1, \ldots, p\}$:

$$H_{0,\mathcal{G}} : \beta_{0,k} = \beta_k^*, \text{ for all } k \in \mathcal{G},$$

versus

$$H_{a,\mathcal{G}} : \beta_{0,k} \neq \beta_k^*, \text{ for some } k \in \mathcal{G},$$

where $\{\beta_k^*\}_{k\in\mathcal{G}}$ are prespecified values. We allow the cardinality of $\mathcal{G}$, denoted by $|\mathcal{G}|$, to grow as fast as p, which itself may grow exponentially with the sample size. To test $H_{0,\mathcal{G}}$, we use two test statistics: the non-studentized statistic $T_{\mathcal{G}} = \max_{k\in\mathcal{G}} \sqrt{N}|\widehat{\beta}_k^{\mathrm{d}} - \beta_k^*|$ and the studentized statistic $\overline{T}_{\mathcal{G}} = \max_{k\in\mathcal{G}} \sqrt{N}|\widehat{\beta}_k^{\mathrm{d}} - \beta_k^*|/\widehat{\omega}_k$, where $\widehat{\omega}_k^2$ is defined in Theorem 5. Inspired by Zhang and Cheng (2017) and Zhu et al. (2019), and based on the asymptotic expansions in Lemma 4, we propose two multiplier bootstrap statistics:

$$\begin{aligned} W_{\mathcal{G}} &= \max_{k\in\mathcal{G}} \Big| \sum_{l=1}^{2} \sum_{i=1}^{n} \widehat{\boldsymbol{\theta}}_k^{\mathrm{T}} (\widehat{\mathbf{x}}_i^{(l)} - \overline{\widehat{\mathbf{x}}}^{(l)}) u_i^{(l)} \Big| \Big/ [\sqrt{12N}\{\widehat{f}^*(0)\}], \\ \overline{W}_{\mathcal{G}} &= \max_{k\in\mathcal{G}} \Big| \sum_{l=1}^{2} \sum_{i=1}^{n} \widehat{\boldsymbol{\theta}}_k^{\mathrm{T}} (\widehat{\mathbf{x}}_i^{(l)} - \overline{\widehat{\mathbf{x}}}^{(l)}) u_i^{(l)} \Big| \Big/ [\sqrt{12N}\{\widehat{f}^*(0)\}\widehat{\omega}_k], \end{aligned}$$

where $\overline{\widehat{\mathbf{x}}}^{(l)} = n^{-1}\sum_{i=1}^{n} \widehat{\mathbf{x}}_i^{(l)}$, and $u_1^{(1)}, \ldots, u_n^{(1)}, u_1^{(2)}, \ldots, u_n^{(2)}$ are independent standard normal random variables, independent of $\{\mathbf{x}_i, Y_i, Z_i\}_{i=1}^N$. Then, the corresponding bootstrap critical values are given by $c_{\mathcal{G}}(\alpha) = \inf\{t \in \mathbb{R} : \mathrm{pr}[W_{\mathcal{G}} \leq t \mid \{\mathbf{x}_i, Y_i, Z_i\}_{i=1}^N] \geq 1-\alpha\}$ and $\overline{c}_{\mathcal{G}}(\alpha) = \inf\{t \in \mathbb{R} : \mathrm{pr}[\overline{W}_{\mathcal{G}} \leq t \mid \{\mathbf{x}_i, Y_i, Z_i\}_{i=1}^N] \geq 1-\alpha\}$ for any $\alpha \in (0,1)$. We reject $H_{0,\mathcal{G}}$ if $T_{\mathcal{G}} > c_{\mathcal{G}}(\alpha)$, or $\overline{T}_{\mathcal{G}} > \overline{c}_{\mathcal{G}}(\alpha)$ at significance level α. Furthermore, we can construct two simultaneous confidence intervals for $\{\beta_{0,k}\}_{k\in\mathcal{G}}$ as $[\widehat{\beta}_k^{\mathrm{d}} - c_{\mathcal{G}}(\alpha)/\sqrt{N}, \widehat{\beta}_k^{\mathrm{d}} + c_{\mathcal{G}}(\alpha)/\sqrt{N}]$, and $[\widehat{\beta}_k^{\mathrm{d}} - \widehat{\omega}_k\overline{c}_{\mathcal{G}}(\alpha)/\sqrt{N}, \widehat{\beta}_k^{\mathrm{d}} + \widehat{\omega}_k\overline{c}_{\mathcal{G}}(\alpha)/\sqrt{N}]$. Note that the first simultaneous confidence intervals have equal lengths across all $k \in \mathcal{G}$, while the second can have varying lengths.

The validity of these bootstrap-based methods is established by the following theorem.

Theorem 7 *Under the conditions of Lemma 4, if further* $\|\mathbf{\Theta}\|_1^3 L_n^3 \{\log(p)\}^{3/2} q^{3/2} n^{-1/2} = o(1)$, $\|\mathbf{\Theta}\|_1^2(\|\boldsymbol{\beta}_0\|_1 \vee 1) r_n[\sqrt{q}\log p \vee \{\log(p)\}^2] = o(1)$, $\|\mathbf{\Theta}\|_1\sqrt{n\log p}(\|\boldsymbol{\beta}_0\|_1 \vee 1) r_n^2 = o(1)$, $\|\mathbf{\Theta}\|_1^6 L_n^{14}/n \leq Mn^{-\iota}$ *for some* $M, \iota > 0$, *and there exists some* $\zeta_n \to +\infty$ *such that* $\zeta_n[\|\mathbf{\Theta}\|_1 q \log p \vee \{\log(p)\}^2]\|\widehat{\mathbf{\Theta}} - \mathbf{\Theta}\|_1 = o_p(1)$, *we have*

$$\sup_{\alpha\in(0,1)} \Big| pr\Big\{ \max_{k\in\mathcal{G}} \sqrt{N}|\widehat{\beta}_k^{\mathrm{d}} - \beta_{0,k}|/\widehat{\omega}_k > \overline{c}_{\mathcal{G}}(\alpha) \Big\} - \alpha \Big| = o(1).$$

If in addition $\max_{k\in\{1,\ldots,p\}} \Theta_{k,k} < +\infty$, *we have*

$$\sup_{\alpha\in(0,1)} \Big| pr\Big\{ \max_{k\in\mathcal{G}} \sqrt{N}|\widehat{\beta}_k^{\mathrm{d}} - \beta_{0,k}| > c_{\mathcal{G}}(\alpha) \Big\} - \alpha \Big| = o(1).$$

Theorem 7 shows that the coverage probabilities of the two types of simultaneous confidence intervals converge to the nominal level $1-\alpha$. It also implies that under $H_{0,\mathcal{G}}$, $\sup_{\alpha\in(0,1)} |\mathrm{pr}\{T_{\mathcal{G}} > c_{\mathcal{G}}(\alpha)\} - \alpha| = o(1)$ and $\sup_{\alpha\in(0,1)} |\mathrm{pr}\{\overline{T}_{\mathcal{G}} > \overline{c}_{\mathcal{G}}(\alpha)\} - \alpha| = o(1)$, indicating that the bootstrap-based tests have correct size under the null hypothesis. Furthermore, Theorem 7 enables power analysis of the two tests, analogous to Zhang and Cheng (2017, Theorem 2.3 and Theorem 2.4).

4. Numerical Studies

This section presents numerical studies to evaluate the performance of the proposed methods. Specifically, we demonstrate the effectiveness of the proposed estimation and inference methods through simulation studies in Section 4.1. In Section 4.2, we examine the performance of the proposed methods in solving real-world problems.

4.1 Simulation Studies

In the simulation study, covariates $\mathbf{x}_i$ $(i = 1, \ldots, N)$ are generated as follows. First, $\mathbf{x}_{0,i}$ is sampled from a p-dimensional normal distribution with mean $\mathbf{0}$ and covariance matrix $\mathbf{\Sigma} = (\Sigma_{k,k'})_{p\times p}$, where $\Sigma_{k,k'} = 0.5^{|k-k'|}$. Then, Z_i is generated from a uniform distribution on the interval $[0, 2]$. Finally, the covariates $X_{i,k}$ are defined as: $X_{i,1} = X_{0,i,1} + 3Z_i$, $X_{i,2} = X_{0,i,2} + 3Z_i^2$, $X_{i,3} = X_{0,i,3} - 3Z_i$, and $X_{i,k} = X_{0,i,k}$ for $4 \leq k \leq p$, for all $1 \leq i \leq N$.

Following Example 1 in Wang et al. (2020), we define the active set as $\mathcal{B} = \{1, 2, 3\}$ and set the regression coefficient $\boldsymbol{\beta}_0$ to be $(\sqrt{3}, \sqrt{3}, \sqrt{3}, 0, \ldots, 0)^{\mathrm{T}}$. For the nonlinear nuisance function, we use $g_0(z) = 1.5\sin(2\pi z)$ in this section. The sensitivity of the proposed inference methods to different forms of $g_0(z)$ is further examined in Section B.3 of the Appendix. To evaluate the performance of the proposed methods, we consider the following error distributions for ϵ_i:

(i) Normal distribution: $\epsilon_i \sim \mathcal{N}(0, 2^2)$;

(ii) Standard log-normal (LogN) distribution: $\log(\epsilon_i) \sim \mathcal{N}(0, 1)$;

(iii) Mixture-Normal (MN) distribution: $\epsilon_i \sim 0.95\mathcal{N}(0, 1) + 0.05\mathcal{N}(0, 10^2)$;

(iv) Scaled t-distribution: $\epsilon_i/\sqrt{2} \sim t(2)$, where $t(2)$ denotes the t-distribution with 2 degrees of freedom;

(v) Standard Cauchy distribution: $\epsilon_i \sim \text{Cauchy}(0, 1)$.

We then simulate Y_i based on model (1). We set the sample size $N \in \{300, 450\}$ and the dimensionality $p = 500$. We consider both the DS rank Lasso estimator (denoted as DSRLasso) and its ANOPE (Fan et al., 2020b) subsampled version (denoted as DSRLasso$^{\text{sub}}$). To estimate $m_0(\cdot)$, we use local linear regression (Fan and Gijbels, 1996). For $\mathbf{m}_{\mathbf{x}}(\cdot)$, where p estimators are required, we apply local constant regression due to its faster computational speed. The sensitivity of the proposed methods to the choice of nonparametric estimators is investigated in Section B.4 of the Appendix.

In Example 1, we compare the proposed DSRLasso estimator with the two-step projection based Lasso estimator (denoted as TSLasso; Zhu, 2017). In Examples 2 and 3, we further compare the performance of two debiased procedures: the proposed debiased DSRLasso and the debiased TSLasso (Zhu et al., 2019), with Example 2 focusing on inference for individual coefficients and Example 3 on simultaneous inference. For the TSLasso-based methods, we follow Zhu et al. (2019) and use smoothing splines for the first-step nonparametric projection. In these examples, we observe that although the moment condition on $\widetilde{Y}_i$ required by local regression fails to hold under $t(2)$ and Cauchy errors, the proposed DSRLasso and debiased DSRLasso estimators still exhibit reasonable performance. This contrasts sharply with the degraded performance of the TSLasso and debiased TSLasso estimators. This discrepancy arises because the moment condition is necessary only for the first-step local regression, whereas the second-step penalized rank regression does not require it. In contrast, the sub-Gaussian assumption on $\widetilde{Y}_i$ is crucial for both the first-step regularized nonparametric least squares and the second-step penalized least squares in the TSLasso methods (Zhu, 2017; Zhu et al., 2019).

For high-dimensional linear models, aside from the debiased rank Lasso estimator (Fan et al., 2020b), another robust inference method that does not rely on sub-Gaussian or moment conditions for random errors is the one based on the debiased Lasso-penalized quantile, particularly median, estimator (Bradic and Kolar, 2017). However, to the best of our knowledge, median-based inference methods have not been developed for high-dimensional partially linear models. Moreover, median-based methods have been shown to incur significant efficiency loss for light-tailed errors (e.g., normal errors) compared to rank-based methods (Hettmansperger and McKean, 2010; Wang et al., 2020; Cai et al., 2025). In Section B.5 of the Appendix, we show this efficiency loss persists when the debiased Lasso-penalized median estimator is extended to high-dimensional partially linear models.

Following the recommendations of Wang et al. (2020) and Fan et al. (2020b), we set the constants $c = 1.01$ and $\alpha_0 = 0.10$ in (3). Additionally, we set the constant $c_0 = 0.01$. These choices lead to satisfactory performance in our practice. Furthermore, a sensitivity analysis for these constants is conducted in Section B.2 of the Appendix.

Example 1. The comparison between the proposed DSRLasso estimator and the TSLasso estimator is conducted using the following four metrics: (i) ℓ_1 error $\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|_1$; (ii) ℓ_2 error $\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|_2$; (iii) number of false positive variables (FP) $\sum_{k \in \mathcal{B}^c} I(\widehat{\beta}_k \neq 0)$; and (iv) number of false negative variables (FN) $\sum_{k \in \mathcal{B}} I(\widehat{\beta}_k = 0)$. The results are summarized in Table 1, which presents the averages and standard errors of these metrics based on 200 simulations.

Error	N	Method	ℓ_1 Error	ℓ_2 Error	FP	FN
$\mathcal{N}(0, 2^2)$	300	TSLasso	1.141(0.050)	0.450(0.008)	10.77(0.871)	0.000(0.000)
		DSRLasso	1.212(0.015)	0.776(0.010)	0.225(0.035)	0.000(0.000)
		DSRLasso$^{\text{sub}}$	1.225(0.016)	0.783(0.011)	0.360(0.043)	0.000(0.000)
	450	TSLasso	0.926(0.040)	0.351(0.006)	11.46(0.828)	0.000(0.000)
		DSRLasso	0.976(0.012)	0.632(0.008)	0.200(0.032)	0.000(0.000)
		DSRLasso$^{\text{sub}}$	0.986(0.013)	0.637(0.008)	0.400(0.047)	0.000(0.000)
LogN	300	TSLasso	1.195(0.057)	0.464(0.011)	11.15(0.775)	0.000(0.000)
		DSRLasso	0.789(0.012)	0.507(0.008)	0.250(0.037)	0.000(0.000)
		DSRLasso$^{\text{sub}}$	0.799(0.013)	0.512(0.008)	0.450(0.047)	0.000(0.000)
	450	TSLasso	0.928(0.037)	0.372(0.008)	10.34(0.756)	0.000(0.000)
		DSRLasso	0.610(0.009)	0.394(0.006)	0.260(0.034)	0.000(0.000)
		DSRLasso$^{\text{sub}}$	0.615(0.010)	0.397(0.006)	0.480(0.045)	0.000(0.000)
MN	300	TSLasso	1.306(0.057)	0.530(0.012)	10.32(0.783)	0.000(0.000)
		DSRLasso	0.840(0.013)	0.539(0.009)	0.260(0.036)	0.000(0.000)
		DSRLasso$^{\text{sub}}$	0.845(0.014)	0.540(0.009)	0.420(0.047)	0.000(0.000)
	450	TSLasso	1.137(0.044)	0.438(0.009)	11.97(0.777)	0.000(0.000)
		DSRLasso	0.636(0.009)	0.412(0.006)	0.200(0.029)	0.000(0.000)
		DSRLasso$^{\text{sub}}$	0.641(0.009)	0.413(0.006)	0.360(0.043)	0.000(0.000)
$t(2)$	300	TSLasso	1.998(0.080)	0.854(0.029)	9.450(0.664)	0.025(0.013)
		DSRLasso	1.401(0.020)	0.886(0.013)	0.280(0.038)	0.000(0.000)
		DSRLasso$^{\text{sub}}$	1.415(0.020)	0.892(0.013)	0.485(0.050)	0.000(0.000)
	450	TSLasso	1.710(0.070)	0.692(0.025)	10.76(0.726)	0.030(0.016)
		DSRLasso	1.051(0.016)	0.666(0.010)	0.275(0.042)	0.000(0.000)
		DSRLasso$^{\text{sub}}$	1.061(0.016)	0.670(0.010)	0.475(0.050)	0.000(0.000)
Cauchy	300	TSLasso	5.331(0.181)	2.542(0.049)	4.995(0.551)	1.685(0.090)
		DSRLasso	2.397(0.062)	1.451(0.033)	0.545(0.058)	0.055(0.025)
		DSRLasso$^{\text{sub}}$	2.433(0.064)	1.462(0.033)	0.815(0.072)	0.045(0.023)
	450	TSLasso	5.185(0.130)	2.515(0.047)	4.910(0.506)	1.680(0.088)
		DSRLasso	1.907(0.066)	1.170(0.035)	0.525(0.054)	0.010(0.007)
		DSRLasso$^{\text{sub}}$	1.943(0.078)	1.186(0.046)	0.820(0.067)	0.010(0.007)

Table 1: Simulation Results for Example 1. Averages (and Standard Errors) of ℓ_1 Error, ℓ_2 Error, # of False Positive Variables (FP) and # of False Negative Variables (FN) Based on 200 Simulations.

We have the following observations. (1) DSRLasso yields considerably fewer false positives than TSLasso across all error distributions. (2) When random errors follow the normal distribution, DSRLasso produces comparable ℓ_1 errors to TSLasso, while under heavy-tailed distributions, such as mixture-normal, log-normal, and $t(2)$, DSRLasso yields significantly smaller ℓ_1 errors. (3) In the case of Cauchy-distributed errors, DSRLasso outperforms TSLasso across all four evaluation metrics. Notably, TSLasso fails to control false negatives: on average, it incorrectly shrinks more than half of the coefficients in the active set to zero. This leads to substantially larger ℓ_1 and ℓ_2 errors compared to DSRLasso. For example, when $N = 450$, the average ℓ_1 error, ℓ_2 error and FN for DSRLasso are 1.907, 1.170 and 0.010, respectively, while the corresponding values for TSLasso are 5.185, 2.515 and 1.680. (4) DSRLasso$^{\text{sub}}$ incurs only a negligible loss of statistical efficiency compared to DSRLasso. (5) Overall, DSRLasso demonstrates greater robustness to heavy-tailed random errors, which aligns with its weaker assumptions on error distributions and is consistent with earlier findings on the robustness of rank Lasso (Wang et al., 2020).

Example 2. We compare the performance of the proposed debiased DSRLasso and the debiased TSLasso in inferring individual coefficients, using the following metrics:

(i) Coverage probability for $\beta_{0,1}$ (CP_1): $\text{pr}(\beta_{0,1} \in \text{CI}_1)$;

(ii) Coverage probability for $\beta_{0,500}$ (CP_{500}): $\text{pr}(\beta_{0,500} \in \text{CI}_{500})$;

(iii) Average coverage probability for $\mathcal{B}$ ($\text{ACP}_{\mathcal{B}}$): $\sum_{k\in\mathcal{B}} \text{pr}(\beta_{0,k} \in \text{CI}_k)/q$;

(iv) Average coverage probability for $\mathcal{B}^c$ ($\text{ACP}_{\mathcal{B}^c}$): $\sum_{k\in\mathcal{B}^c} \text{pr}(0 \in \text{CI}_k)/(p-q)$;

(v) Average rejection probability/power for $\mathcal{B}$ ($\text{ARP}_{\mathcal{B}}$): $1 - \sum_{k\in\mathcal{B}} \text{pr}(0 \in \text{CI}_k)/q$;

(vi) Rejection and coverage probability for $\beta_{0,1}$ (RCP_1): $\text{pr}(0 \notin \text{CI}_1, \beta_{0,1} \in \text{CI}_1)$;

(vii) Average interval length for $\mathcal{B}$ ($\text{AIL}_{\mathcal{B}}$): $\sum_{k\in\mathcal{B}} \text{length}(\text{CI}_k)/q$;

(viii) Average interval length for $\mathcal{B}^c$ ($\text{AIL}_{\mathcal{B}^c}$): $\sum_{k\in\mathcal{B}^c} \text{length}(\text{CI}_k)/(p-q)$,

where CI_k denotes the two-sided 95% confidence interval for $\beta_{0,k}$.

The results reported in Table 2 are based on 200 simulations. Both debiased DSRLasso and debiased TSLasso show coverage probabilities close to the nominal 95% level for $\mathcal{B}^c$. For $\mathcal{B}$, however, debiased DSRLasso exhibits coverage probabilities closer to 95% in most cases compared to debiased TSLasso. Furthermore, debiased DSRLasso demonstrates greater robustness to heavy-tailed errors. For heavy-tailed error distributions such as mixture-normal, log-normal, $t(2)$, and Cauchy, the confidence intervals of debiased DSRLasso are significantly shorter than those of debiased TSLasso. In particular, debiased DSRLasso consistently maintains interval lengths below 1.5 across all settings, whereas debiased TSLasso's intervals exceed 10 under Cauchy errors. When errors follow $t(2)$ or Cauchy distributions, debiased DSRLasso achieves significantly higher power in detecting nonzero coefficients. For example, under Cauchy errors with $N = 450$, debiased DSRLasso attains values of 0.992 and 0.940 for metrics (v) and (vi), respectively, whereas debiased TSLasso yields substantially lower values of 0.438 and 0.360. This implies that under heavy-tailed errors, the excessively wide confidence intervals of debiased TSLasso can lead to the

Error	N	Method	CP_1	CP_{500}	$ACP_{\mathcal{B}}$	$ACP_{\mathcal{B}^c}$	$ARP_{\mathcal{B}}$	RCP_1	$AIL_{\mathcal{B}}$	$AIL_{\mathcal{B}^c}$
$\mathcal{N}(0,2^2)$	300	TSLasso	0.870	0.940	0.848	0.948	1.000	0.870	0.454	0.457
		DSRLasso	0.915	0.925	0.898	0.947	1.000	0.915	0.617	0.600
		$\text{DSRLasso}^{\text{sub}}$	0.910	0.930	0.893	0.947	1.000	0.910	0.618	0.601
	450	TSLasso	0.925	0.940	0.928	0.949	1.000	0.925	0.384	0.388
		DSRLasso	0.955	0.955	0.942	0.949	1.000	0.955	0.483	0.476
		$\text{DSRLasso}^{\text{sub}}$	0.960	0.950	0.943	0.949	1.000	0.960	0.483	0.477
LogN	300	TSLasso	0.885	0.945	0.872	0.949	1.000	0.885	0.467	0.470
		DSRLasso	0.950	0.950	0.925	0.952	1.000	0.950	0.486	0.473
		$\text{DSRLasso}^{\text{sub}}$	0.965	0.950	0.933	0.952	1.000	0.965	0.487	0.474
	450	TSLasso	0.915	0.935	0.903	0.949	1.000	0.915	0.406	0.410
		DSRLasso	0.960	0.935	0.953	0.953	1.000	0.960	0.363	0.359
		$\text{DSRLasso}^{\text{sub}}$	0.950	0.935	0.953	0.953	1.000	0.950	0.364	0.359
MN	300	TSLasso	0.830	0.955	0.858	0.948	1.000	0.830	0.549	0.552
		DSRLasso	0.900	0.960	0.902	0.949	1.000	0.900	0.435	0.424
		$\text{DSRLasso}^{\text{sub}}$	0.900	0.955	0.903	0.949	1.000	0.900	0.436	0.424
	450	TSLasso	0.905	0.950	0.875	0.949	1.000	0.905	0.464	0.470
		DSRLasso	0.945	0.950	0.937	0.948	1.000	0.945	0.329	0.326
		$\text{DSRLasso}^{\text{sub}}$	0.955	0.950	0.942	0.948	1.000	0.955	0.330	0.326
$t(2)$	300	TSLasso	0.875	0.925	0.860	0.949	0.985	0.870	0.936	0.942
		DSRLasso	0.940	0.940	0.918	0.951	1.000	0.940	0.754	0.752
		$\text{DSRLasso}^{\text{sub}}$	0.940	0.945	0.927	0.951	1.000	0.940	0.755	0.753
	450	TSLasso	0.905	0.960	0.903	0.948	0.985	0.885	0.827	0.836
		DSRLasso	0.950	0.960	0.940	0.949	1.000	0.950	0.555	0.560
		$\text{DSRLasso}^{\text{sub}}$	0.965	0.960	0.943	0.949	1.000	0.965	0.556	0.562
Cauchy	300	TSLasso	0.945	0.960	0.922	0.949	0.445	0.425	10.30	10.34
		DSRLasso	0.950	0.925	0.955	0.947	0.973	0.930	1.457	1.454
		$\text{DSRLasso}^{\text{sub}}$	0.940	0.935	0.952	0.948	0.973	0.920	1.461	1.458
	450	TSLasso	0.915	0.965	0.913	0.949	0.438	0.360	12.90	13.10
		DSRLasso	0.950	0.945	0.937	0.951	0.992	0.940	1.104	1.116
		$\text{DSRLasso}^{\text{sub}}$	0.945	0.935	0.933	0.952	0.987	0.930	1.137	1.150

Table 2: Simulation Results for Example 2. Empirical Coverage Probabilities, Power and Lengths of Confidence Intervals Based on 200 Simulations.

omission of important variables. Finally, debiased DSRLasso$^{\text{sub}}$ and debiased DSRLasso show comparable performance.

Example 3. We compare the performance of the proposed debiased DSRLasso and the debiased TSLasso in simultaneous inference, using the following three metrics:

(a) Coverage probability (CP): $\text{pr}[\cap_{k\in\mathcal{G}}\{\beta_{0,k} \in \text{SCI}_k\}]$;

(b) Rejection probability/power (RP): $\text{pr}[\cup_{k\in\mathcal{G}}\{0 \notin \text{SCI}_k\}]$;

(c) Average interval length (AIL): $\sum_{k\in\mathcal{G}} \text{length}(\text{SCI}_k)/|\mathcal{G}|$,

where $\{\text{SCI}_k\}_{k\in\mathcal{G}}$ denotes the 95% simultaneous confidence interval (SCI) for $\{\beta_{0,k}\}_{k\in\mathcal{G}}$.

We set $\mathcal{G} = \{1, \ldots, p\}$ and consider both non-studentized (NST) and studentized (ST) SCIs. The NST intervals have equal lengths across all $k \in \mathcal{G}$, whereas the ST intervals have varying lengths. Empirical values of the three metrics, estimated from 200 simulations, are summarized in Table 3. The overall pattern aligns with Table 2. Generally, SCIs based on the debiased DSRLasso exhibit coverage probabilities closer to the nominal 95% level than those based on the debiased TSLasso. Moreover, the debiased DSRLasso-based intervals are significantly shorter under heavy-tailed error distributions. When errors follow $t(2)$ or Cauchy distributions, the debiased DSRLasso-based method has significantly better power performance than the debiased TSLasso-based method in testing the simultaneous hypothesis $H_{0,\mathcal{G}} : \beta_{0,k} = 0$ for all $k \in \mathcal{G}$. Additionally, the performance of ST-based methods is comparable to NST-based methods, though the NST intervals tend to be slightly longer. Finally, DSRLasso$^{\text{sub}}$-based and DSRLasso-based SCIs show similar performance.

4.2 A Real Example

In this section, we apply the proposed methods to analyze the breast cancer dataset GSE93601 from the GEO database (`https://www.ncbi.nlm.nih.gov/geo`). This dataset was analyzed in Wang et al. (2015) and Zhao et al. (2022). It includes clinical information, such as age, body mass index (BMI), and alcohol consumption, from participants of the Nurses' Health Study I and II who were diagnosed with invasive breast cancer, along with gene expression profiling of 602 breast tumors and 508 tumor-adjacent normal tissues. Among the genes studied, TP53 (tumor protein p53) plays a critical role in inducing apoptosis under various stress conditions, including oncogene activation or DNA damage. Loss of p53 tumor-suppressor activity, whether through mutation/deletion of the TP53 gene or inhibition of the p53 protein, promotes cancer development. Our objective is to identify genes whose expression levels are associated with TP53. Hence, we use TP53 expression as the response variable and the expression levels of other genes as explanatory variables.

We restrict our analysis to breast tumor tissues. Individuals with missing age, BMI, or alcohol consumption data are excluded, resulting in a final sample size of $n = 596$. Genes with a missing rate exceeding 10% are removed, and the remaining missing gene expression values are imputed using the "impute.knn" from the R package `impute`. Among the retained genes, the MDM2 (mouse double minute 2 homolog) protein is a known negative regulator of p53. By binding to p53, it suppresses its transcriptional activity, promotes nuclear export, and enhances degradation. Overexpression of MDM2 in various tumors inhibits p53, thereby facilitating uncontrolled cell proliferation (Chène, 2003). Therefore, we include MDM2 in

Error	N	Method	TSLasso			DSRLasso			DSRLasso^{sub}		
			CP	RP	AIL	CP	RP	AIL	CP	RP	AIL
$\mathcal{N}(0,2^2)$	300	NST	0.855	1.000	0.912	0.940	1.000	1.207	0.940	1.000	1.209
		ST	0.850	1.000	0.903	0.920	1.000	1.189	0.940	1.000	1.191
	450	NST	0.880	1.000	0.773	0.935	1.000	0.951	0.935	1.000	0.952
		ST	0.900	1.000	0.768	0.940	1.000	0.944	0.940	1.000	0.944
LogN	300	NST	0.840	1.000	0.938	0.955	1.000	1.014	0.960	1.000	1.016
		ST	0.855	1.000	0.929	0.955	1.000	0.997	0.950	1.000	0.999
	450	NST	0.920	1.000	0.817	0.955	1.000	0.713	0.960	1.000	0.714
		ST	0.910	1.000	0.812	0.960	1.000	0.706	0.960	1.000	0.707
MN	300	NST	0.870	1.000	1.103	0.900	1.000	0.851	0.905	1.000	0.852
		ST	0.850	1.000	1.093	0.940	1.000	0.838	0.945	1.000	0.839
	450	NST	0.870	1.000	0.934	0.955	1.000	0.651	0.950	1.000	0.651
		ST	0.880	1.000	0.929	0.945	1.000	0.646	0.945	1.000	0.646
$t(2)$	300	NST	0.905	0.980	1.879	0.945	1.000	1.514	0.950	1.000	1.516
		ST	0.880	0.980	1.861	0.950	1.000	1.487	0.950	1.000	1.489
	450	NST	0.930	0.985	1.663	0.965	1.000	1.119	0.965	1.000	1.121
		ST	0.935	0.985	1.652	0.965	1.000	1.107	0.960	1.000	1.109
Cauchy	300	NST	0.890	0.325	20.64	0.915	0.915	3.769	0.900	0.910	3.777
		ST	0.910	0.320	20.39	0.905	0.765	3.583	0.905	0.775	3.590
	450	NST	0.905	0.320	26.01	0.930	0.985	2.868	0.935	0.980	2.900
		ST	0.910	0.330	25.81	0.960	0.955	2.810	0.960	0.945	2.841

Table 3: Simulation Results of Example 3. Empirical Coverage Probabilities, Power and Average Interval Lengths for Non-studentized (NST) and Studentized (ST) Methods Based on 200 Simulations.

our analysis and assume a functional relationship between MDM2 and TP53. For the remaining genes, we further filter the genes by including only those with expression values that fall within the top 800 in terms of absolute Spearman correlation with TP53.

Given that aging, excess body weight, and alcohol consumption are widely recognized as significant risk factors for many cancers (Renehan et al., 2008; Yoo et al., 2022; Montégut et al., 2024), we also include age, BMI, and alcohol intake as explanatory variables. We define two dummy variables for BMI and two for alcohol consumption. Specifically, for $i = 1, \ldots, 596$, let $Y_i = \text{TP53}_i$, $Z_i = \text{MDM2}_i$, and $\mathbf{x}_i = \{X_{i,k}\}_{k=1}^{805} \in \mathbb{R}^{805}$, where $X_{i,1} = \text{Age}_i$, $X_{i,2} = I(\text{BMI}_i > 25 \text{ kg/m}^2)$, $X_{i,3} = I(\text{BMI}_i > 30 \text{ kg/m}^2)$, $X_{i,4} = I(0 \text{ g/day} < \text{Alcohol}_i \leq 10 \text{ g/day})$, $X_{i,5} = I(\text{Alcohol}_i > 10 \text{ g/day})$, and $\{X_{i,k}\}_{k=6}^{805}$ represents the expression values of the 800 selected genes for the i-th sample. We consider the following partially linear model:

$$Y_i = g_0(Z_i) + \sum_{k=1}^{805} \beta_{0,k} X_{i,k} + \epsilon_i, \ i = 1, \ldots, 596.$$

The kurtosis of TP53 expression values is 2.36, suggesting a potentially heavy-tailed distribution. Let p_k^{TSLasso} and p_k^{DSRLasso} denote the p-values from the debiased two-step projection-based Lasso estimator and the proposed debiased DS rank Lasso estimator, respectively, for testing the significance of $\beta_{0,k}$. At the 1% significance level, DSRLasso indicates a significant positive effect of aging ($p_k^{\text{DSRLasso}} = 0.003$) consistent with previous findings that aging is a significant risk factor for cancer, whereas the Lasso-based method does not ($p_k^{\text{TSLasso}} = 0.059$). A summary of the p-values for the regression coefficients of the genes from both methods is provided in Table 4.

Overall, the genes selected by the two methods show considerable overlap at the 5% significance level. In particular, the p-values for the regression coefficients of the genes {C17orf85, POLR2B, COL10A1, KIAA0368, PRODH2} are all below 0.01 in both methods, and these genes have been previously implicated in tumorigenesis and prognosis (Zacharakis et al., 2018; Zhang et al., 2019; Hu et al., 2025; Wang et al., 2025; Gong et al., 2025). Moreover, the proposed debiased DSRLasso method identifies several genes not detected by the debiased TSLasso method. Specifically, the p-values for {PAPOLA, HBP1, SMG7, EP300, WDR75} are all below 0.05 for DSRLasso, whereas they exceed 0.05 for TSLasso. These genes have been linked to p53 or breast cancer in existing literature (Yang et al., 2020; Komini et al., 2021; Moudry et al., 2022; Zhou et al., 2023; Gronkowska and Robaszkiewicz, 2024). This demonstrates that the proposed debiased DSRLasso method is effective in identifying genes that the debiased TSLasso method fails to detect.

We also apply the proposed methods to a worker wage dataset in Section B.6 of the Appendix.

5. Discussion

In this study, we propose the DS rank Lasso, a novel regularized method for high-dimensional partially linear models. We also develop a debiased DS rank Lasso estimator to enable inference for individual coefficients as well as simultaneous inference. All proposed methods exhibit robustness against heavy-tailed random errors. To preserve the near-oracle rate

	$p_k^{\text{TSLasso}} \in [0, 0.01)$	$p_k^{\text{TSLasso}} \in [0.01, 0.05)$	$p_k^{\text{TSLasso}} \in [0.05, 1]$
$p_k^{\text{DSRLasso}} \in [0, 0.01)$	C17orf85(NCBP3), POLR2B, COL10A1, KIAA0368 PRODH2(NPHS1)	C13orf23, N-PAC PIBF1(C13orf24), STAU2, RBM39	PAPOLA, Q6ZU24(FER1L5)
$p_k^{\text{DSRLasso}} \in [0.01, 0.05)$	BRWD1, SFPQ, WDR57, TRAPPC1, STAB2, ZC3H11A, NCKAP1	SET, NAT10, UBA3, CNOT6, GOPC, ATP2A2, TMEM131, NUP188, SETD2, WAC	WDR42A, ZZEF1, HBP1, SPTLC2, SMG7, WDR75, NUPL1, EP300, ZNF23(ZNF19), NUP214, RPN1, ZC3H7A, KIAA0494, UBQLN1, SLC33A1
$p_k^{\text{DSRLasso}} \in [0.05, 1]$	GFM1, DNAH2, DDX5	WNK1, POLR2A, BIRC6, AKAP11, MYOM1, TUBA1B, LOC645159(MLL3), KPNB1, KIAA0196, ELF1, SEC31A, TRAM1, PSMD1, NRAP, CEP350, ANXA7, ZNF124, LRRC16A(LRRC16), GNPTAB, NRD1	Other Genes

Table 4: A Summary of the p-values for the Regression Coefficients of the 800 Genes.

of rank Lasso under mild conditions on nonparametric estimators, we incorporate data-splitting into the regularized regression method. The two-split procedure is also used for inference, where the proposed debiased DS rank Lasso estimator is constructed as an average of two debiased estimators. Furthermore, we demonstrate that data splitting asymptotically incurs no loss of statistical efficiency in the inference procedure.

Acknowledgments

The authors would like to thank the editor and reviewers for their helpful comments. Yang's research is supported by the BRAIN of Renmin University of China and NSFC 12301389. Zhao's research is supported by the Natural Science Foundation of Fujian Province, China (No. 2026J008060) and the Startup Fund of Fuzhou University (No. 511839).

Appendix A. Technical Details

In this appendix, we provide proofs for Lemma 1 and Lemma 4, and Theorem 2, Theorem 5, and Theorem 7 in the main text, assuming without loss of generality that each subsample has an equal size of n.

We define the event

$$\mathscr{N} = \left\{ \max_{k\in\{0,1,\ldots,p\}} \sup_{z\in\mathcal{Z}} |\widehat{m}_k^{(1)}(z) - m_k(z)| \le r_n \right\}.$$

Consequently, we have $\mathrm{pr}(\mathscr{N}^c) \le \rho_n$ by Condition (C1). For notational simplicity, we denote $a_n = 2(\|\boldsymbol{\beta}_0\|_1 \vee 1) r_n$, $\widehat{\Delta}_k^{(1)}(Z_i^{(2)}) = \widehat{m}_k^{(1)}(Z_i^{(2)}) - m_k(Z_i^{(2)})$, $\widehat{\Delta}_g^{(1)}(Z_i^{(2)}) = \widehat{g}(Z_i^{(2)}, \boldsymbol{\beta}_0) - g_0(Z_i^{(2)})$, and use $E_*(\cdot)$ and $\mathrm{pr}_*(\cdot)$ to represent the conditional expectation and conditional probability given $\mathcal{D}_1$.

A.1 Proof of Lemma 1

Proof We will prove the results only for $l = 1$, as the results for $l = 2$ can be obtained directly by swapping the roles of the two subsamples.

For the first part of Lemma 1, we will derive a high probability upper bound for $|\|\mathbf{S}^{(2)}(\mathbf{0})\|_\infty - \|\mathbf{S}_0^{(2)}\|_\infty| \le \|\mathbf{S}^{(2)}(\mathbf{0}) - \mathbf{S}_0^{(2)}\|_\infty$, where

$$\begin{aligned}
\mathbf{S}^{(2)}(\boldsymbol{\gamma}) &= -2\{n(n-1)\}^{-1} \sum_{i=1}^n \widehat{\mathbf{x}}_i^{(2)} \sum_{j=1, j\neq i}^n \mathrm{sign}\big\{(e_i^{(2)} - e_j^{(2)}) - (\widehat{\mathbf{x}}_i^{(2)} - \widehat{\mathbf{x}}_j^{(2)})^{\mathrm{T}}\boldsymbol{\gamma}\big\} \\
&= -\{n(n-1)\}^{-1} \sum_{i\neq j} (\widehat{\mathbf{x}}_i^{(2)} - \widehat{\mathbf{x}}_j^{(2)}) \mathrm{sign}\big\{(e_i^{(2)} - e_j^{(2)}) - (\widehat{\mathbf{x}}_i^{(2)} - \widehat{\mathbf{x}}_j^{(2)})^{\mathrm{T}}\boldsymbol{\gamma}\big\},
\end{aligned}$$

and

$$\begin{aligned}
\mathbf{S}_0^{(2)} &= -2\{n(n-1)\}^{-1} \sum_{i=1}^n \widehat{\mathbf{x}}_i^{(2)} \sum_{j=1, j\neq i}^n \mathrm{sign}(\epsilon_i^{(2)} - \epsilon_j^{(2)}) \\
&= -\{n(n-1)\}^{-1} \sum_{i\neq j} (\widehat{\mathbf{x}}_i^{(2)} - \widehat{\mathbf{x}}_j^{(2)}) \mathrm{sign}(\epsilon_i^{(2)} - \epsilon_j^{(2)}).
\end{aligned}$$

This can be achieved by upper bounding $\|\mathbf{S}_1\|_\infty = \|\{n(n-1)\}^{-1} \sum_{i\neq j} \mathbf{S}_{1,i,j}\|_\infty$ and $\|\mathbf{S}_2\|_\infty = \|\{n(n-1)\}^{-1} \sum_{i\neq j} \mathbf{S}_{2,i,j}\|_\infty$, where $\mathbf{S}_1 = (S_{1,1}, \ldots, S_{1,p})^{\mathrm{T}}$, $\mathbf{S}_{1,i,j} = (S_{1,i,j,1}, \ldots, S_{1,i,j,p})^{\mathrm{T}}$, $S_{1,i,j,k} = (\widetilde{X}_{i,k}^{(2)} - \widetilde{X}_{j,k}^{(2)})\{\mathrm{sign}(e_i^{(2)} - e_j^{(2)}) - \mathrm{sign}(\epsilon_i^{(2)} - \epsilon_j^{(2)})\}$, $\mathbf{S}_2 = (S_{2,1}, \ldots, S_{2,p})^{\mathrm{T}}$, $\mathbf{S}_{2,i,j} = (S_{2,i,j,1}, \ldots, S_{2,i,j,p})^{\mathrm{T}}$, and $S_{2,i,j,k} = \{\widehat{\Delta}_k^{(1)}(Z_i^{(2)}) - \widehat{\Delta}_k^{(1)}(Z_j^{(2)})\}\{\mathrm{sign}(e_i^{(2)} - e_j^{(2)}) - \mathrm{sign}(\epsilon_i^{(2)} - \epsilon_j^{(2)})\}$.

Note that for any $t > 0$,

$$\mathrm{pr}_*(|S_{1,k}| \ge t) = \mathrm{pr}_*(S_{1,k} \ge t) + \mathrm{pr}_*(S_{1,k} \le -t).$$

We denote $M_n = \lfloor n/2 \rfloor$, the largest integer smaller than or equal to $n/2$, and π the permutations of $\{1, \ldots, n\}$. For $\mathrm{pr}_*(S_{1,k} \geq t)$, we have for any $\lambda > 0$, that

$$\begin{aligned}
\mathrm{pr}_*\big(S_{1,k} \geq t\big) &\leq E_*\Big(\exp\Big[\lambda\{n(n-1)\}^{-1} \sum_{i \neq j} S_{1,i,j,k}\Big]\Big) \Big/ \exp(\lambda t) \\
&= E_*\Big\{ \exp\Big(\lambda \frac{1}{n!} \sum_{\pi} M_n^{-1} \sum_{i=1}^{M_n} S_{1,\pi(i),\pi(M_n+i),k}\Big)\Big\} \Big/ \exp(\lambda t) \\
&\leq \frac{1}{n!} \sum_{\pi} E_*\Big\{ \exp\Big(\lambda M_n^{-1} \sum_{i=1}^{M_n} S_{1,\pi(i),\pi(M_n+i),k}\Big)\Big\} \Big/ \exp(\lambda t).
\end{aligned}$$

The two inequalities follow from Markov's and Jensen's inequalities, respectively, and the equality is due to the property of U-statistic. It's thus sufficient to deal with the summand for any given π. We will frequently use the above arguments in our proofs. Because $\widetilde{X}^{(2)}_{\pi(i),k} - \widetilde{X}^{(2)}_{\pi(M_n+i),k}$ is zero-mean sub-Gaussian with parameter $\sqrt{2}\sigma$ under Condition (C2), $E_*|\widetilde{X}^{(2)}_{\pi(i),k} - \widetilde{X}^{(2)}_{\pi(M_n+i),k}|^t \leq (4\sigma^2)^{t/2} t\Gamma(t/2)$ by using Proposition 2.5.2 (ii) in Vershynin (2018), where $\Gamma(\cdot)$ represents the Gamma function. Therefore, on $\mathscr{N}$, $E_*|S_{1,\pi(i),\pi(M_n+i),k}|^2 \leq 32\sigma^2(2\kappa_u a_n)$, and for $t \geq 3$,

$$\begin{aligned}
E_*|S_{1,\pi(i),\pi(M_n+i),k}|^t &\leq (4\sigma^2)^{t/2} t\Gamma(t/2) 2^t (4\kappa_u a_n) \\
&\leq \{t\Gamma(t/2)/2\}(64\sigma^2\kappa_u a_n)(4\sigma)^{t-2},
\end{aligned}$$

so that $S_{1,\pi(i),\pi(M_n+i),k}$ satisfies Bernstein's condition (see, for example, Wainwright, 2019, Chapter 2.1) with parameters $(64\sigma^2\kappa_u a_n, 4\sigma)$. Therefore, following similar arguments to Wainwright (2019, Proposition 2.10),

$$E_*\Big\{ \exp\Big(\lambda M_n^{-1} \sum_{i=1}^{M_n} S_{1,\pi(i),\pi(M_n+i),k}\Big)\Big\} \leq \exp[32\lambda^2\sigma^2\kappa_u a_n / \{M_n(1 - 4\sigma|\lambda|M_n^{-1})\}],$$

for all $|\lambda|M_n^{-1} < 1/(4\sigma)$. By letting $\lambda = M_n t/(4\sigma t + 64\sigma^2\kappa_u a_n)$, we have

$$\mathrm{pr}_*(S_{1,k} \geq t) \leq \exp[-M_n t^2/\{2(4\sigma t + 64\sigma^2\kappa_u a_n)\}],$$

and hence

$$\begin{aligned}
\mathrm{pr}_*(|S_{1,k}| \geq t) &\leq 2\exp[-M_n t^2/\{2(4\sigma t + 64\sigma^2\kappa_u a_n)\}] \\
&\leq 2\exp[-M_n t^2/\{4(4\sigma t \vee 64\sigma^2\kappa_u a_n)\}],
\end{aligned}$$

by processing $\mathrm{pr}_*(S_{1,k} \leq -t)$ in similar steps. By the union bound, we have that

$$\mathrm{pr}_*(\|\mathbf{S}_1\|_\infty \geq t) \leq 2\exp[-M_n t^2/\{4(4\sigma t \vee 64\sigma^2\kappa_u a_n)\} + \log p].$$

Denote $\mathscr{S}_1 = [\|\mathbf{S}_1\|_\infty \leq 16\sqrt{2}C\sigma\sqrt{\log p/n}\{(\kappa_u a_n)^{1/2} \vee (2\log p/n)^{1/4}\}]$, and then $\mathrm{pr}_*(\mathscr{S}_1^c) \leq 2\exp(-[\{C^2 \wedge C(2\log p/n)^{-1/4}\} - 1]\log p)$ on $\mathscr{N}$.

As for $|S_{2,k}|$, we can follow a similar paradigm to $|S_{1,k}|$. Specifically,

$$\begin{aligned}|E_*(S_{2,\pi(i),\pi(M_n+i),k})| &= \Big|E_*\Big(\{\widehat{\Delta}_k^{(1)}(Z_i^{(2)}) - \widehat{\Delta}_k^{(1)}(Z_j^{(2)})\} \\ &\qquad \times[2F^*\{\widehat{\Delta}_g^{(1)}(Z_i^{(2)}) - \widehat{\Delta}_g^{(1)}(Z_j^{(2)})\} - 1]\Big)\Big| \le 8\kappa_u r_n a_n,\end{aligned}$$

under Condition (C4) and on $\mathscr{N}$. Moreover, we note that

$$\begin{aligned}& E_*|S_{2,\pi(i),\pi(M_n+i),k} - E_*(S_{2,\pi(i),\pi(M_n+i),k})|^2 \\ \le\ & E_*|S_{2,\pi(i),\pi(M_n+i),k}|^2 \le 32 r_n^2 \kappa_u a_n,\end{aligned}$$

and for $t \ge 3$,

$$\begin{aligned}& E_*|S_{2,\pi(i),\pi(M_n+i),k} - E_*(S_{2,\pi(i),\pi(M_n+i),k})|^t \\ \le\ & 2^t E_*|S_{2,\pi(i),\pi(M_n+i),k}|^t \le (32 r_n^2 \kappa_u a_n)(4r_n)^{t-2} 2^t \\ =\ & 2^{t-2}(128 r_n^2 \kappa_u a_n)(4r_n)^{t-2} \le t!/2 (128 r_n^2 \kappa_u a_n)(4r_n)^{t-2},\end{aligned}$$

on $\mathscr{N}$, which implies $S_{2,\pi(i),\pi(M_n+i),k}$ satisfies Bernstein's condition (see, for example, Wainwright, 2019, Chapter 2.1) with parameters $(128 r_n^2 \kappa_u a_n, 4r_n)$. Following a similar argument to $\mathscr{S}_1$, $\mathrm{pr}_*(\mathscr{S}_2^c) \le 2\exp(-[\{C^2 \wedge C(\log p/n)^{-1/4}\} - 1]\log p)$ when $\mathcal{D}_1 \in \mathscr{N}$, where $\mathscr{S}_2 = [\|\mathbf{S}_2\|_\infty \le 8\kappa_u r_n a_n + 32 C r_n \sqrt{\log p/n}\{(\kappa_u a_n)^{1/2} \vee (\log p/n)^{1/4}\}]$.

To conclude,

$$\|\mathbf{S}_1\|_\infty + \|\mathbf{S}_2\|_\infty \le 8\kappa_u r_n a_n + 32 C(\sigma + r_n)\sqrt{\log p/n}\{(\kappa_u a_n)^{1/2} \vee (2\log p/n)^{1/4}\},$$

on $\mathscr{S}_1 \cap \mathscr{S}_2$, and $\mathrm{pr}_*\{(\mathscr{S}_1 \cap \mathscr{S}_2)^c\} \le 4\exp(-[\{C^2 \wedge C(2\log p/n)^{-1/4}\} - 1]\log p)$ on $\mathscr{N}$. This implies for any constant $c_0 > 0$, by letting

$$C = \Big[\{c_0 - 8\kappa_u r_n a_n (\log p/n)^{-1/2}\}/\{32(\sigma + r_n)\}\Big]\{(\kappa_u a_n)^{1/2} \vee (2\log p/n)^{1/4}\}^{-1},$$

we have $\|\mathbf{S}_1\|_\infty + \|\mathbf{S}_2\|_\infty \le c_0\sqrt{\log p/n}$. Therefore, $\mathrm{pr}_*\{(\mathscr{S}_1 \cap \mathscr{S}_2)^c\} \le 4p^{-(C_n^2 - 1)}$ on $\mathscr{N}$, where

$$C_n = \Big(\Big[\{c_0 - 8\kappa_u r_n a_n (\log p/n)^{-1/2}\}/\{32(\sigma + r_n)\}\Big] \wedge 1\Big)\{(\kappa_u a_n)^{1/2} \vee (2\log p/n)^{1/4}\}^{-1}.$$

Denote $\widehat{\boldsymbol{\gamma}}^{(1)}(\widehat{\lambda}^{(1)})$ by $\widehat{\boldsymbol{\gamma}}^{(1)}$ for simplicity. By the definition of $\widehat{\boldsymbol{\beta}}^{(1)}(\widehat{\lambda}^{(1)})$, we have,

$$Q^{(2)}(\widehat{\boldsymbol{\gamma}}^{(1)}) + \widehat{\lambda}^{(1)}\|\widehat{\boldsymbol{\beta}}^{(1)}\|_1 \le Q^{(2)}(\mathbf{0}) + \widehat{\lambda}^{(1)}\|\boldsymbol{\beta}_0\|_1.$$

This implies

$$Q^{(2)}(\widehat{\boldsymbol{\gamma}}^{(1)}) - Q^{(2)}(\mathbf{0}) \le \widehat{\lambda}^{(1)}\big(\|\boldsymbol{\beta}_0\|_1 - \|\widehat{\boldsymbol{\beta}}^{(1)}\|_1\big) \le \widehat{\lambda}^{(1)}\big(\|\widehat{\boldsymbol{\gamma}}^{(1)}_{\mathcal{B}}\|_1 - \|\widehat{\boldsymbol{\gamma}}^{(1)}_{\mathcal{B}^c}\|_1\big).$$

By simple calculation, the subgradient of $Q^{(2)}(\boldsymbol{\gamma})$ is

$$\begin{aligned}& -\{n(n-1)\}^{-1}\sum_{i\neq j}(\widehat{\mathbf{x}}_i^{(2)} - \widehat{\mathbf{x}}_j^{(2)})\mathrm{sign}\{(e_i^{(2)} - e_j^{(2)}) - (\widehat{\mathbf{x}}_i^{(2)} - \widehat{\mathbf{x}}_j^{(2)})^{\mathrm{T}}\boldsymbol{\gamma}\} \\ & -\{n(n-1)\}^{-1}\sum_{i\neq j}(\widehat{\mathbf{x}}_i^{(2)} - \widehat{\mathbf{x}}_j^{(2)})v_{i,j}^{(2)} I\{e_i^{(2)} - e_j^{(2)} = (\widehat{\mathbf{x}}_i^{(2)} - \widehat{\mathbf{x}}_j^{(2)})^{\mathrm{T}}\boldsymbol{\gamma}\},\end{aligned}$$

where $v_{i,j}^{(2)} \in [-1,1]$. We note that with probability one, the subgradient of $Q^{(2)}(\boldsymbol{\gamma})$ at $\boldsymbol{\gamma}=\mathbf{0}$ is exactly $\mathbf{S}^{(2)}(\mathbf{0})$ because the distribution of $\epsilon_i^{(2)}$ has no point masses. By the convexity of $Q(\cdot)$, the definition of subdifferential and the above discussion,

$$\begin{aligned} Q^{(2)}(\widehat{\boldsymbol{\gamma}}^{(1)}) - Q^{(2)}(\mathbf{0}) &\geq -\|\widehat{\boldsymbol{\gamma}}^{(1)}\|_1 \|\mathbf{S}^{(2)}(\mathbf{0})\|_\infty \\ &\geq -\|\widehat{\boldsymbol{\gamma}}^{(1)}\|_1 \big(\|\mathbf{S}_0^{(2)}\|_\infty + c_0\sqrt{\log p/n}\big) \\ &\geq -\big(\|\widehat{\boldsymbol{\gamma}}_{\mathcal{B}}^{(1)}\|_1 + \|\widehat{\boldsymbol{\gamma}}_{\mathcal{B}^c}^{(1)}\|_1\big)(\widehat{\lambda}^{(1)}/c), \end{aligned}$$

where the second inequality holds on $\mathscr{S}_1 \cap \mathscr{S}_2$, and the third inequality holds with probability at least $1-\alpha_0$ by the definition of $\widehat{\lambda}^{(1)}$. This implies $\mathrm{pr}_*\Big(\|\widehat{\boldsymbol{\gamma}}_{\mathcal{B}^c}^{(1)}\|_1 \leq \bar{c}\|\widehat{\boldsymbol{\gamma}}_{\mathcal{B}}^{(1)}\|_1\Big) = \mathrm{pr}_*\Big\{\widehat{\boldsymbol{\gamma}}^{(1)}(\widehat{\lambda}^{(1)}) \in \Gamma\Big\} \geq 1-\alpha_0-4p^{-(C_n^2-1)}$. Here, $C_n \to \infty$ as $n\to\infty$ since $a_n = o(1)$, $a_n r_n = o\big(\sqrt{\log p/n}\big)$ and $\log p/n = o(1)$ by Condition (C1).

For the second part of Lemma 1, it suffices to upper bound $\|\mathbf{S}_0^{(2)}\|_\infty = \|\{n(n-1)\}^{-1}\sum_{i\neq j}\mathbf{S}_{0,i,j}\|_\infty$, conditioning on $\{\widehat{\mathbf{x}}_i^{(2)}\}_{i=1}^n$, where $\mathbf{S}_0^{(2)} = (S_{0,1},\ldots,S_{0,p})^{\mathrm{T}}$, $\mathbf{S}_{0,i,j} = (S_{0,i,j,1},\ldots,S_{0,i,j,p})^{\mathrm{T}}$ and $S_{0,i,j,k} = (\widehat{X}_{i,k}^{(2)} - \widehat{X}_{j,k}^{(2)})\mathrm{sign}(\epsilon_i^{(2)} - \epsilon_j^{(2)})$.

We note that $\mathcal{D}_1$ and $\{\mathbf{x}_i^{(2)}, Z_i^{(2)}\}_{i=1}^n$ are independent of $\{\epsilon_i^{(2)}\}_{i=1}^n$, then conditioning on $\mathcal{D}_1$ and $\{\mathbf{x}_i^{(2)}, Z_i^{(2)}\}_{i=1}^n$, $M_n^{-1}\sum_{i=1}^{M_n} S_{0,\pi(i),\pi(M_n+i),k}$ for any given π is zero-mean sub-Gaussian, with parameter at most

$$M_n^{-1}\Big\{\sum_{i=1}^{M_n}(\widehat{X}_{\pi(i),k}^{(2)} - \widehat{X}_{\pi(M_n+i),k}^{(2)})^2\Big\}^{1/2} \leq \Big[(16/n)\Big\{n^{-1}\sum_{i=1}^n(\widetilde{X}_{i,k}^{(2)})^2 + r_n^2\Big\}\Big]^{1/2},$$

on $\mathscr{N}$. Denote the event $\mathscr{M}_1 = \{\max_k n^{-1}\sum_{i=1}^n(\widetilde{X}_{i,k}^{(2)})^2 \leq 12\sigma^2\}$, and then the above equation is upper bounded by $\{(16/n)(12\sigma^2+r_n^2)\}^{1/2}$ on $\mathscr{M}_1\cap\mathscr{N}$. Using Honorio and Jaakkola (2014, Appendix B), since $\widetilde{X}_{i,k}^{(2)}$ is sub-Gaussian with parameter σ, its square is sub-exponential with parameter $\big(4\sqrt{2}\sigma^2, 4\sigma^2\big)$. Besides, $E\{(\widetilde{X}_{i,k}^{(2)})^2\} \leq 4\sigma^2$ by Vershynin (2018, Proposition 2.5.2), then

$$\begin{aligned} &\mathrm{pr}\Big\{\max_k n^{-1}\sum_{i=1}^n(\widetilde{X}_{i,k}^{(2)})^2 \geq t\Big\} \leq \sum_{k=1}^p \mathrm{pr}\Big\{n^{-1}\sum_{i=1}^n(\widetilde{X}_{i,k}^{(2)})^2 \geq t\Big\} \\ \leq\ & \sum_{k=1}^p \inf_{\lambda\in[0,(4\sigma^2/n)^{-1})} E\Big(\exp\Big[(\lambda/n)\sum_{i=1}^n[(\widetilde{X}_{i,k}^{(2)})^2 - E\{(\widetilde{X}_{i,k}^{(2)})^2\}] + E\{(\widetilde{X}_{i,k}^{(2)})^2\}\Big]\Big)\Big/\exp(\lambda t) \\ \leq\ & \sum_{k=1}^p \inf_{\lambda\in[0,(4\sigma^2/n)^{-1})} \exp\{16\sigma^2\lambda^2/n - \lambda(t-4\sigma^2)\} \\ \leq\ & \exp\Big[-n(t-4\sigma^2)^2/\max\{64\sigma^4, 8\sigma^2(t-4\sigma^2)\} + \log p\Big]. \end{aligned}$$

Consequently, $\mathrm{pr}(\mathscr{M}_1^c) \leq p\exp(-n)$. Then, by properties of sub-Gaussian random variables and similar arguments to $\|\mathbf{S}_1\|_\infty$ and $\|\mathbf{S}_2\|_\infty$, it follows that

$$\mathrm{pr}[\|\mathbf{S}_0^{(2)}\|_\infty \geq t \mid \mathcal{D}_1, \{\mathbf{x}_i^{(2)}, Z_i^{(2)}\}_{i=1}^n] \leq 2\exp\{-(n/32)t^2/(12\sigma^2+r_n^2) + \log p\},$$

on $\mathscr{N} \cap \mathscr{M}_1$, where $\boldsymbol{\epsilon}^{(2)} = (\epsilon_1^{(2)}, \ldots, \epsilon_n^{(2)})^{\mathrm{T}}$. Consequently, for any $C > 1$,

$$\begin{aligned} & \mathrm{pr}[\|\mathbf{S}_0^{(2)}\|_\infty \geq 4\sqrt{2}C(2\sqrt{3}\sigma + r_n)\sqrt{\log p/n} \mid \mathcal{D}_1, \{\mathbf{x}_i^{(2)}, Z_i^{(2)}\}_{i=1}^n] \\ \leq \; & 2\exp\{-(C^2-1)\log p\}, \end{aligned}$$

on $\mathscr{N} \cap \mathscr{M}_1$, which implies, for $\alpha_0 > 2p^{-(C^2-1)}$, $\mathrm{pr}[\|\mathbf{S}_0^{(2)}\|_\infty \geq 4\sqrt{2}C(2\sqrt{3}\sigma + r_n)\sqrt{\log p/n} \mid \mathcal{D}_1, \{\mathbf{x}_i^{(2)}, Z_i^{(2)}\}_{i=1}^n] < \alpha_0$. By the definition of $\widehat{\lambda}^{(1)}$, $\widehat{\lambda}^{(1)}$ is the infimum of $\lambda^{(1)}$ such that

$$\begin{aligned} & \mathrm{pr}\Big[c(\|\mathbf{S}_0^{(2)}\|_\infty + c_0\sqrt{\log p/n}) > \lambda^{(1)} \;\Big|\; \mathcal{D}_1, \{\mathbf{x}_i^{(2)}, Z_i^{(2)}\}_{i=1}^n\Big] \\ = \; & \mathrm{pr}\Big[c(\|\mathbf{S}_0^{(2)}\|_\infty + c_0\sqrt{\log p/n}) > \lambda^{(1)} \;\Big|\; \{\widehat{\mathbf{x}}_i^{(2)}\}_{i=1}^n\Big] < \alpha_0. \end{aligned}$$

We can then conclude that $\widehat{\lambda}^{(1)} \leq c\{c_0 + 4\sqrt{2}C(2\sqrt{3}\sigma + r_n)\}\sqrt{\log p/n}$ on $\mathscr{N} \cap \mathscr{M}_1$. Thus, on $\mathscr{N}$, we have $\widehat{\lambda}^{(1)} \leq c\{c_0 + 4\sqrt{2}C(2\sqrt{3}\sigma + r_n)\}\sqrt{\log p/n}$ with probability at least $1 - p\exp(-n)$, which tends to one as $n \to \infty$. Then the proof is completed. ■

A.2 Proof of Theorem 2

Proof We focus on proving the properties of $\widehat{\boldsymbol{\beta}}^{(1)}$, because the properties of $\widehat{\boldsymbol{\beta}}^{(2)}$ follow directly by swapping the roles of the two subsamples.

We write $L^{(2)}(\boldsymbol{\gamma}) = Q^{(2)}(\boldsymbol{\gamma}) + \widehat{\lambda}^{(1)}\|\boldsymbol{\beta}_0 + \boldsymbol{\gamma}\|_1$, and then by definition,

$$\widehat{\boldsymbol{\gamma}}^{(1)}(\widehat{\lambda}^{(1)}) := \widehat{\boldsymbol{\beta}}^{(1)}(\widehat{\lambda}^{(1)}) - \boldsymbol{\beta}_0 = \operatorname*{arg\,min}_{\boldsymbol{\gamma} \in \mathbb{R}^p} L^{(2)}(\boldsymbol{\gamma}).$$

Recall that $\Gamma = \{\boldsymbol{\gamma} \in \mathbb{R}^p : \|\boldsymbol{\gamma}_{\mathcal{B}^c}\|_1 \leq \bar{c}\|\boldsymbol{\gamma}_{\mathcal{B}}\|_1\}$ and $\widehat{\lambda}^{(1)} = c\{G_{(1)}^{-1}(1-\alpha_0) + c_0\sqrt{\log p/n}\}$. It follows from Lemma 1 that $\mathrm{pr}_*\{\widehat{\boldsymbol{\gamma}}^{(1)}(\widehat{\lambda}^{(1)}) \in \Gamma\} \geq 1 - \alpha_0 - \alpha_n$ and $\widehat{\lambda}^{(1)} \leq c_1\sqrt{\log p/n}$ on $\mathscr{M}_1 \cap \mathscr{N}$, where $c_1 = c\{c_0 + 4\sqrt{2}C(2\sqrt{3}\sigma + r_n)\}$ and $\alpha_n = 4p^{-(C_n^2-1)} \to 0$ as $n \to \infty$. Let $h_n = \sqrt{q\log p/n}$, and denote

$$\Gamma^* = \{\boldsymbol{\gamma} \in \mathbb{R}^p : \boldsymbol{\gamma} \in \Gamma, \|\boldsymbol{\gamma}\|_2 = \delta h_n\},$$

where δ is a constant to be set large enough later. We note that

$$\begin{aligned} \mathrm{pr}\Big\{\|\widehat{\boldsymbol{\gamma}}^{(1)}(\widehat{\lambda}^{(1)})\|_2 \leq \delta h_n \;\Big|\; \mathscr{N}\Big\} \; \geq \; & 1 - \mathrm{pr}\Big\{\|\widehat{\boldsymbol{\gamma}}^{(1)}(\widehat{\lambda}^{(1)})\|_2 \geq \delta h_n, \widehat{\boldsymbol{\gamma}}^{(1)}(\widehat{\lambda}^{(1)}) \in \Gamma \;\Big|\; \mathscr{N}\Big\} \\ & - \mathrm{pr}\Big\{\widehat{\boldsymbol{\gamma}}^{(1)}(\widehat{\lambda}^{(1)}) \notin \Gamma \;\Big|\; \mathscr{N}\Big\} \\ \geq \; & \mathrm{pr}\Big\{\inf_{\boldsymbol{\gamma} \in \Gamma, \|\boldsymbol{\gamma}\|_2 \geq \delta h_n} L^{(2)}(\boldsymbol{\gamma}) > L^{(2)}(\mathbf{0}) \;\Big|\; \mathscr{N}\Big\} - \alpha_0 - \alpha_n \\ = \; & \mathrm{pr}\Big\{\inf_{\boldsymbol{\gamma} \in \Gamma^*} L^{(2)}(\boldsymbol{\gamma}) > L^{(2)}(\mathbf{0}) \;\Big|\; \mathscr{N}\Big\} - \alpha_0 - \alpha_n, \end{aligned}$$

where the third inequality is due to the convexity of $L^{(2)}(\cdot)$. To see this, for an arbitrary point $\mathbf{s}$ outside the ℓ_2 ball centered at $\mathbf{0}$ with radius δh_n, we can write it as $\mathbf{s} = l\mathbf{u}$, where

$\mathbf{u}$ is a unit vector and $l > \delta h_n$ is a positive constant. The convexity of $L^{(2)}(\cdot)$ implies that $(1 - l^{-1}\delta h_n)L^{(2)}(\mathbf{0}) + l^{-1}\delta h_n L^{(2)}(\mathbf{s}) \geq L^{(2)}(\delta h_n \mathbf{u})$. Hence,

$$\begin{aligned} l^{-1}\delta h_n\{L^{(2)}(\mathbf{s}) - L^{(2)}(\mathbf{0})\} &\geq L^{(2)}(\delta h_n \mathbf{u}) - L^{(2)}(\mathbf{0}) \\ &\geq \inf_{\|\boldsymbol{\gamma}\|_2 = \delta h_n} \{L^{(2)}(\boldsymbol{\gamma}) - L^{(2)}(\mathbf{0})\}. \end{aligned}$$

Besides, on $\mathscr{N} \cap \mathscr{M}_1$,

$$\begin{aligned} & \widehat{\lambda}^{(1)} |\|\boldsymbol{\beta}_0 + \boldsymbol{\gamma}\|_1 - \|\boldsymbol{\beta}_0\|_1| \\ \leq\ & \widehat{\lambda}^{(1)} \|\boldsymbol{\gamma}\|_1 = \widehat{\lambda}^{(1)}(\|\boldsymbol{\gamma}_{\mathcal{B}}\|_1 + \|\boldsymbol{\gamma}_{\mathcal{B}^c}\|_1) \\ \leq\ & \widehat{\lambda}^{(1)}(1 + \bar{c})\|\boldsymbol{\gamma}_{\mathcal{B}}\|_1 \leq \widehat{\lambda}^{(1)}(1 + \bar{c})\sqrt{q}\|\boldsymbol{\gamma}_{\mathcal{B}}\|_2 \\ \leq\ & \widehat{\lambda}^{(1)}(1 + \bar{c})\sqrt{q}\delta h_n \leq c_1(1 + \bar{c})\delta h_n^2, \end{aligned}$$

for any $\boldsymbol{\gamma} \in \Gamma^*$. We denote $Q^{(2)}(\boldsymbol{\gamma}) = \{n(n-1)\}^{-1}\sum_{i \neq j} h_{i,j}(\boldsymbol{\gamma})$ with $h_{i,j}(\boldsymbol{\gamma}) = |(e_i^{(2)} - e_j^{(2)}) - (\widehat{\mathbf{x}}_i^{(2)} - \widehat{\mathbf{x}}_j^{(2)})^{\mathrm{T}}\boldsymbol{\gamma}|$. We further define for simplicity that $\zeta_{i,j}^{(2)} = e_i^{(2)} - e_j^{(2)}$, and by the Knight's identity (Koenker, 2005), we have

$$\begin{aligned} Q^{(2)}(\boldsymbol{\gamma}) - Q^{(2)}(\mathbf{0}) &= 2\{n(n-1)\}^{-1}\sum_{i \neq j}(\widehat{\mathbf{x}}_i^{(2)} - \widehat{\mathbf{x}}_j^{(2)})^{\mathrm{T}}\boldsymbol{\gamma}\{I(\zeta_{i,j}^{(2)} \leq 0) - 1/2\} \\ &\quad + 2\{n(n-1)\}^{-1}\sum_{i \neq j}\int_0^{(\widehat{\mathbf{x}}_i^{(2)} - \widehat{\mathbf{x}}_j^{(2)})^{\mathrm{T}}\boldsymbol{\gamma}} \{I(\zeta_{i,j}^{(2)} \leq u) - I(\zeta_{i,j}^{(2)} \leq 0)\}\,\mathrm{d}u \\ &= 2Q_1^{(1)}(\boldsymbol{\gamma}) + 2Q_2^{(1)}(\boldsymbol{\gamma}). \end{aligned}$$

We further denote $E_{\boldsymbol{\epsilon}^{(2)}}(\cdot)$, $\mathrm{var}_{\boldsymbol{\epsilon}^{(2)}}(\cdot)$ and $\mathrm{pr}_{\boldsymbol{\epsilon}^{(2)}}(\cdot)$ as taking expectation, variance and probability with respect to $\boldsymbol{\epsilon}^{(2)} = (\epsilon_1^{(2)}, \ldots, \epsilon_n^{(2)})^{\mathrm{T}}$ only.

We first note that for any $\boldsymbol{\gamma} \in \Gamma^*$,

$$\begin{aligned} E_{\boldsymbol{\epsilon}^{(2)}}\{Q_1^{(1)}(\boldsymbol{\gamma})\} &= \{n(n-1)\}^{-1}\sum_{i \neq j}\Big\{(\widehat{\mathbf{x}}_i^{(2)} - \widehat{\mathbf{x}}_j^{(2)})^{\mathrm{T}}\boldsymbol{\gamma} \\ &\qquad \times [F^*\{\widehat{\Delta}_g^{(1)}(Z_i^{(2)}) - \widehat{\Delta}_g^{(1)}(Z_j^{(2)})\} - 1/2]\Big\} \\ &\geq -\Big\|\{n(n-1)\}^{-1}\sum_{i \neq j}\Big\{[F^*\{\widehat{\Delta}_g^{(1)}(Z_i^{(2)}) - \widehat{\Delta}_g^{(1)}(Z_j^{(2)})\} - 1/2] \\ &\qquad \times (\widehat{\mathbf{x}}_i^{(2)} - \widehat{\mathbf{x}}_j^{(2)})\Big\}\Big\|_\infty (1 + \bar{c})\sqrt{q}\delta h_n. \end{aligned}$$

On the one hand, by Condition (C4) and the mean value theorem, there exists some $\mu_{i,j}$ lying between 0 and $\widehat{\Delta}_g^{(1)}(Z_i^{(2)}) - \widehat{\Delta}_g^{(1)}(Z_j^{(2)})$, such that,

$$\begin{aligned} & \Big\|\{n(n-1)\}^{-1}\sum_{i \neq j}\Big([F^*\{\widehat{\Delta}_g^{(1)}(Z_i^{(2)}) - \widehat{\Delta}_g^{(1)}(Z_j^{(2)})\} - 1/2] \\ &\qquad \times \{\widehat{\boldsymbol{\Delta}}_{\mathbf{x}}^{(1)}(Z_i^{(2)}) - \widehat{\boldsymbol{\Delta}}_{\mathbf{x}}^{(1)}(Z_j^{(2)})\}\Big)\Big\|_\infty \\ =\ & \Big\|\{n(n-1)\}^{-1}\sum_{i \neq j}\Big[f^*(\mu_{i,j})\{\widehat{\Delta}_g^{(1)}(Z_i^{(2)}) - \widehat{\Delta}_g^{(1)}(Z_j^{(2)})\} \\ &\qquad \times \{\widehat{\boldsymbol{\Delta}}_{\mathbf{x}}^{(1)}(Z_i^{(2)}) - \widehat{\boldsymbol{\Delta}}_{\mathbf{x}}^{(1)}(Z_j^{(2)})\}\Big]\Big\|_\infty \leq 4\kappa_u r_n a_n, \end{aligned}$$

under Condition (C1) and on $\mathscr{N}$, where $\widehat{\boldsymbol{\Delta}}_{\mathbf{x}}^{(1)}(Z_i^{(2)}) = (\widehat{\Delta}_1^{(1)}(Z_i^{(2)}), \ldots, \widehat{\Delta}_p^{(1)}(Z_i^{(2)}))^{\mathrm{T}}$.

On the other hand, denote $H_{i,j,k} = [F^*\{\widehat{\Delta}_g^{(1)}(Z_i^{(2)}) - \widehat{\Delta}_g^{(1)}(Z_j^{(2)})\} - 1/2](\widetilde{X}_{i,k}^{(2)} - \widetilde{X}_{j,k}^{(2)})$. Then we claim that conditioning on $\mathcal{D}_1$, $H_{\pi(i),\pi(M_n+i),k}$ is zero-mean sub-Gaussian with parameter $4\sqrt{2}\kappa_u a_n \sigma$ under Condition (C2) and on $\mathscr{N}$. To see this, recall that $E_*|\widetilde{X}_{\pi(i),k}^{(2)} - \widetilde{X}_{\pi(M_n+i),k}^{(2)}|^t \le (4\sigma^2)^{t/2} t\Gamma(t/2)$ by Proposition 2.5.2 (ii) in Vershynin (2018), and then

$$\begin{aligned} & E_*\Big|[F^*\{\widehat{\Delta}_g^{(1)}(Z_{\pi(i)}^{(2)}) - \widehat{\Delta}_g^{(1)}(Z_{\pi(M_n+i)}^{(2)})\} - 1/2](\widetilde{X}_{\pi(i),k}^{(2)} - \widetilde{X}_{\pi(M_n+i),k}^{(2)})\Big|^t \\ \le\ & (4\kappa_u a_n \sigma)^t t\Gamma(t/2). \end{aligned}$$

Therefore,

$$\begin{aligned} & E_*\{\exp(\lambda^2 H_{\pi(i),\pi(M_n+i),k}^2)\} = E_*\Big\{1 + \sum_{t=1}^{\infty} (\lambda^2 H_{\pi(i),\pi(M_n+i),k}^2)^t / t!\Big\} \\ =\ & 1 + \sum_{t=1}^{\infty} \lambda^{2t} E_*(H_{\pi(i),\pi(M_n+i),k}^{2t})/t! \le 1 + 2\sum_{t=1}^{\infty} (4\lambda\kappa_u a_n\sigma)^{2t} \\ \le\ & \sum_{t=0}^{\infty} (4\sqrt{2}\lambda\kappa_u a_n\sigma)^{2t} = 1/(1 - 32\lambda^2\kappa_u^2 a_n^2\sigma^2). \end{aligned}$$

Similar to the proof of Vershynin (2018, Proposition 2.5.2), $E_*\{\exp(\lambda^2 H_{\pi(i),\pi(M_n+i),k}^2)\} \le \exp(64\lambda^2\kappa_u^2 a_n^2\sigma^2)$ for $|\lambda| \le (8\kappa_u a_n\sigma)^{-1}$, since $1/(1-x) \le \exp(2x)$ for $x \in [0, 1/2]$. Furthermore, when $\lambda \le (8\kappa_u a_n\sigma)^{-1}$,

$$\begin{aligned} E_*\{\exp(\lambda H_{\pi(i),\pi(M_n+i),k})\} &\le E_*\{\lambda H_{\pi(i),\pi(M_n+i),k} + \exp(\lambda^2 H_{\pi(i),\pi(M_n+i),k}^2)\} \\ &\le \exp(64\lambda^2\kappa_u^2 a_n^2\sigma^2), \end{aligned}$$

as $\exp(x) \le x + \exp(x^2)$. When $\lambda > (8\kappa_u a_n\sigma)^{-1}$,

$$\begin{aligned} E_*\{\exp(\lambda H_{\pi(i),\pi(M_n+i),k})\} &\le \exp(32\lambda^2\kappa_u^2 a_n^2\sigma^2) E_*[\exp\{H_{\pi(i),\pi(M_n+i),k}^2/(128\kappa_u^2 a_n^2\sigma^2)\}] \\ &\le \exp(32\lambda^2\kappa_u^2 a_n^2\sigma^2)\exp(1/2) \le \exp(64\lambda^2\kappa_u^2 a_n^2\sigma^2). \end{aligned}$$

That is, $H_{\pi(i),\pi(M_n+i),k}$ is sub-Gaussian with parameter at most $8\sqrt{2}\kappa_u a_n\sigma$. Denote the event

$$\begin{aligned} \mathscr{M}_2 =\ & \Big[\Big\|\{n(n-1)\}^{-1}\sum_{i\ne j}\Big\{[F^*\{\widehat{\Delta}_g^{(1)}(Z_i^{(2)}) - \widehat{\Delta}_g^{(1)}(Z_j^{(2)})\} - 1/2] \\ & \times(\widetilde{\mathbf{x}}_i^{(2)} - \widetilde{\mathbf{x}}_j^{(2)})\Big\}\Big\|_\infty \le 16\sqrt{2}C\kappa_u\sigma a_n\sqrt{\log p/n}\Big], \end{aligned}$$

then $\mathrm{pr}_*(\mathscr{M}_2^c) \le 2\exp\{-(C^2-1)\log p\}$ on $\mathscr{N}$.

To sum up, $\inf_{\boldsymbol{\gamma}\in\Gamma^*} E_{\boldsymbol{\epsilon}^{(2)}}\{Q_1^{(1)}(\boldsymbol{\gamma})\} \ge -4(1+\bar{c})\sqrt{q}\delta h_n\kappa_u a_n(r_n + 4\sqrt{2}C\sigma\sqrt{\log p/n})$ on $\mathscr{M}_2 \cap \mathscr{N}$.

As for $Q_2^{(1)}(\boldsymbol{\gamma})$, we have

$$E_{\boldsymbol{\epsilon}^{(2)}}\{Q_2^{(1)}(\boldsymbol{\gamma})\} = \{n(n-1)\}^{-1}\sum_{i\ne j}\int_0^{(\widehat{\mathbf{x}}_i^{(2)} - \widehat{\mathbf{x}}_j^{(2)})^{\mathrm{T}}\boldsymbol{\gamma}} \{F^*(u + E_{i,j}^0) - F^*(E_{i,j}^0)\}\,\mathrm{d}u,$$

where $E^0_{i,j} = \widehat{\Delta}^{(1)}_g(Z^{(2)}_i) - \widehat{\Delta}^{(1)}_g(Z^{(2)}_j)$. Again by the mean value theorem, there exists some $\mu_{i,j}$ lying between $E^0_{i,j}$ and $E^0_{i,j} + (\widehat{\mathbf{x}}^{(2)}_i - \widehat{\mathbf{x}}^{(2)}_j)^{\mathrm{T}}\boldsymbol{\gamma}$, such that

$$E_{\boldsymbol{\epsilon}^{(2)}}\{Q^{(1)}_2(\boldsymbol{\gamma})\} = \{n(n-1)\}^{-1}\sum_{i\neq j}\int_0^{|(\widehat{\mathbf{x}}^{(2)}_i - \widehat{\mathbf{x}}^{(2)}_j)^{\mathrm{T}}\boldsymbol{\gamma}|} f^*(\mu_{i,j})u\,\mathrm{d}u,$$

which we claim is lower bounded by $(\kappa_l/2)\{n(n-1)\}^{-1}\sum_{i\neq j}\{(\widehat{\mathbf{x}}^{(2)}_i - \widehat{\mathbf{x}}^{(2)}_j)^{\mathrm{T}}\boldsymbol{\gamma}\}^2$ on $\mathscr{N}$, provided that $2\{(2C\sigma L_n + r_n)(1+\bar{c})\sqrt{q}\delta h_n + a_n\} \leq \delta_0$. To see this, note that

$$\begin{aligned}
&\mathrm{pr}\Big(\max_i \|\widetilde{\mathbf{x}}^{(2)}_i\|_\infty \geq t\Big) \leq \inf_{\lambda>0} E\Big\{\exp\Big(\lambda \max_i \|\widetilde{\mathbf{x}}^{(2)}_i\|_\infty\Big)\Big\}\Big/\exp(\lambda t)\\
=\;& \inf_{\lambda>0} E\Big[\sum_{i=1}^n\sum_{k=1}^p\{\exp(\lambda\widetilde{X}^{(2)}_{i,k}) + \exp(-\lambda\widetilde{X}^{(2)}_{i,k})\}\Big]\Big/\exp(\lambda t)\\
\leq\;& 2\inf_{\lambda>0}\exp\Big\{\lambda^2\sigma^2/2 - \lambda t + \log(np)\Big\} \leq 2\exp\Big\{-t^2/(2\sigma^2) + 2\log(p\vee n)\Big\}.
\end{aligned}$$

Therefore,

$$\max_{i\neq j}|(\widehat{\mathbf{x}}^{(2)}_i - \widehat{\mathbf{x}}^{(2)}_j)^{\mathrm{T}}\boldsymbol{\gamma}| \leq 2(2C\sigma L_n + r_n)(1+\bar{c})\sqrt{q}\delta h_n,$$

under Condition (C1), and on $\mathscr{N}$ and $\mathscr{M}_3 = \{\max_i \|\widetilde{\mathbf{x}}^{(2)}_i\|_\infty \leq 2C\sigma L_n\}$, where $\mathrm{pr}(\mathscr{M}_3^c) \leq 2\exp\{-2(C^2-1)L_n^2\}$ and $L_n = \sqrt{\log(p\vee n)}$. Consequently, we have $f^*(\mu_{i,j}) \geq \kappa_l$ under Condition (C4), provided that $2\{(2C\sigma L_n + r_n)(1+\bar{c})\sqrt{q}\delta h_n + a_n\} \leq \delta_0$.

Then, we observe that

$$\begin{aligned}
\{(\widehat{\mathbf{x}}^{(2)}_i - \widehat{\mathbf{x}}^{(2)}_j)^{\mathrm{T}}\boldsymbol{\gamma}\}^2 =\;& \{(\widetilde{\mathbf{x}}^{(2)}_i - \widetilde{\mathbf{x}}^{(2)}_j)^{\mathrm{T}}\boldsymbol{\gamma}\}^2 + [\{\widehat{\boldsymbol{\Delta}}^{(1)}_{\mathbf{x}}(Z^{(2)}_i) - \widehat{\boldsymbol{\Delta}}^{(1)}_{\mathbf{x}}(Z^{(2)}_j)\}^{\mathrm{T}}\boldsymbol{\gamma}]^2\\
&-2[\boldsymbol{\gamma}^{\mathrm{T}}(\widetilde{\mathbf{x}}^{(2)}_i - \widetilde{\mathbf{x}}^{(2)}_j)\{\widehat{\boldsymbol{\Delta}}^{(1)}_{\mathbf{x}}(Z^{(2)}_i) - \widehat{\boldsymbol{\Delta}}^{(1)}_{\mathbf{x}}(Z^{(2)}_j)\}^{\mathrm{T}}\boldsymbol{\gamma}].
\end{aligned}$$

On the one hand, we have

$$\begin{aligned}
&\Big|2\{n(n-1)\}^{-1}\sum_{i\neq j}\boldsymbol{\gamma}^{\mathrm{T}}(\widetilde{\mathbf{x}}^{(2)}_i - \widetilde{\mathbf{x}}^{(2)}_j)\{\widehat{\boldsymbol{\Delta}}^{(1)}_{\mathbf{x}}(Z^{(2)}_i) - \widehat{\boldsymbol{\Delta}}^{(1)}_{\mathbf{x}}(Z^{(2)}_j)\}^{\mathrm{T}}\boldsymbol{\gamma}\Big|\\
\leq\;& 2\Big\|\{n(n-1)\}^{-1}\sum_{i\neq j}(\widetilde{\mathbf{x}}^{(2)}_i - \widetilde{\mathbf{x}}^{(2)}_j)\{\widehat{\boldsymbol{\Delta}}^{(1)}_{\mathbf{x}}(Z^{(2)}_i) - \widehat{\boldsymbol{\Delta}}^{(1)}_{\mathbf{x}}(Z^{(2)}_j)\}^{\mathrm{T}}\Big\|_\infty\|\boldsymbol{\gamma}\|_1^2\\
\leq\;& 64C(1+\bar{c})^2\sigma r_n q\delta^2 h_n^2\sqrt{\log p/n},
\end{aligned}$$

on $\mathscr{N}$ and the event

$$\begin{aligned}
\mathscr{M}_4 =\;& \Big[\Big\|\{n(n-1)\}^{-1}\sum_{i\neq j}(\widetilde{\mathbf{x}}^{(2)}_i - \widetilde{\mathbf{x}}^{(2)}_j)\{\widehat{\boldsymbol{\Delta}}^{(1)}_{\mathbf{x}}(Z^{(2)}_i) - \widehat{\boldsymbol{\Delta}}^{(1)}_{\mathbf{x}}(Z^{(2)}_j)\}^{\mathrm{T}}\Big\|_\infty\\
&\leq 32C\sigma r_n\sqrt{\log p/n}\Big].
\end{aligned}$$

Following similar arguments to $\mathscr{M}_2$, we note that $[(\widetilde{X}^{(2)}_{\pi(i),k} - \widetilde{X}^{(2)}_{\pi(M_n+i),k})\{\widehat{\Delta}^{(1)}_{k'}(Z^{(2)}_{\pi(i)}) - \widehat{\Delta}^{(1)}_{k'}(Z^{(2)}_{\pi(M_n+i)})\}]$, conditioning on $\mathcal{D}_1$, is zero-mean sub-Gaussian with parameter $8\sqrt{2}\sigma r_n$ under Condition (C2) and on $\mathscr{N}$. We then have $\mathrm{pr}_*(\mathscr{M}_4^c) \leq 2\exp\{-2(C^2-1)\log p\}$ on $\mathscr{N}$.

On the other hand, we note that for any $\boldsymbol{\gamma}' \in \mathbb{R}^p$,

$$\begin{aligned} n^{-1}\sum_{i=1}^n (\boldsymbol{\gamma}')^{\mathrm{T}}\widetilde{\mathbf{x}}_i^{(2)}(\widetilde{\mathbf{x}}_i^{(2)})^{\mathrm{T}}\boldsymbol{\gamma}' &= (\boldsymbol{\gamma}')^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{\gamma}' + (\boldsymbol{\gamma}')^{\mathrm{T}}\Big\{n^{-1}\sum_{i=1}^n \widetilde{\mathbf{x}}_i^{(2)}(\widetilde{\mathbf{x}}_i^{(2)})^{\mathrm{T}} - \boldsymbol{\Sigma}\Big\}\boldsymbol{\gamma}' \\ &\geq \lambda_{\min}(\boldsymbol{\Sigma})\|\boldsymbol{\gamma}'\|_2^2 - \Big\|n^{-1}\sum_{i=1}^n \widetilde{\mathbf{x}}_i^{(2)}(\widetilde{\mathbf{x}}_i^{(2)})^{\mathrm{T}} - \boldsymbol{\Sigma}\Big\|_\infty \|\boldsymbol{\gamma}'\|_1^2, \end{aligned}$$

and similarly,

$$n^{-1}\sum_{i=1}^n (\boldsymbol{\gamma}')^{\mathrm{T}}\widetilde{\mathbf{x}}_i^{(2)}(\widetilde{\mathbf{x}}_i^{(2)})^{\mathrm{T}}\boldsymbol{\gamma}' \leq \lambda_{\max}(\boldsymbol{\Sigma})\|\boldsymbol{\gamma}'\|_2^2 + \Big\|n^{-1}\sum_{i=1}^n \widetilde{\mathbf{x}}_i^{(2)}(\widetilde{\mathbf{x}}_i^{(2)})^{\mathrm{T}} - \boldsymbol{\Sigma}\Big\|_\infty \|\boldsymbol{\gamma}'\|_1^2.$$

It is well-known that the product of sub-Gaussian is sub-exponential (Vershynin, 2018, Lemma 2.7.7). In particular, since for all $k, k' = 1, \ldots, p$ and $t \geq 1$,

$$\begin{aligned} & E\big|\widetilde{X}_{i,k}^{(2)}\widetilde{X}_{i,k'}^{(2)} - E(\widetilde{X}_{i,k}^{(2)}\widetilde{X}_{i,k'}^{(2)})\big|^t = E\big|E(\widetilde{X}_{i,k}^{(2)}\widetilde{X}_{i,k'}^{(2)} - \widetilde{X}_{i,k}^{\prime(2)}\widetilde{X}_{i,k'}^{\prime(2)} \mid \widetilde{X}_{i,k}^{(2)}\widetilde{X}_{i,k'}^{(2)})\big|^t \\ \leq\ & E\big(|\widetilde{X}_{i,k}^{(2)}\widetilde{X}_{i,k'}^{(2)} - \widetilde{X}_{i,k}^{\prime(2)}\widetilde{X}_{i,k'}^{\prime(2)}|^t\big) = 2^t E\big\{|(\widetilde{X}_{i,k}^{(2)}\widetilde{X}_{i,k'}^{(2)} - \widetilde{X}_{i,k}^{\prime(2)}\widetilde{X}_{i,k'}^{\prime(2)})/2|^t\big\} \\ \leq\ & 2^t E\big(|\widetilde{X}_{i,k}^{(2)}\widetilde{X}_{i,k'}^{(2)}|^t\big) \leq 2^t\big\{E(|\widetilde{X}_{i,k}^{(2)}|^{2t})E(|\widetilde{X}_{i,k'}^{(2)}|^{2t})\big\}^{1/2}, \end{aligned}$$

where $(\widetilde{X}_{i,k}^{\prime(2)}, \widetilde{X}_{i,k'}^{\prime(2)})$ is an independent copy of $(\widetilde{X}_{i,k}^{(2)}, \widetilde{X}_{i,k'}^{(2)})$, the first two inequalities follow from Jensen's inequality, and the last inequality follows from Cauchy-Schwarz inequality. Then following a similar argument to Honorio and Jaakkola (2014, Appendix B), $\widetilde{X}_{i,k}^{(2)}\widetilde{X}_{i,k'}^{(2)} - E(\widetilde{X}_{i,k}\widetilde{X}_{i,k'})$ is sub-exponential with parameter $\big(8\sqrt{2}\sigma^2, 8\sigma^2\big)$, so that

$$\begin{aligned} & \mathrm{pr}\Big(\Big\|(1/n)\sum_{i=1}^n \widetilde{\mathbf{x}}_i^{(2)}(\widetilde{\mathbf{x}}_i^{(2)})^{\mathrm{T}} - \boldsymbol{\Sigma}\Big\|_\infty \geq t\Big) \\ \leq\ & \sum_{k'=1}^p\sum_{k=1}^p \inf_{\lambda\in[0,(8\sigma^2/n)^{-1})} E\Big\{\exp\Big(\lambda\Big|(1/n)\sum_{i=1}^n \widetilde{X}_{i,k}^{(2)}\widetilde{X}_{i,k'}^{(2)} - E(\widetilde{X}_{i,k}^{(2)}\widetilde{X}_{i,k'}^{(2)})\Big|\Big)\Big\}\Big/\exp(\lambda t) \\ \leq\ & 2\sum_{k'=1}^p\sum_{k=1}^p \exp\Big\{-nt^2/\max(256\sigma^4, 16\sigma^2 t)\Big\} \\ =\ & 2\exp\Big\{-nt^2/\max(256\sigma^4, 16\sigma^2 t) + 2\log p\Big\}. \end{aligned}$$

For notational simplicity, denote

$$\mathscr{L} := \mathscr{L}(t) = \Big[\Big\|(1/n)\sum_{i=1}^n \widetilde{\mathbf{x}}_i^{(2)}(\widetilde{\mathbf{x}}_i^{(2)})^{\mathrm{T}} - \boldsymbol{\Sigma}\Big\|_\infty \leq t\Big].$$

This means when $t = \lambda_{\min}(\boldsymbol{\Sigma})/\{4(1+\bar{c})^2 q\}$, we have

$$\begin{aligned} n^{-1}\sum_{i=1}^n (\boldsymbol{\gamma}')^{\mathrm{T}}\widetilde{\mathbf{x}}_i^{(2)}(\widetilde{\mathbf{x}}_i^{(2)})^{\mathrm{T}}\boldsymbol{\gamma}' &\geq \lambda_{\min}(\boldsymbol{\Sigma})\|\boldsymbol{\gamma}'\|_2^2 - \lambda_{\min}(\boldsymbol{\Sigma})/\{4(1+\bar{c})^2 q\}\|\boldsymbol{\gamma}'\|_1^2, \\ n^{-1}\sum_{i=1}^n (\boldsymbol{\gamma}')^{\mathrm{T}}\widetilde{\mathbf{x}}_i^{(2)}(\widetilde{\mathbf{x}}_i^{(2)})^{\mathrm{T}}\boldsymbol{\gamma}' &\leq \lambda_{\max}(\boldsymbol{\Sigma})\|\boldsymbol{\gamma}'\|_2^2 + \lambda_{\min}(\boldsymbol{\Sigma})/\{4(1+\bar{c})^2 q\}\|\boldsymbol{\gamma}'\|_1^2, \end{aligned}$$

on $\mathscr{L}$, and

$$\begin{aligned}\mathrm{pr}(\mathscr{L}^c) &\le 2\exp\Big\{-n[\lambda_{\min}(\boldsymbol{\Sigma})/\{4(1+\bar{c})^2q\}]^2 \\ &\qquad \Big/\max[256\sigma^4, 4\sigma^2\lambda_{\min}(\boldsymbol{\Sigma})/\{(1+\bar{c})^2q\}] + 2\log p\Big\}.\end{aligned}$$

Because $\boldsymbol{\gamma}\in\Gamma^*$, we have $\|\boldsymbol{\gamma}\|_1^2 \le (1+\bar{c})^2q\|\boldsymbol{\gamma}\|_2^2$, and

$$\begin{aligned}n^{-1}\sum_{i=1}^n\{(\widetilde{\mathbf{x}}_i^{(2)})^{\mathrm{T}}\boldsymbol{\gamma}\}^2 &\ge (3/4)\lambda_{\min}(\boldsymbol{\Sigma})\|\boldsymbol{\gamma}\|_2^2 = (3/4)\lambda_{\min}(\boldsymbol{\Sigma})\delta^2h_n^2,\\ n^{-1}\sum_{i=1}^n\{(\widetilde{\mathbf{x}}_i^{(2)})^{\mathrm{T}}\boldsymbol{\gamma}\}^2 &\le (5/4)\lambda_{\max}(\boldsymbol{\Sigma})\|\boldsymbol{\gamma}\|_2^2 \le 5\sigma^2\delta^2h_n^2,\end{aligned}$$

where the last inequality holds because for any unit $\boldsymbol{\alpha}\in\mathbb{R}^p$,

$$\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{\alpha} = E(\boldsymbol{\alpha}^{\mathrm{T}}\widetilde{\mathbf{x}}\widetilde{\mathbf{x}}^{\mathrm{T}}\boldsymbol{\alpha}) = E\{(\widetilde{\mathbf{x}}^{\mathrm{T}}\boldsymbol{\alpha})^2\} \le 4\sigma^2,$$

under Condition (C2) (see, for example, Vershynin, 2018, Proposition 2.5.2).

Besides, denote

$$\mathscr{M}_5 := \mathscr{M}_5(t) = \Big[\Big\|\{n(n-1)\}^{-1}\sum_{i\ne j}\widetilde{\mathbf{x}}_i^{(2)}(\widetilde{\mathbf{x}}_j^{(2)})^{\mathrm{T}}\Big\|_\infty \le t\Big],$$

for simplicity. This means when $t = \lambda_{\min}(\boldsymbol{\Sigma})/\{4(1+\bar{c})^2q\}$, we have

$$\begin{aligned}&\Big|2\{n(n-1)\}^{-1}\sum_{i\ne j}\boldsymbol{\gamma}^{\mathrm{T}}\widetilde{\mathbf{x}}_i^{(2)}(\widetilde{\mathbf{x}}_j^{(2)})^{\mathrm{T}}\boldsymbol{\gamma}\Big|\\ \le\ & 2\Big\|\{n(n-1)\}^{-1}\sum_{i\ne j}\widetilde{\mathbf{x}}_i^{(2)}(\widetilde{\mathbf{x}}_j^{(2)})^{\mathrm{T}}\Big\|_\infty\|\boldsymbol{\gamma}\|_1^2 \le (1/2)\lambda_{\min}(\boldsymbol{\Sigma})\delta^2h_n^2,\end{aligned}$$

for any $\boldsymbol{\gamma}\in\Gamma^*$ and on $\mathscr{M}_5$. Because

$$\begin{aligned}&\mathrm{pr}\Big(\Big\|\{n(n-1)\}^{-1}\sum_{i\ne j}\widetilde{\mathbf{x}}_i^{(2)}(\widetilde{\mathbf{x}}_j^{(2)})^{\mathrm{T}}\Big\|_\infty \ge t\Big)\\ \le\ & \sum_{k'=1}^p\sum_{k=1}^p\inf_{\lambda\in[0,(2\sqrt{2}\sigma^2/n)^{-1})}\frac{1}{n!}\\ &\qquad\sum_{\pi}E\Big\{\exp\Big(\lambda\Big|M_n^{-1}\sum_{i=1}^{M_n}\widetilde{X}_{\pi(i),k}^{(2)}\widetilde{X}_{\pi(M_n+i),k'}^{(2)}\Big|\Big)\Big\}\Big/\exp(\lambda t)\\ \le\ & 2\sum_{k'=1}^p\sum_{k=1}^p\exp\Big\{-nt^2/\max(8\sigma^4, 4\sqrt{2}\sigma^2t)\Big\}\\ =\ & 2\exp\Big\{-nt^2/\max(8\sigma^4, 4\sqrt{2}\sigma^2t) + 2\log p\Big\},\end{aligned}$$

where the second inequality follows from the fact that the product of two independent sub-Gaussian random variables with parameter σ is sub-exponential with parameter $\big(\sqrt{2}\sigma^2, \sqrt{2}\sigma^2\big)$ (Wainwright, 2019, Exercise 2.13), we then have

$$\begin{aligned}\mathrm{pr}(\mathscr{M}_5^c) \;\leq\; & 2\exp\Big\{ - n[\lambda_{\min}(\mathbf{\Sigma})/\{4(1+\bar{c})^2 q\}]^2 \\ & \qquad \Big/ \max[8\sigma^4, \sqrt{2}\sigma^2\lambda_{\min}(\mathbf{\Sigma})/\{(1+\bar{c})^2 q\}] + 2\log p\Big\}.\end{aligned}$$

Putting the pieces together, we have $\inf_{\boldsymbol{\gamma}\in\Gamma^*} E_{\boldsymbol{\epsilon}^{(2)}}\{Q_2^{(1)}(\boldsymbol{\gamma})\} \geq \kappa_l \delta^2 h_n^2\{(1/2)\lambda_{\min}(\mathbf{\Sigma}) - 32C(1+\bar{c})^2\sigma r_n q\sqrt{\log p/n}\}$ on $\mathscr{L}\cap\mathscr{N}\cap\mathscr{M}_3\cap\mathscr{M}_4\cap\mathscr{M}_5$.

Since we have known that

$$\begin{aligned} & \inf_{\boldsymbol{\gamma}\in\Gamma^*}\{L^{(2)}(\boldsymbol{\gamma}) - L^{(2)}(\mathbf{0})\} \\ \geq\; & \inf_{\boldsymbol{\gamma}\in\Gamma^*} E_{\boldsymbol{\epsilon}^{(2)}}\{Q^{(2)}(\boldsymbol{\gamma}) - Q^{(2)}(\mathbf{0})\} \\ & - \sup_{\boldsymbol{\gamma}\in\Gamma^*} |Q^{(2)}(\boldsymbol{\gamma}) - Q^{(2)}(\mathbf{0}) - E_{\boldsymbol{\epsilon}^{(2)}}\{Q^{(2)}(\boldsymbol{\gamma}) - Q^{(2)}(\mathbf{0})\}| \\ & - \sup_{\boldsymbol{\gamma}\in\Gamma^*} \widehat{\lambda}^{(1)} |\|\boldsymbol{\beta}_0 + \boldsymbol{\gamma}\|_1 - \|\boldsymbol{\beta}_0\|_1|,\end{aligned}$$

it remains to deal with the following term

$$W_n = \sup_{\boldsymbol{\gamma}\in\Gamma^*} |Q^{(2)}(\boldsymbol{\gamma}) - Q^{(2)}(\mathbf{0}) - E_{\boldsymbol{\epsilon}^{(2)}}\{Q^{(2)}(\boldsymbol{\gamma}) - Q^{(2)}(\mathbf{0})\}|.$$

By Markov's inequality, for any $\lambda > 0$,

$$\begin{aligned} & \mathrm{pr}_{\boldsymbol{\epsilon}^{(2)}}(W_n \geq t) \\ \leq\; & E_{\boldsymbol{\epsilon}^{(2)}}\Big\{ \exp\Big(\lambda \sup_{\boldsymbol{\gamma}\in\Gamma^*} \Big| \{n(n-1)\}^{-1} \sum_{i\neq j} \Big[h_{i,j}(\boldsymbol{\gamma}) - h_{i,j}(\mathbf{0}) \\ & \qquad\qquad - E_{\boldsymbol{\epsilon}^{(2)}}\{h_{i,j}(\boldsymbol{\gamma}) - h_{i,j}(\mathbf{0})\}\Big]\Big|\Big)\Big\} \Big/ \exp(\lambda t) \\ =\; & E_{\boldsymbol{\epsilon}^{(2)}}\Big\{ \exp\Big(\lambda \sup_{\boldsymbol{\gamma}\in\Gamma^*} \Big| \frac{1}{n!}\sum_{\pi} M_n^{-1} \sum_{i=1}^{M_n} \Big[h_{\pi(i),\pi(M_n+i)}(\boldsymbol{\gamma}) - h_{\pi(i),\pi(M_n+i)}(\mathbf{0}) \\ & \qquad\qquad - E_{\boldsymbol{\epsilon}^{(2)}}\{h_{\pi(i),\pi(M_n+i)}(\boldsymbol{\gamma}) - h_{\pi(i),\pi(M_n+i)}(\mathbf{0})\}\Big]\Big|\Big)\Big\} \Big/ \exp(\lambda t) \\ \leq\; & \frac{1}{n!}\sum_{\pi} E_{\boldsymbol{\epsilon}^{(2)}}\Big\{ \exp\Big(\lambda \sup_{\boldsymbol{\gamma}\in\Gamma^*} \Big| M_n^{-1} \sum_{i=1}^{M_n} \Big[h_{\pi(i),\pi(M_n+i)}(\boldsymbol{\gamma}) - h_{\pi(i),\pi(M_n+i)}(\mathbf{0}) \\ & \qquad\qquad - E_{\boldsymbol{\epsilon}^{(2)}}\{h_{\pi(i),\pi(M_n+i)}(\boldsymbol{\gamma}) - h_{\pi(i),\pi(M_n+i)}(\mathbf{0})\}\Big]\Big|\Big)\Big\} \Big/ \exp(\lambda t).\end{aligned}$$

Write

$$\begin{aligned}\omega_+(\pi) &= \sup_{\boldsymbol{\gamma}\in\Gamma^*} M_n^{-1}\sum_{i=1}^{M_n}\Big[h_{\pi(i),\pi(M_n+i)}(\boldsymbol{\gamma}) - h_{\pi(i),\pi(M_n+i)}(\mathbf{0}) \\ &\qquad -E_{\boldsymbol{\epsilon}^{(2)}}\{h_{\pi(i),\pi(M_n+i)}(\boldsymbol{\gamma}) - h_{\pi(i),\pi(M_n+i)}(\mathbf{0})\}\Big], \\ \omega_-(\pi) &= \sup_{\boldsymbol{\gamma}\in\Gamma^*} -M_n^{-1}\sum_{i=1}^{M_n}\Big[h_{\pi(i),\pi(M_n+i)}(\boldsymbol{\gamma}) - h_{\pi(i),\pi(M_n+i)}(\mathbf{0}) \\ &\qquad -E_{\boldsymbol{\epsilon}^{(2)}}\{h_{\pi(i),\pi(M_n+i)}(\boldsymbol{\gamma}) - h_{\pi(i),\pi(M_n+i)}(\mathbf{0})\}\Big].\end{aligned}$$

Given the set Γ^*, we define its ε-covering $\Gamma^*(\varepsilon)$ following Wainwright (2019, Chapter 5), and denote the covering number by $|\Gamma^*(\varepsilon)| = N(\varepsilon)$. We will show that the concentration inequalities for the supremum over $\Gamma^*(\varepsilon)$ are independent of $N(\varepsilon)$. Therefore, by letting $\varepsilon \to 0$, we can extend the results to Γ^*. For simplicity, we will not distinguish between Γ^* and $\Gamma^*(\varepsilon)$ in the following analysis. Using Massart (2000, Lemma 8), we have

$$\begin{aligned}&\lambda E_{\boldsymbol{\epsilon}^{(2)}}[\omega_\cdot(\pi)\exp\{\lambda\omega_\cdot(\pi)\}] - E_{\boldsymbol{\epsilon}^{(2)}}[\exp\{\lambda\omega_\cdot(\pi)\}]\log E_{\boldsymbol{\epsilon}^{(2)}}[\exp\{\lambda\omega_\cdot(\pi)\}] \\ \le\ &\sum_{i=1}^{M_n} E_{\boldsymbol{\epsilon}^{(2)}}(\exp\{\lambda\omega_\cdot(\pi)\}\phi[-\lambda\{\omega_\cdot(\pi) - \omega_\cdot^{\setminus i}(\pi)\}]),\end{aligned}$$

where $\phi(u) = \exp(u) - u - 1$, and $\omega_\cdot^{\setminus i}(\pi)$ is defined similarly to $\omega_\cdot(\pi)$, but with $h_{\pi(i),\pi(M_n+i)}(\boldsymbol{\gamma}) - h_{\pi(i),\pi(M_n+i)}(\mathbf{0}) - E_{\boldsymbol{\epsilon}^{(2)}}\{h_{\pi(i),\pi(M_n+i)}(\boldsymbol{\gamma}) - h_{\pi(i),\pi(M_n+i)}(\mathbf{0})\}$ replaced by $-2|(\widehat{\mathbf{x}}^{(2)}_{\pi(i)} - \widehat{\mathbf{x}}^{(2)}_{\pi(M_n+i)})^{\mathrm{T}}\boldsymbol{\gamma}|$ for $\omega_+^{\setminus i}(\pi)$ and $2|(\widehat{\mathbf{x}}^{(2)}_{\pi(i)} - \widehat{\mathbf{x}}^{(2)}_{\pi(M_n+i)})^{\mathrm{T}}\boldsymbol{\gamma}|$ for $\omega_-^{\setminus i}(\pi)$, respectively. Since

$$|h_{\pi(i),\pi(M_n+i)}(\boldsymbol{\gamma}) - h_{\pi(i),\pi(M_n+i)}(\mathbf{0})| \le |(\widehat{\mathbf{x}}^{(2)}_{\pi(i)} - \widehat{\mathbf{x}}^{(2)}_{\pi(M_n+i)})^{\mathrm{T}}\boldsymbol{\gamma}|,$$

we have

$$\begin{aligned}&|h_{\pi(i),\pi(M_n+i)}(\boldsymbol{\gamma}) - h_{\pi(i),\pi(M_n+i)}(\mathbf{0}) - E_{\boldsymbol{\epsilon}^{(2)}}\{h_{\pi(i),\pi(M_n+i)}(\boldsymbol{\gamma}) - h_{\pi(i),\pi(M_n+i)}(\mathbf{0})\}| \\ \le\ &2|(\widehat{\mathbf{x}}^{(2)}_{\pi(i)} - \widehat{\mathbf{x}}^{(2)}_{\pi(M_n+i)})^{\mathrm{T}}\boldsymbol{\gamma}|.\end{aligned}$$

Therefore, for $\boldsymbol{\gamma}^*$ such that

$$\begin{aligned}\omega_+(\pi) &= M_n^{-1}\sum_{i=1}^{M_n}\Big[h_{\pi(i),\pi(M_n+i)}(\boldsymbol{\gamma}^*) - h_{\pi(i),\pi(M_n+i)}(\mathbf{0}) \\ &\qquad -E_{\boldsymbol{\epsilon}^{(2)}}\{h_{\pi(i),\pi(M_n+i)}(\boldsymbol{\gamma}^*) - h_{\pi(i),\pi(M_n+i)}(\mathbf{0})\}\Big],\end{aligned}$$

or

$$\begin{aligned}\omega_-(\pi) &= -M_n^{-1}\sum_{i=1}^{M_n}\Big[h_{\pi(i),\pi(M_n+i)}(\boldsymbol{\gamma}^*) - h_{\pi(i),\pi(M_n+i)}(\mathbf{0}) \\ &\qquad -E_{\boldsymbol{\epsilon}^{(2)}}\{h_{\pi(i),\pi(M_n+i)}(\boldsymbol{\gamma}^*) - h_{\pi(i),\pi(M_n+i)}(\mathbf{0})\}\Big],\end{aligned}$$

we have

$$0 \leq \omega_{\cdot}(\pi) - \omega_{\cdot}^{\setminus i}(\pi) \leq 4M_n^{-1}|(\widehat{\mathbf{x}}^{(2)}_{\pi(i)} - \widehat{\mathbf{x}}^{(2)}_{\pi(M_n+i)})^{\mathrm{T}}\boldsymbol{\gamma}^*|.$$

Note that $\phi(u)/u^2$ is nondecreasing on $\mathbb{R}$, $\phi(u) \leq u^2/2$ for $u \leq 0$, which implies

$$\begin{aligned} & \sum_{i=1}^{M_n} E_{\boldsymbol{\epsilon}^{(2)}}(\exp\{\lambda\omega_{\cdot}(\pi)\}\phi[-\lambda\{\omega_{\cdot}(\pi) - \omega_{\cdot}^{\setminus i}(\pi)\}]) \\ \leq & \sum_{i=1}^{M_n} 8M_n^{-2}\lambda^2\{(\widehat{\mathbf{x}}^{(2)}_{\pi(i)} - \widehat{\mathbf{x}}^{(2)}_{\pi(M_n+i)})^{\mathrm{T}}\boldsymbol{\gamma}^*\}^2 E_{\boldsymbol{\epsilon}^{(2)}}[\exp\{\lambda\omega_{\cdot}(\pi)\}], \end{aligned}$$

and

$$\begin{aligned} & \lambda E_{\boldsymbol{\epsilon}^{(2)}}[\omega_{\cdot}(\pi)\exp\{\lambda\omega_{\cdot}(\pi)\}] - E_{\boldsymbol{\epsilon}^{(2)}}[\exp\{\lambda\omega_{\cdot}(\pi)\}]\log E_{\boldsymbol{\epsilon}^{(2)}}[\exp\{\lambda\omega_{\cdot}(\pi)\}] \\ \leq & \sup_{\boldsymbol{\gamma}\in\Gamma^*} M_n^{-2}\sum_{i=1}^{M_n}\{(\widehat{\mathbf{x}}^{(2)}_{\pi(i)} - \widehat{\mathbf{x}}^{(2)}_{\pi(M_n+i)})^{\mathrm{T}}\boldsymbol{\gamma}\}^2 8\lambda^2 E_{\boldsymbol{\epsilon}^{(2)}}[\exp\{\lambda\omega_{\cdot}(\pi)\}]. \end{aligned}$$

On the one hand, on $\mathscr{N}$, we have that

$$M_n^{-1}\sum_{i=1}^{M_n}[\{\widehat{\boldsymbol{\Delta}}^{(1)}_{\mathbf{x}}(Z^{(2)}_{\pi(i)}) - \widehat{\boldsymbol{\Delta}}^{(1)}_{\mathbf{x}}(Z^{(2)}_{\pi(M_n+i)})\}^{\mathrm{T}}\boldsymbol{\gamma}]^2 \leq 4(1+\bar{c})^2 r_n^2 q\delta^2 h_n^2.$$

On the other hand, on $\mathscr{L}$,

$$M_n^{-1}\sum_{i=1}^{M_n}\{(\widetilde{\mathbf{x}}^{(2)}_{\pi(i)} - \widetilde{\mathbf{x}}^{(2)}_{\pi(M_n+i)})^{\mathrm{T}}\boldsymbol{\gamma}\}^2 \leq (4/n)\sum_{i=1}^{n}\{(\widetilde{\mathbf{x}}^{(2)}_i)^{\mathrm{T}}\boldsymbol{\gamma}\}^2 \leq 20\sigma^2\delta^2 h_n^2.$$

Consequently, on $\mathscr{L}\cap\mathscr{N}$,

$$\begin{aligned} & \lambda E_{\boldsymbol{\epsilon}^{(2)}}[\omega_{\cdot}(\pi)\exp\{\lambda\omega_{\cdot}(\pi)\}] - E_{\boldsymbol{\epsilon}^{(2)}}[\exp\{\lambda\omega_{\cdot}(\pi)\}]\log E_{\boldsymbol{\epsilon}^{(2)}}[\exp\{\lambda\omega_{\cdot}(\pi)\}] \\ \leq & 128\{(1+\bar{c})^2 r_n^2 q + 5\sigma^2\}(1/n)\delta^2 h_n^2\lambda^2 E_{\boldsymbol{\epsilon}^{(2)}}[\exp\{\lambda\omega_{\cdot}(\pi)\}]. \end{aligned}$$

This inequality holds for all positive λ and therefore yields the differential inequality

$$(1/\lambda)\{T'(\lambda)/T(\lambda)\} - (1/\lambda^2)\log T(\lambda) \leq 128\{(1+\bar{c})^2 r_n^2 q + 5\sigma^2\}(1/n)\delta^2 h_n^2,$$

for $T(\lambda) = E_{\boldsymbol{\epsilon}^{(2)}}\{\exp(\lambda[\omega_{\cdot}(\pi) - E_{\boldsymbol{\epsilon}^{(2)}}\{\omega_{\cdot}(\pi)\}])\}$. Setting $H(\lambda) = (1/\lambda)\log T(\lambda)$, the differential inequality becomes $H'(\lambda) \leq 128\{(1+\bar{c})^2 r_n^2 q + 5\sigma^2\}(1/n)\delta^2 h_n^2$, which in turn implies since $H(\lambda)$ tends to 0 as λ tends to 0, $H(\lambda) \leq 128\{(1+\bar{c})^2 r_n^2 q + 5\sigma^2\}(1/n)\delta^2 h_n^2\lambda$. Equivalently, on $\mathscr{L}\cap\mathscr{N}$,

$$E_{\boldsymbol{\epsilon}^{(2)}}\{\exp(\lambda[\omega_{\cdot}(\pi) - E_{\boldsymbol{\epsilon}^{(2)}}\{\omega_{\cdot}(\pi)\}])\} \leq \exp[128\{(1+\bar{c})^2 r_n^2 q + 5\sigma^2\}(1/n)\delta^2 h_n^2\lambda^2].$$

Putting the pieces together, on $\mathscr{L}\cap\mathscr{N}$,

$$\begin{aligned} \mathrm{pr}_{\boldsymbol{\epsilon}^{(2)}}(W_n \geq t) \quad \leq \quad & \frac{2}{n!}\sum_{\pi}\exp\Big(128\{(1+\bar{c})^2 r_n^2 q + 5\sigma^2\}(1/n)\delta^2 h_n^2\lambda^2 \\ & -\lambda[t - E_{\boldsymbol{\epsilon}^{(2)}}\{\omega_{\cdot}(\pi)\}]\Big). \end{aligned}$$

It thus suffices to deal with $E_{\boldsymbol{\epsilon}^{(2)}}\{\omega_\cdot(\pi)\}$:

$$\begin{aligned}
E_{\boldsymbol{\epsilon}^{(2)}}\{\omega_\cdot(\pi)\} &\le 2E_{\boldsymbol{\epsilon}^{(2)},\boldsymbol{\sigma}}\Big[\sup_{\boldsymbol{\gamma}\in\Gamma^*}\Big|M_n^{-1}\sum_{i=1}^{M_n}\sigma_i\{h_{\pi(i),\pi(M_n+i)}(\boldsymbol{\gamma})-h_{\pi(i),\pi(M_n+i)}(\mathbf{0})\}\Big|\Big]\\
&\le 4E_{\boldsymbol{\epsilon}^{(2)},\boldsymbol{\sigma}}\Big\{\sup_{\boldsymbol{\gamma}\in\Gamma^*}\Big|M_n^{-1}\sum_{i=1}^{M_n}\sigma_i(\widehat{\mathbf{x}}^{(2)}_{\pi(i)}-\widehat{\mathbf{x}}^{(2)}_{\pi(M_n+i)})^{\mathrm{T}}\boldsymbol{\gamma}\Big|\Big\}\\
&\le 4\sup_{\boldsymbol{\gamma}\in\Gamma^*}\|\boldsymbol{\gamma}\|_1 E_{\boldsymbol{\epsilon}^{(2)},\boldsymbol{\sigma}}\Big\{\max_k\Big|M_n^{-1}\sum_{i=1}^{M_n}\sigma_i(\widehat{X}^{(2)}_{\pi(i),k}-\widehat{X}^{(2)}_{\pi(M_n+i),k})\Big|\Big\}\\
&\le 32\sup_{\boldsymbol{\gamma}\in\Gamma^*}\|\boldsymbol{\gamma}\|_1\Big\{\max_k n^{-1}\sum_{i=1}^{n}(\widetilde{X}^{(2)}_{i,k})^2+r_n^2\Big\}^{1/2}\sqrt{\log p/n},
\end{aligned}$$

where $\boldsymbol{\sigma}=(\sigma_1,\ldots,\sigma_{M_n})^{\mathrm{T}}$ are i.i.d. Rademacher random variables, the first inequality follows from the property of U-statistic, the symmetrization technique (Wainwright, 2019) and Ledoux and Talagrand (1991, Theorem 4.12), and the last equality follows from Wainwright (2019, Exercise 2.12). On $\mathscr{M}_1\cap\mathscr{N}$, $E_{\boldsymbol{\epsilon}^{(2)}}\{\omega_\cdot(\pi)\}\le 32(1+\bar{c})(12\sigma^2+r_n^2)^{1/2}\delta h_n^2$. As a result, on $\mathscr{L}\cap\mathscr{N}\cap\mathscr{M}_1$,

$$\begin{aligned}
\mathrm{pr}_{\boldsymbol{\epsilon}^{(2)}}(W_n\ge t) &\le 2\inf_{\lambda>0}\exp\Big[128\{(1+\bar{c})^2r_n^2q+5\sigma^2\}(1/n)\delta^2h_n^2\lambda^2\\
&\qquad -\lambda\{t-32(1+\bar{c})(12\sigma^2+r_n^2)^{1/2}\delta h_n^2\}\Big]\\
&\le 2\exp\Big\{-n\{t-32(1+\bar{c})(12\sigma^2+r_n^2)^{1/2}\delta h_n^2\}^2\\
&\qquad \Big/[512\{(1+\bar{c})^2r_n^2q+5\sigma^2\}\delta^2h_n^2]\Big\}.
\end{aligned}$$

That is, define

$$\begin{aligned}
\mathscr{W} &= \Big\{W_n\le 16\sqrt{2}C\{(1+\bar{c})^2r_n^2q+5\sigma^2\}^{1/2}\delta h_n\sqrt{\log p/n}\\
&\qquad +32(1+\bar{c})(12\sigma^2+r_n^2)^{1/2}\delta h_n^2\Big\},
\end{aligned}$$

we have $\mathrm{pr}(\mathscr{W}^c\cap\mathscr{L}\cap\mathscr{N}\cap\mathscr{M}_1)\le 2\exp(-C^2\log p)$.

Putting the pieces together, we have

$$\begin{aligned}
&\inf_{\boldsymbol{\gamma}\in\Gamma^*}\{L^{(2)}(\boldsymbol{\gamma})-L^{(2)}(\mathbf{0})\}\\
>\ & \kappa_l\delta^2h_n^2\{\lambda_{\min}(\boldsymbol{\Sigma})-64C(1+\bar{c})^2\sigma r_nq\sqrt{\log p/n}\}\\
&-8(1+\bar{c})\sqrt{q}\delta h_n\kappa_u a_n(r_n+4\sqrt{2}C\sigma\sqrt{\log p/n})\\
&-64(1+\bar{c})C(3\sigma+r_n)\delta h_n^2-c_1(1+\bar{c})\delta h_n^2\\
\ge\ & \kappa_l\delta^2h_n^2\lambda_{\min}(\boldsymbol{\Sigma})-64C(1+\bar{c})\sigma\kappa_u\delta_0\delta h_n^2-c_0(1+\bar{c})\delta h_n^2\\
&-64C(1+\bar{c})(3\sigma+r_n)\delta h_n^2-c_1(1+\bar{c})\delta h_n^2\\
\ge\ & \kappa_l\delta^2h_n^2\lambda_{\min}(\boldsymbol{\Sigma})-64C(1+\bar{c})\{4\sigma(\kappa_u\delta_0\vee 1)+r_n\}\delta h_n^2-2c_1(1+\bar{c})\delta h_n^2,
\end{aligned}$$

on $\mathscr{W} \cap \mathscr{N} \cap \mathscr{L} \cap \mathscr{M}_1 \cap \mathscr{M}_2 \cap \mathscr{M}_3 \cap \mathscr{M}_4 \cap \mathscr{M}_5$, provided that $2\{(2C\sigma L_n + r_n)(1+\bar{c})\sqrt{q}\delta h_n + a_n\} \leq \delta_0$ and $8\kappa_u r_n a_n \leq c_0\sqrt{\log p/n}$. By letting $\delta = 2(1+\bar{c})[c_1 + 32C\{4\sigma(\kappa_u\delta_0 \vee 1) + r_n\}]/\{\kappa_l\lambda_{\min}(\mathbf{\Sigma})\}$, $\inf_{\boldsymbol{\gamma}\in\Gamma^*}\{L^{(2)}(\boldsymbol{\gamma}) - L^{(2)}(\mathbf{0})\} > 0$. Therefore, for each $l \in \{1,2\}$, we have

$$\begin{aligned}
& \mathrm{pr}[\|\widehat{\boldsymbol{\gamma}}^{(l)}(\widehat{\lambda}^{(l)})\|_2 \leq \delta\sqrt{q\log p/n}] \\
\geq\ & 1 - \mathrm{pr}\{(\mathscr{W} \cap \mathscr{L} \cap \mathscr{M}_1 \cap \mathscr{M}_2 \cap \mathscr{M}_3 \cap \mathscr{M}_4 \cap \mathscr{M}_5)^c \cap \mathscr{N}\} - \mathrm{pr}(\mathscr{N}^c) - \alpha_0 - \alpha_n \\
\geq\ & 1 - 8\exp\{-(C^2-1)\log p\} \\
& -c_2\exp(-c_3 n[\{\lambda_{\min}(\mathbf{\Sigma})/q\}^2/\sigma^4 \wedge 1] + \log p) - \rho_n - \alpha_0 - \alpha_n,
\end{aligned}$$

for some universal constants $c_2, c_3 > 0$. Furthermore,

$$\|\widehat{\boldsymbol{\gamma}}^{(l)}(\widehat{\lambda}^{(l)})\|_1 \leq (1+\bar{c})\sqrt{q}\|\widehat{\boldsymbol{\gamma}}^{(l)}(\widehat{\lambda}^{(l)})\|_2 \leq (1+\bar{c})\delta q\sqrt{\log p/n},$$

with probability at least $1 - 8\exp\{-(C^2-1)\log p\} - c_2\exp(-c_3 n[\{\lambda_{\min}(\mathbf{\Sigma})/q\}^2/\sigma^4 \wedge 1] + \log p) - \rho_n - \alpha_0 - \alpha_n$ as well. Then the proof is completed. ■

Remark 8 *The proof of Theorem 2 can be adapted to obtain similar results for the ANOPE (Fan et al., 2020b) subsampled version of DS Rank Lasso. Specifically, the loss function now becomes*

$$Q^{(2)}(\boldsymbol{\gamma}) = (M_n|\mathcal{S}|)^{-1}\sum_{\pi\in\mathcal{S}}\sum_{i=1}^{M_n}|(e^{(2)}_{\pi(i)} - e^{(2)}_{\pi(M_n+i)}) - (\widehat{\mathbf{x}}^{(2)}_{\pi(i)} - \widehat{\mathbf{x}}^{(2)}_{\pi(M_n+i)})^T\boldsymbol{\gamma}|,$$

where $\mathcal{S}$ is sampled with replacement from permutations of $\{1,\ldots,n\}$, and $|\mathcal{S}|$ denotes its cardinality. Recall that

$$W_n = \sup_{\boldsymbol{\gamma}\in\Gamma^*}|Q^{(2)}(\boldsymbol{\gamma}) - Q^{(2)}(\mathbf{0}) - E_{\boldsymbol{\epsilon}^{(2)}}\{Q^{(2)}(\boldsymbol{\gamma}) - Q^{(2)}(\mathbf{0})\}|,$$

we can similarly derive that

$$\begin{aligned}
& pr_{\boldsymbol{\epsilon}^{(2)}}(W_n \geq t) \\
\leq\ & E_{\boldsymbol{\epsilon}^{(2)}}\Big\{\exp\Big(\lambda\sup_{\boldsymbol{\gamma}\in\Gamma^*}\Big|(M_n|\mathcal{S}|)^{-1}\sum_{\pi\in\mathcal{S}}\sum_{i=1}^{M_n}[h_{\pi(i),\pi(M_n+i)}(\boldsymbol{\gamma}) - h_{\pi(i),\pi(M_n+i)}(\mathbf{0}) \\
& \qquad -E_{\boldsymbol{\epsilon}^{(2)}}\{h_{\pi(i),\pi(M_n+i)}(\boldsymbol{\gamma}) - h_{\pi(i),\pi(M_n+i)}(\mathbf{0})\}]\Big|\Big)\Big\}\Big/\exp(\lambda t) \\
=\ & E_{\boldsymbol{\epsilon}^{(2)}}\Big\{\exp\Big(\lambda\sup_{\boldsymbol{\gamma}\in\Gamma^*}\Big|\frac{1}{|\mathcal{S}|}\sum_{\pi\in\mathcal{S}}M_n^{-1}\sum_{i=1}^{M_n}\big[h_{\pi(i),\pi(M_n+i)}(\boldsymbol{\gamma}) - h_{\pi(i),\pi(M_n+i)}(\mathbf{0}) \\
& \qquad -E_{\boldsymbol{\epsilon}^{(2)}}\{h_{\pi(i),\pi(M_n+i)}(\boldsymbol{\gamma}) - h_{\pi(i),\pi(M_n+i)}(\mathbf{0})\}\big]\Big|\Big)\Big\}\Big/\exp(\lambda t) \\
\leq\ & \frac{1}{|\mathcal{S}|}\sum_{\pi\in\mathcal{S}}E_{\boldsymbol{\epsilon}^{(2)}}\Big\{\exp\Big(\lambda\sup_{\boldsymbol{\gamma}\in\Gamma^*}\Big|M_n^{-1}\sum_{i=1}^{M_n}\big[h_{\pi(i),\pi(M_n+i)}(\boldsymbol{\gamma}) - h_{\pi(i),\pi(M_n+i)}(\mathbf{0}) \\
& \qquad -E_{\boldsymbol{\epsilon}^{(2)}}\{h_{\pi(i),\pi(M_n+i)}(\boldsymbol{\gamma}) - h_{\pi(i),\pi(M_n+i)}(\mathbf{0})\}\big]\Big|\Big)\Big\}\Big/\exp(\lambda t),
\end{aligned}$$

with $h_{i,j}(\boldsymbol{\gamma}) = |(e^{(2)}_i - e^{(2)}_j) - (\widehat{\mathbf{x}}^{(2)}_i - \widehat{\mathbf{x}}^{(2)}_j)^T\boldsymbol{\gamma}|$. Then it is also sufficient to deal with the summand for any given π, and further details are thus omitted here.

A.3 Proof of Lemma 4

Proof For consistency in notation, we next deal with $\widehat{\boldsymbol{\beta}}^{\mathrm{d},(2)}$. Now recall that

$$\mathbf{S}^{(2)}(\boldsymbol{\gamma}) = -\{n(n-1)\}^{-1}\sum_{i\neq j}(\widehat{\mathbf{x}}_i^{(2)} - \widehat{\mathbf{x}}_j^{(2)})\mathrm{sign}\big\{(e_i^{(2)} - e_j^{(2)}) - (\widehat{\mathbf{x}}_i^{(2)} - \widehat{\mathbf{x}}_j^{(2)})^{\mathrm{T}}\boldsymbol{\gamma}\big\}.$$

Given any $C' > 0$, we note that uniformly for $\boldsymbol{\gamma} \in \mathbb{R}^p$ such that $\|\boldsymbol{\gamma}\|_0 \leq C'q$, the number of (i,j) pairs such that $e_i^{(2)} - e_j^{(2)} = (\widehat{\mathbf{x}}_i^{(2)} - \widehat{\mathbf{x}}_j^{(2)})^{\mathrm{T}}\boldsymbol{\gamma}$ is at most of order $O(q^2)$ with probability one, because the distribution of $\epsilon_i^{(2)}$ has no point masses (see, for exmaple, Koenker, 2005, Section 2.2). Consequently, with probability one,

$$\begin{aligned}
&\sup_{\boldsymbol{\gamma}\in\mathbb{R}^p:\|\boldsymbol{\gamma}\|_0\leq C'q}\Big\|\{n(n-1)\}^{-1}\sum_{i\neq j}(\widehat{\mathbf{x}}_i^{(2)} - \widehat{\mathbf{x}}_j^{(2)})I\{e_i^{(2)} - e_j^{(2)} = (\widehat{\mathbf{x}}_i^{(2)} - \widehat{\mathbf{x}}_j^{(2)})^{\mathrm{T}}\boldsymbol{\gamma}\}\Big\|_\infty \\
= \;& O\{(q/n)^2 L_n\},
\end{aligned}$$

on $\mathscr{M}_3$ defined in the proof of Theorem 2. We shall see this quantity is negligible. Thus, it is sufficient to consider the term $2\{n(n-1)\}^{-1}\sum_{i\neq j}(\widehat{\mathbf{x}}_i^{(2)} - \widehat{\mathbf{x}}_j^{(2)})[I\{(e_i^{(2)} - e_j^{(2)}) \leq (\widehat{\mathbf{x}}_i^{(2)} - \widehat{\mathbf{x}}_j^{(2)})^{\mathrm{T}}\boldsymbol{\gamma}\} - 1/2]$, which we also denote by $\mathbf{S}^{(2)}(\boldsymbol{\gamma})$ for notational simplicity. Then the debiased estimator can be represented as

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}^{\mathrm{d},(2)} - \boldsymbol{\beta}_0 \;=\;& -\{n(n-1)\}^{-1}\sum_{i\neq j}\widehat{\boldsymbol{\Theta}}^{\mathrm{T}}(\widehat{\mathbf{x}}_i^{(2)} - \widehat{\mathbf{x}}_j^{(2)})\{I(\zeta_{i,j}^{(2)} < 0) - 1/2\}/\{2\widehat{f}^*(0)\} \\
&-\frac{f^*(0)}{\widehat{f}^*(0)}\widehat{\boldsymbol{\Theta}}^{\mathrm{T}}\mathbf{G}^{(2)}(\widehat{\boldsymbol{\gamma}}^{(2)}) - \frac{f^*(0)}{\widehat{f}^*(0)}\widehat{\boldsymbol{\Theta}}^{\mathrm{T}}E_{\boldsymbol{\epsilon}^{(2)}}\{\boldsymbol{\nu}^{(2)}(\boldsymbol{\gamma})\}\big|_{\boldsymbol{\gamma}=\widehat{\boldsymbol{\gamma}}^{(2)}} + \widehat{\boldsymbol{\gamma}}^{(2)} \\
=\;& -\widehat{\boldsymbol{\Theta}}^{\mathrm{T}}\mathbf{S}^{(2)}(\mathbf{0})/\{4\widehat{f}^*(0)\} - \frac{f^*(0)}{\widehat{f}^*(0)}\widehat{\boldsymbol{\Theta}}^{\mathrm{T}}\mathbf{G}^{(2)}(\widehat{\boldsymbol{\gamma}}^{(2)}) \\
&-\frac{f^*(0)}{\widehat{f}^*(0)}\widehat{\boldsymbol{\Theta}}^{\mathrm{T}}E_{\boldsymbol{\epsilon}^{(2)}}\{\boldsymbol{\nu}^{(2)}(\widehat{\boldsymbol{\gamma}}^{(2)})\} + \widehat{\boldsymbol{\gamma}}^{(2)},
\end{aligned}$$

where $\mathbf{G}^{(2)}(\boldsymbol{\gamma})$ denotes the empirical process

$$\mathbf{G}^{(2)}(\boldsymbol{\gamma}) = \boldsymbol{\nu}^{(2)}(\boldsymbol{\gamma}) - E_{\boldsymbol{\epsilon}^{(2)}}\{\boldsymbol{\nu}^{(2)}(\boldsymbol{\gamma})\},$$

and for $\boldsymbol{\gamma} \in \mathbb{R}^p$,

$$\boldsymbol{\nu}^{(2)}(\boldsymbol{\gamma}) = \{n(n-1)\}^{-1}\sum_{i\neq j}(\widehat{\mathbf{x}}_i^{(2)} - \widehat{\mathbf{x}}_j^{(2)})/\{2f^*(0)\}[I\{\zeta_{i,j}^{(2)} \leq (\widehat{\mathbf{x}}_i^{(2)} - \widehat{\mathbf{x}}_j^{(2)})^{\mathrm{T}}\boldsymbol{\gamma}\} - I(\zeta_{i,j}^{(2)} \leq 0)].$$

Denote

$$\widehat{P}_{i,j}(\boldsymbol{\gamma}) = I\{\zeta_{0,i,j}^{(2)} \leq (\widehat{\mathbf{x}}_i^{(2)} - \widehat{\mathbf{x}}_j^{(2)})^{\mathrm{T}}\boldsymbol{\gamma} + \widehat{\Delta}_g^{(1)}(Z_i^{(2)}) - \widehat{\Delta}_g^{(1)}(Z_j^{(2)})\},$$

and

$$P_{i,j}(\boldsymbol{\gamma}) = F^*\{(\widehat{\mathbf{x}}_i^{(2)} - \widehat{\mathbf{x}}_j^{(2)})^{\mathrm{T}}\boldsymbol{\gamma} + \widehat{\Delta}_g^{(1)}(Z_i^{(2)}) - \widehat{\Delta}_g^{(1)}(Z_j^{(2)})\},$$

where $\zeta_{0,i,j}^{(2)} = \epsilon_i^{(2)} - \epsilon_j^{(2)}$. Then

$$\boldsymbol{\nu}^{(2)}(\boldsymbol{\gamma}) = \{n(n-1)\}^{-1} \sum_{i \neq j} (\widehat{\mathbf{x}}_i^{(2)} - \widehat{\mathbf{x}}_j^{(2)})/(2f^*(0))\{\widehat{P}_{i,j}(\boldsymbol{\gamma}) - \widehat{P}_{i,j}(\mathbf{0})\}.$$

We let $\{\widetilde{\boldsymbol{\gamma}}_s\}_{s\in[N_{\boldsymbol{\gamma}}]}$ be centers of the balls of radius $\delta_n\xi_n$ that cover the set $\mathcal{C}(\delta_n, q') = \{\boldsymbol{\gamma} \in \mathbb{R}^p : \|\boldsymbol{\gamma}\|_2 \leq \delta_n, \|\boldsymbol{\gamma}\|_0 \leq q'\}$. Such a cover can be constructed with $N_{\boldsymbol{\gamma}} \leq C_p^{q'}(3/\xi_n)^{q'}$ (see, for exmaple, Wainwright, 2019, Example 5.8). Furthermore, let

$$\mathcal{B}(\widetilde{\boldsymbol{\gamma}}_s, r) = \{\boldsymbol{\gamma} \in \mathbb{R}^p : \|\boldsymbol{\gamma} - \widetilde{\boldsymbol{\gamma}}_s\|_2 \leq r \wedge \mathrm{supp}(\boldsymbol{\gamma}) \subseteq \mathrm{supp}(\widetilde{\boldsymbol{\gamma}}_s)\}$$

be a ball or radius r centered at $\widetilde{\boldsymbol{\gamma}}_s$ with points that have the same support as $\widetilde{\boldsymbol{\gamma}}_s$ for $s = 1, \ldots, N_{\boldsymbol{\gamma}}$. We define $G_k^{(2)}(\boldsymbol{\gamma})$ as the k-th component of $\mathbf{G}^{(2)}(\boldsymbol{\gamma})$, and in what follows we shall bound $\sup_{\boldsymbol{\gamma}\in\mathcal{C}(\delta_n,q')} \|\mathbf{G}^{(2)}(\boldsymbol{\gamma})\|_\infty = \max_k \sup_{\boldsymbol{\gamma}\in\mathcal{C}(\delta_n,q')} |G_k^{(2)}(\boldsymbol{\gamma})|$ using a standard discretization argument. In particular, using the notation introduced above, we have the following decomposition

$$\begin{aligned}
&\sup_{\boldsymbol{\gamma}\in\mathcal{C}(\delta_n,q')} |G_k^{(2)}(\boldsymbol{\gamma})| \\
= &\max_{s\in[N_{\boldsymbol{\gamma}}]} \sup_{\boldsymbol{\gamma}\in\mathcal{B}(\widetilde{\boldsymbol{\gamma}}_s,\delta_n\xi_n)} |G_k^{(2)}(\boldsymbol{\gamma})| \\
\leq &\max_{s\in[N_{\boldsymbol{\gamma}}]} |G_k^{(2)}(\widetilde{\boldsymbol{\gamma}}_s)| + \max_{s\in[N_{\boldsymbol{\gamma}}]} \sup_{\boldsymbol{\gamma}\in\mathcal{B}(\widetilde{\boldsymbol{\gamma}}_s,\delta_n\xi_n)} |G_k^{(2)}(\boldsymbol{\gamma}) - G_k^{(2)}(\widetilde{\boldsymbol{\gamma}}_s)| = T_{1,k} + T_{2,k}.
\end{aligned}$$

Notice that the term $T_{1,k}$ arises from discretization of the set $\mathcal{C}(\delta_n, q')$. To control it, we will apply the union bound for $s \in [N_{\boldsymbol{\gamma}}]$. The term $T_{2,k}$ captures the deviation of the empirical process in a small neighborhood around the fixed center $\widetilde{\boldsymbol{\gamma}}_s$. We first bound $T_{1,k}$. Let $a_{i,j,k} = (\widehat{X}_{i,k}^{(2)} - \widehat{X}_{j,k}^{(2)})/(2f^*(0))$, and

$$D_{i,j,k,s} = a_{i,j,k}[\{\widehat{P}_{i,j}(\widetilde{\boldsymbol{\gamma}}_s) - P_{i,j}(\widetilde{\boldsymbol{\gamma}}_s)\} - \{\widehat{P}_{i,j}(\mathbf{0}) - P_{i,j}(\mathbf{0})\}].$$

That is,

$$T_{1,k} = \max_{s\in[N_{\boldsymbol{\gamma}}]} \Big|\{n(n-1)\}^{-1} \sum_{i\neq j} D_{i,j,k,s}\Big|.$$

We note that $E_{\boldsymbol{\epsilon}^{(2)}}(D_{i,j,k,s}) = 0$ and for some $\eta_{i,j} \in [0,1]$,

$$\begin{aligned}
\mathrm{var}_{\boldsymbol{\epsilon}^{(2)}}(D_{i,j,k,s}) &= a_{i,j,k}^2\Big[P_{i,j}(\widetilde{\boldsymbol{\gamma}}_s) - P_{i,j}^2(\widetilde{\boldsymbol{\gamma}}_s) + P_{i,j}(\mathbf{0}) - P_{i,j}^2(\mathbf{0}) \\
&\qquad -2\{P_{i,j}(\mathbf{0}) \wedge P_{i,j}(\widetilde{\boldsymbol{\gamma}}_s)\} + 2P_{i,j}(\widetilde{\boldsymbol{\gamma}}_s)P_{i,j}(\mathbf{0})\Big] \\
&\leq a_{i,j,k}^2[P_{i,j}(\widetilde{\boldsymbol{\gamma}}_s) + P_{i,j}(\mathbf{0}) - 2\{P_{i,j}(\mathbf{0}) \wedge P_{i,j}(\widetilde{\boldsymbol{\gamma}}_s)\}] \\
&\leq a_{i,j,k}^2|(\widehat{\mathbf{x}}_i^{(2)} - \widehat{\mathbf{x}}_j^{(2)})^{\mathrm{T}}\widetilde{\boldsymbol{\gamma}}_s| f^*\Big\{\widehat{\Delta}_g^{(1)}(Z_i^{(2)}) - \widehat{\Delta}_g^{(1)}(Z_j^{(2)}) \\
&\qquad +\eta_{i,j}(\widehat{\mathbf{x}}_i^{(2)} - \widehat{\mathbf{x}}_j^{(2)})^{\mathrm{T}}\widetilde{\boldsymbol{\gamma}}_s\Big\} \\
&\leq \kappa_u a_{i,j,k}^2|(\widehat{\mathbf{x}}_i^{(2)} - \widehat{\mathbf{x}}_j^{(2)})^{\mathrm{T}}\widetilde{\boldsymbol{\gamma}}_s| := \sigma_{i,j,k,s}^2,
\end{aligned}$$

where the first inequality follows by dropping a negative term, the second inequality follows by the mean value theorem, and the last inequality is from Condition (C4). Furthermore, $b_{k,s} := \max_{i,j} |D_{i,j,k,s}| \leq 2 \max_{i,j,k} |a_{i,j,k}|$. Therefore, for a fixed s, we have that

$$\begin{aligned}
& E_{\boldsymbol{\epsilon}^{(2)}} \Big[\exp\{\lambda(D_{\pi(i),\pi(M_n+i),k,s})\} \Big] \\
= \;& 1 + \lambda^2 \sigma^2_{\pi(i),\pi(M_n+i),k,s}/2 + \sum_{t=3}^{\infty} \lambda^t E_{\boldsymbol{\epsilon}^{(2)}} (D^t_{\pi(i),\pi(M_n+i),k,s})/t! \\
\leq \;& 1 + \lambda^2 \sigma^2_{\pi(i),\pi(M_n+i),k,s}/2 + \lambda^2 \sigma^2_{\pi(i),\pi(M_n+i),k,s}/2 \sum_{t=3}^{\infty} (|\lambda| b_{k,s})^{t-2} \\
= \;& 1 + \frac{\lambda^2 \sigma^2_{\pi(i),\pi(M_n+i),k,s}/2}{1 - b_{k,s}|\lambda|} \leq \exp\Big(\frac{\lambda^2 \sigma^2_{\pi(i),\pi(M_n+i),k,s}/2}{1 - b_{k,s}|\lambda|} \Big),
\end{aligned}$$

for $|\lambda| \leq 1/b_{k,s}$. Let $C_0 > 0$ be a constant that may vary from line to line in the proof. On $\mathscr{M}_1 \cap \mathscr{M}_3 \cap \mathscr{L} \cap \mathscr{N}$ and under Condition (C1), we have that

$$b_{k,s} \leq 2 \max_{i,j,k} |a_{i,j,k}| \leq (2/f^*(0))(2C\sigma L_n + r_n) \leq C_0 L_n,$$

and by the Cauchy-Schwarz inequality,

$$M_n^{-1} \sum_{i=1}^{M_n} \sigma^2_{\pi(i),\pi(M_n+i),k,s} = M_n^{-1} \sum_{i=1}^{M_n} a^2_{\pi(i),\pi(M_n+i),k} |(\widehat{\mathbf{x}}^{(2)}_{\pi(i)} - \widehat{\mathbf{x}}^{(2)}_{\pi(M_n+i)})^{\mathrm{T}} \widetilde{\boldsymbol{\gamma}}_s| \leq C_0 \delta_n L_n.$$

Then for arbitrary $\varepsilon > 0$, following similar arguments to Theorem 2 and Bernstein's inequality (Wainwright, 2019, Proposition 2.10), we have

$$\Big| \{n(n-1)\}^{-1} \sum_{i \neq j} D_{i,j,k,s} \Big| \leq C_0 \big[\sqrt{\{\log(2/\varepsilon)/n\} \delta_n L_n} \vee (L_n/n) \log(2/\varepsilon) \big],$$

with probability $1 - \varepsilon$. Therefore, using the union bound over $s \in [N_{\boldsymbol{\gamma}}]$, with probability $1 - \varepsilon$, we have

$$T_{1,k} \leq C_0 \big[\{\log(2N_{\boldsymbol{\gamma}}/\varepsilon) \delta_n L_n/n\}^{1/2} \vee (L_n/n) \log(2N_{\boldsymbol{\gamma}}/\varepsilon) \big].$$

Let us now focus on bounding $T_{2,k}$ term. For a fixed s, we denote

$$T_{2,k,s} = \sup_{\boldsymbol{\gamma} \in \mathcal{B}(\widetilde{\boldsymbol{\gamma}}_s, \delta_n \xi_n)} |G^{(2)}_k(\boldsymbol{\gamma}) - G^{(2)}_k(\widetilde{\boldsymbol{\gamma}}_s)|.$$

We will use the fact that $I(\cdot < x)$ and $\mathrm{pr}(\cdot < x)$ are monotone function in x. Note that

$$\begin{aligned}
& \max_{s \in [N_{\boldsymbol{\gamma}}]} \max_{i,j} \sup_{\boldsymbol{\gamma} \in \mathcal{B}(\widetilde{\boldsymbol{\gamma}}_s, \delta_n \xi_n)} |(\widehat{\mathbf{x}}^{(2)}_i - \widehat{\mathbf{x}}^{(2)}_j)^{\mathrm{T}} (\boldsymbol{\gamma} - \widetilde{\boldsymbol{\gamma}}_s)| \\
\leq \;& \delta_n \xi_n \sqrt{q'} \max_{i,j} \|\widehat{\mathbf{x}}^{(2)}_i - \widehat{\mathbf{x}}^{(2)}_j\|_\infty \leq C_0 \delta_n \xi_n \sqrt{q' L_n^2} := \widetilde{L}_n,
\end{aligned}$$

on $\mathscr{M}_3 \cap \mathscr{N}$, because

$$\begin{aligned} & |(\widehat{\mathbf{x}}_i^{(2)} - \widehat{\mathbf{x}}_j^{(2)})^{\mathrm{T}}(\boldsymbol{\gamma} - \widetilde{\boldsymbol{\gamma}}_s)| \\ \leq \; & \|\boldsymbol{\gamma} - \widetilde{\boldsymbol{\gamma}}_s\|_2 \sqrt{\|\boldsymbol{\gamma} - \widetilde{\boldsymbol{\gamma}}_s\|_0} \max_{i,j} \|\widehat{\mathbf{x}}_i^{(2)} - \widehat{\mathbf{x}}_j^{(2)}\|_\infty. \end{aligned}$$

We further denote $E_{i,j}^s = (\widehat{\mathbf{x}}_i^{(2)} - \widehat{\mathbf{x}}_j^{(2)})^{\mathrm{T}}\widetilde{\boldsymbol{\gamma}}_s + \widehat{\Delta}_g^{(1)}(Z_i^{(2)}) - \widehat{\Delta}_g^{(1)}(Z_j^{(2)})$ and $E_{i,j}^0 = \widehat{\Delta}_g^{(1)}(Z_i^{(2)}) - \widehat{\Delta}_g^{(1)}(Z_j^{(2)})$, and then

$$\begin{aligned} T_{2,k,s} \;\leq\; & \{n(n-1)\}^{-1} \sum_{i \neq j} \Big[|a_{i,j,k}| \Big\{ I(\zeta_{0,i,j}^{(2)} \leq E_{i,j}^s + \widetilde{L}_n) - I(\zeta_{0,i,j}^{(2)} \leq E_{i,j}^0 - \widetilde{L}_n) \\ & - I(\zeta_{0,i,j}^{(2)} \leq E_{i,j}^s - \widetilde{L}_n) + I(\zeta_{0,i,j}^{(2)} \leq E_{i,j}^0 + \widetilde{L}_n) \Big\} \\ & + |a_{i,j,k}| \Big\{ - F^*(E_{i,j}^s - \widetilde{L}_n) + F^*(E_{i,j}^0 + \widetilde{L}_n) \\ & + F^*(E_{i,j}^s + \widetilde{L}_n) - F^*(E_{i,j}^0 - \widetilde{L}_n) \Big\} \Big]. \end{aligned}$$

Furthermore,

$$\begin{aligned} T_{2,k,s} \;\leq\; & \{n(n-1)\}^{-1} \sum_{i \neq j} \Big[|a_{i,j,k}| \Big\{ I(\zeta_{0,i,j}^{(2)} \leq E_{i,j}^s + \widetilde{L}_n) - I(\zeta_{0,i,j}^{(2)} \leq E_{i,j}^0 - \widetilde{L}_n) \\ & - I(\zeta_{0,i,j}^{(2)} \leq E_{i,j}^s - \widetilde{L}_n) + I(\zeta_{0,i,j}^{(2)} \leq E_{i,j}^0 + \widetilde{L}_n) \Big\} \\ & + |a_{i,j,k}| \Big\{ - F^*(E_{i,j}^s + \widetilde{L}_n) + F^*(E_{i,j}^0 - \widetilde{L}_n) \\ & + F^*(E_{i,j}^s - \widetilde{L}_n) - F^*(E_{i,j}^0 + \widetilde{L}_n) \Big\} \Big] \\ & + 2\{n(n-1)\}^{-1} \sum_{i \neq j} \Big[|a_{i,j,k}| \Big\{ F^*(E_{i,j}^s + \widetilde{L}_n) - F^*(E_{i,j}^0 - \widetilde{L}_n) \\ & - F^*(E_{i,j}^s - \widetilde{L}_n) + F^*(E_{i,j}^0 + \widetilde{L}_n) \Big\} \Big]. \end{aligned}$$

The first term in this equation can be bounded in a similar way to $T_{1,k}$ by applying Bernstein's inequality and hence the details are omitted. For the second term, we have a bound $C_0 L_n \widetilde{L}_n$, since $|a_{i,j,k}| \leq C_0 L_n$ and $F^*(x+\widetilde{L}_n) - F^*(x-\widetilde{L}_n) \leq 2\kappa_u \widetilde{L}_n$ for all $x \in \mathbb{R}$. Therefore, with probability $1-\varepsilon$,

$$T_{2,k,s} \leq C_0 [\{L_n \widetilde{L}_n \log(2/\varepsilon)/n\}^{1/2} \vee L_n \{\log(2/\varepsilon)/n\} \vee L_n \widetilde{L}_n].$$

A bound on $T_{2,k}$ now follows using a union bound over $s \in [N_{\boldsymbol{\gamma}}]$. We choose $\xi_n = n^{-1}$, which gives $N_{\boldsymbol{\gamma}} \leq (3np)^{q'}$. Then we let $q' = C'q$, and $\delta_n = \delta h_n$ with δ, h_n given in the proof of Theorem 2. Using a union bound over $k \in \{1, \ldots, p\}$, we have, on $\mathscr{M}_3 \cap \mathscr{L} \cap \mathscr{N}$, that

$$\max_k \sup_{\boldsymbol{\gamma} \in \mathcal{C}(\delta_n, q')} |G_k^{(2)}(\boldsymbol{\gamma})| \leq C_0 \Big[\sqrt{\{\log(np/\varepsilon)/n\} \delta_n q' L_n} \vee (L_n q'/n) \log(np/\varepsilon) \Big],$$

with probability $1 - 2\varepsilon$. That is,

$$\sup_{\boldsymbol{\gamma} \in \mathcal{C}(\delta_n, q')} \|\mathbf{G}^{(2)}(\boldsymbol{\gamma})\|_\infty = O_p[\lambda_{\min}^{-1/2}(\boldsymbol{\Sigma}) L_n^{3/2} \{\log(p)\}^{1/4} (q/n)^{3/4} \vee L_n^3 q/n].$$

Thus, we have

$$\sup_{\boldsymbol{\gamma}\in\mathcal{C}(\delta_n,q')}\|\widehat{\boldsymbol{\Theta}}^{\mathrm{T}}\mathbf{G}^{(2)}(\boldsymbol{\gamma})\|_\infty = O_p[\|\boldsymbol{\Theta}\|_1^{3/2}L_n^{3/2}\{\log(p)\}^{1/4}(q/n)^{3/4}],$$

since $\|\boldsymbol{\Theta}\|_1 L_n q\sqrt{\log p/n} = o(1)$ and $\|\widehat{\boldsymbol{\Theta}}-\boldsymbol{\Theta}\|_1 = o_p(1)$ by assumption, and since $1/(4\sigma^2) \le \lambda_{\min}^{-1}(\boldsymbol{\Sigma}) = \lambda_{\max}(\boldsymbol{\Theta}) = \|\boldsymbol{\Theta}\|_2 \le \|\boldsymbol{\Theta}\|_1$, where $\|\boldsymbol{\Theta}\|_1 = \max_k\|\boldsymbol{\theta}_k\|_1$ and $\|\boldsymbol{\Theta}\|_2 = \lambda_{\max}^{1/2}(\boldsymbol{\Theta}^{\mathrm{T}}\boldsymbol{\Theta})$ are the ℓ_1 norm and spectral norm of $\boldsymbol{\Theta}$, respectively.

Next, denote $\widetilde{\Gamma}^* = \{\boldsymbol{\gamma}\in\mathbb{R}^p : \boldsymbol{\gamma}\in\Gamma, \|\boldsymbol{\gamma}\|_2 \le \delta h_n\}$. Using the mean value theorem, we note that there exists some $\mu_{i,j}$ lying between $\widehat{\Delta}_g^{(1)}(Z_i^{(2)}) - \widehat{\Delta}_g^{(1)}(Z_j^{(2)})$ and $(\widehat{\mathbf{x}}_i^{(2)}-\widehat{\mathbf{x}}_j^{(2)})^{\mathrm{T}}\boldsymbol{\gamma} + \widehat{\Delta}_g^{(1)}(Z_i^{(2)}) - \widehat{\Delta}_g^{(1)}(Z_j^{(2)})$, such that

$$\begin{aligned}
E_{\boldsymbol{\epsilon}^{(2)}}\{\boldsymbol{\nu}^{(2)}(\boldsymbol{\gamma})\} &= \{n(n-1)\}^{-1}\sum_{i\neq j}\Big\{(\widehat{\mathbf{x}}_i^{(2)}-\widehat{\mathbf{x}}_j^{(2)})/(2f^*(0)) \\
&\qquad \times\Big[F^*\Big\{(\widehat{\mathbf{x}}_i^{(2)}-\widehat{\mathbf{x}}_j^{(2)})^{\mathrm{T}}\boldsymbol{\gamma} + \widehat{\Delta}_g^{(1)}(Z_i^{(2)}) - \widehat{\Delta}_g^{(1)}(Z_j^{(2)})\Big\} \\
&\qquad\qquad -F^*\{\widehat{\Delta}_g^{(1)}(Z_i^{(2)}) - \widehat{\Delta}_g^{(1)}(Z_j^{(2)})\}\Big]\Big\} \\
&= \{n(n-1)\}^{-1}\sum_{i\neq j}\Big[(\widehat{\mathbf{x}}_i^{(2)}-\widehat{\mathbf{x}}_j^{(2)})/(2f^*(0)) \\
&\qquad \times\{(\widehat{\mathbf{x}}_i^{(2)}-\widehat{\mathbf{x}}_j^{(2)})^{\mathrm{T}}\boldsymbol{\gamma}\}f^*\{\widehat{\Delta}_g^{(1)}(Z_i^{(2)}) - \widehat{\Delta}_g^{(1)}(Z_j^{(2)})\}\Big] \\
&\quad +\{n(n-1)\}^{-1}\sum_{i\neq j}\Big[(\widehat{\mathbf{x}}_i^{(2)}-\widehat{\mathbf{x}}_j^{(2)})/(2f^*(0)) \\
&\qquad \times[f^*(\mu_{i,j}) - f^*\{\widehat{\Delta}_g^{(1)}(Z_i^{(2)}) - \widehat{\Delta}_g^{(1)}(Z_j^{(2)})\}]\{(\widehat{\mathbf{x}}_i^{(2)}-\widehat{\mathbf{x}}_j^{(2)})^{\mathrm{T}}\boldsymbol{\gamma}\}\Big] \\
&:= \mathbf{T}_3(\boldsymbol{\gamma}) + \mathbf{T}_4(\boldsymbol{\gamma}).
\end{aligned}$$

For $\mathbf{T}_4(\boldsymbol{\gamma})$, by Condition (C1), we have known that $\|\widehat{\mathbf{x}}_i^{(2)}-\widehat{\mathbf{x}}_j^{(2)}\|_\infty \le C_0 L_n$ on $\mathscr{M}_3\cap\mathscr{N}$. Furthermore, we have

$$\begin{aligned}
&\sup_{\boldsymbol{\gamma}\in\widetilde{\Gamma}^*}\{n(n-1)\}^{-1}\sum_{i\neq j}[f^*(\mu_{i,j}) - f^*\{\widehat{\Delta}_g^{(1)}(Z_i^{(2)}) - \widehat{\Delta}_g^{(1)}(Z_j^{(2)})\}]^2 \\
\le\ & L_1^2\sup_{\boldsymbol{\gamma}\in\widetilde{\Gamma}^*}\{n(n-1)\}^{-1}\sum_{i\neq j}\{(\widehat{\mathbf{x}}_i^{(2)}-\widehat{\mathbf{x}}_j^{(2)})^{\mathrm{T}}\boldsymbol{\gamma}\}^2 \le C_0\{\lambda_{\min}^{-2}(\boldsymbol{\Sigma})q\log(p)/n\},
\end{aligned}$$

as well as

$$\sup_{\boldsymbol{\gamma}\in\widetilde{\Gamma}^*}\{n(n-1)\}^{-1}\sum_{i\neq j}\{(\widehat{\mathbf{x}}_i^{(2)}-\widehat{\mathbf{x}}_j^{(2)})^{\mathrm{T}}\boldsymbol{\gamma}\}^2 \le C_0\{\lambda_{\min}^{-2}(\boldsymbol{\Sigma})q\log(p)/n\},$$

on $\mathscr{L}\cap\mathscr{N}$, then

$$\sup_{\boldsymbol{\gamma}\in\widetilde{\Gamma}^*}\|\widehat{\boldsymbol{\Theta}}^{\mathrm{T}}\mathbf{T}_4(\boldsymbol{\gamma})\|_\infty = O_p\big[\{\|\boldsymbol{\Theta}\|_1 + o_p(1)\}\lambda_{\min}^{-2}(\boldsymbol{\Sigma})L_n q\log(p)/n\big] = O_p\big\{\|\boldsymbol{\Theta}\|_1^3 L_n q\log(p)/n\big\}.$$

For $\mathbf{T}_3(\boldsymbol{\gamma})$, we further decompose it into

$$\begin{aligned}\mathbf{T}_{3,1}(\boldsymbol{\gamma}) &= \{n(n-1)\}^{-1}\sum_{i\neq j}\Big\{(\widehat{\mathbf{x}}_i^{(2)}-\widehat{\mathbf{x}}_j^{(2)})/(2f^*(0))\{(\widehat{\mathbf{x}}_i^{(2)}-\widehat{\mathbf{x}}_j^{(2)})^{\mathrm{T}}\boldsymbol{\gamma}\}\\ &\qquad\times[f^*\{\widehat{\Delta}_g^{(1)}(Z_i^{(2)})-\widehat{\Delta}_g^{(1)}(Z_j^{(2)})\}-f^*(0)]\Big\},\\ \mathbf{T}_{3,2}(\boldsymbol{\gamma}) &= \{n(n-1)\}^{-1}\sum_{i\neq j}(\widehat{\mathbf{x}}_i^{(2)}-\widehat{\mathbf{x}}_j^{(2)})\{(\widehat{\mathbf{x}}_i^{(2)}-\widehat{\mathbf{x}}_j^{(2)})^{\mathrm{T}}\boldsymbol{\gamma}\}/2.\end{aligned}$$

Because on $\mathscr{M}_1\cap\mathscr{L}\cap\mathscr{N}$ and under Conditions (C1) and (C4),

$$\begin{aligned}&\max_k\{n(n-1)\}^{-1}\sum_{i\neq j}\Big\{(\widehat{X}_{i,k}^{(2)}-\widehat{X}_{j,k}^{(2)})^2\\ &\qquad\times[f^*\{\widehat{\Delta}_g^{(1)}(Z_i^{(2)})-\widehat{\Delta}_g^{(1)}(Z_j^{(2)})\}-f^*(0)]^2\Big\}\le C_0a_n^2,\end{aligned}$$

and

$$\sup_{\boldsymbol{\gamma}\in\widetilde{\Gamma}^*}\{n(n-1)\}^{-1}\sum_{i\neq j}\{(\widehat{\mathbf{x}}_i^{(2)}-\widehat{\mathbf{x}}_j^{(2)})^{\mathrm{T}}\boldsymbol{\gamma}\}^2\le C_0\{\lambda_{\min}^{-2}(\boldsymbol{\Sigma})q\log(p)/n\},$$

we have, by the Cauchy-Schwarz inequality, that

$$\sup_{\boldsymbol{\gamma}\in\widetilde{\Gamma}^*}\|\widehat{\boldsymbol{\Theta}}^{\mathrm{T}}\mathbf{T}_{3,1}(\boldsymbol{\gamma})\|_\infty=O_p[a_n\lambda_{\min}^{-1}(\boldsymbol{\Sigma})\|\boldsymbol{\Theta}\|_1\{q\log(p)/n\}^{1/2}]=O_p[a_n\|\boldsymbol{\Theta}\|_1^2\{q\log(p)/n\}^{1/2}].$$

On the other hand, recall $\mathscr{L}(t)$ and $\mathscr{M}_5(t)$ defined in the proof of Theorem 2, we note let $t_1=16\sqrt{2}C\sigma^2\sqrt{\log p/n}(\sqrt{2}C\sqrt{\log p/n}\vee 1)$ and $t_2=4C\sigma^2\sqrt{\log p/n}(2\sqrt{2}C\sqrt{\log p/n}\vee 1)$, so that

$$\begin{aligned}&\sup_{\boldsymbol{\gamma}\in\widetilde{\Gamma}^*}\|\mathbf{T}_{3,2}(\boldsymbol{\gamma})-\boldsymbol{\Sigma}\boldsymbol{\gamma}\|_\infty\\ \le\ &\sup_{\boldsymbol{\gamma}\in\widetilde{\Gamma}^*}\Big\|\Big[\{n(n-1)\}^{-1}\sum_{i\neq j}(\widehat{\mathbf{x}}_i^{(2)}-\widehat{\mathbf{x}}_j^{(2)})(\widehat{\mathbf{x}}_i^{(2)}-\widehat{\mathbf{x}}_j^{(2)})^{\mathrm{T}}-2\boldsymbol{\Sigma}\Big]\boldsymbol{\gamma}\Big\|_\infty\Big/2\\ \le\ &C_0\lambda_{\min}^{-1}(\boldsymbol{\Sigma})q\sqrt{\log p/n}\Big\|\{n(n-1)\}^{-1}\\ &\qquad\sum_{i\neq j}(\widehat{\mathbf{x}}_i^{(2)}-\widehat{\mathbf{x}}_j^{(2)})(\widehat{\mathbf{x}}_i^{(2)}-\widehat{\mathbf{x}}_j^{(2)})^{\mathrm{T}}-2\boldsymbol{\Sigma}\Big\|_\infty\le C_0\lambda_{\min}^{-1}(\boldsymbol{\Sigma})\{q\log(p)/n\},\end{aligned}$$

on $\mathscr{L}(t_1)\cap\mathscr{M}_4\cap\mathscr{M}_5(t_2)\cap\mathscr{N}$. This implies that

$$\begin{aligned}\sup_{\boldsymbol{\gamma}\in\widetilde{\Gamma}^*}\|\widehat{\boldsymbol{\Theta}}^{\mathrm{T}}\mathbf{T}_{3,2}(\boldsymbol{\gamma})-\widehat{\boldsymbol{\Theta}}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{\gamma}\|_\infty &\le \|\widehat{\boldsymbol{\Theta}}\|_1\sup_{\boldsymbol{\gamma}\in\widetilde{\Gamma}^*}\|\mathbf{T}_{3,2}(\boldsymbol{\gamma})-\boldsymbol{\Sigma}\boldsymbol{\gamma}\|_\infty\\ &\le O_p[\|\boldsymbol{\Theta}\|_1\lambda_{\min}^{-1}(\boldsymbol{\Sigma})\{q\log(p)/n\}].\end{aligned}$$

Moreover,

$$\begin{aligned}\sup_{\boldsymbol{\gamma}\in\widetilde{\Gamma}^*}\|\widehat{\boldsymbol{\Theta}}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{\gamma}-\boldsymbol{\gamma}\|_\infty &\le \|\widehat{\boldsymbol{\Theta}}-\boldsymbol{\Theta}\|_1\|\boldsymbol{\Sigma}\|_\infty\sup_{\boldsymbol{\gamma}\in\widetilde{\Gamma}^*}\|\boldsymbol{\gamma}\|_1\\ &= O_p[\|\widehat{\boldsymbol{\Theta}}-\boldsymbol{\Theta}\|_1\lambda_{\min}^{-1}(\boldsymbol{\Sigma})q\{\log(p)/n\}^{1/2}],\end{aligned}$$

and

$$\sup_{\boldsymbol{\gamma}\in\widetilde{\Gamma}^*}\Big\|\frac{\widehat{f}^*(0)}{f^*(0)}\boldsymbol{\gamma}-\boldsymbol{\gamma}\Big\|_\infty=O_p\Big\{\Big|\frac{\widehat{f}^*(0)}{f^*(0)}-1\Big|\lambda_{\min}^{-1}(\boldsymbol{\Sigma})\sqrt{q\log p/n}\Big\}.$$

Thus,

$$\sup_{\boldsymbol{\gamma}\in\widetilde{\Gamma}^*}\|\widehat{\boldsymbol{\Theta}}^{\mathrm{T}}\mathbf{T}_{3,2}(\boldsymbol{\gamma})-\boldsymbol{\gamma}\|_\infty=O_p\Big\{\|\boldsymbol{\Theta}\|_1 q\sqrt{\log p/n}(\|\widehat{\boldsymbol{\Theta}}-\boldsymbol{\Theta}\|_1\vee\|\boldsymbol{\Theta}\|_1\sqrt{\log p/n})\Big\},$$

and

$$\begin{aligned}&\sup_{\boldsymbol{\gamma}\in\widetilde{\Gamma}^*}\Big\|\widehat{\boldsymbol{\Theta}}^{\mathrm{T}}E_{\boldsymbol{\epsilon}^{(2)}}\{\boldsymbol{\nu}^{(2)}(\boldsymbol{\gamma})\}-\frac{\widehat{f}^*(0)}{f^*(0)}\boldsymbol{\gamma}\Big\|_\infty\\ =\;& O_p\Big[\|\boldsymbol{\Theta}\|_1\sqrt{q\log p/n}\Big\{\Big|\frac{\widehat{f}^*(0)}{f^*(0)}-1\Big|\vee\sqrt{q}\|\widehat{\boldsymbol{\Theta}}-\boldsymbol{\Theta}\|_1\vee\|\boldsymbol{\Theta}\|_1^2L_n\sqrt{q\log p/n}\vee\|\boldsymbol{\Theta}\|_1 a_n\Big\}\Big].\end{aligned}$$

Next we derive the convergence rate of $\widehat{f}^*(0)=\big(\widehat{f}^{*,(1)}(0)+\widehat{f}^{*,(2)}(0)\big)/2$ to $f^*(0)$. Denote

$$\begin{aligned}T_{5,1}(\boldsymbol{\gamma}) \;=\;& \{n(n-1)\}^{-1}\sum_{i\neq j}\Big(K[\{e_i^{(2)}-e_j^{(2)}-(\widehat{\mathbf{x}}_i^{(2)}-\widehat{\mathbf{x}}_j^{(2)})^{\mathrm{T}}\boldsymbol{\gamma}\}/h]\\ &-K\{(e_i^{(2)}-e_j^{(2)})/h\}\Big)\Big/h,\end{aligned}$$

and

$$T_{5,2}=\{n(n-1)\}^{-1}\sum_{i\neq j}K\{(e_i^{(2)}-e_j^{(2)})/h\}/h.$$

We first upper bound $\sup_{\boldsymbol{\gamma}\in\widetilde{\Gamma}^*}|T_{5,1}(\boldsymbol{\gamma})|$. On the one hand, following an argument analogous to that used for W_n in the proof of Theorem 2, we have that

$$\sup_{\boldsymbol{\gamma}\in\widetilde{\Gamma}^*}|T_{5,1}(\boldsymbol{\gamma})-E_{\boldsymbol{\epsilon}^{(2)}}\{T_{5,1}(\boldsymbol{\gamma})\}|=O_p\{\lambda_{\min}^{-1}(\boldsymbol{\Sigma})q\log p/(nh^2)\},$$

under Condition (C5). On the other hand,

$$\begin{aligned}&\sup_{\boldsymbol{\gamma}\in\widetilde{\Gamma}^*}|E_{\boldsymbol{\epsilon}^{(2)}}\{T_{5,1}(\boldsymbol{\gamma})\}|\\ \leq\;& \sup_{\boldsymbol{\gamma}\in\widetilde{\Gamma}^*}\Big|\{n(n-1)\}^{-1}\sum_{i\neq j}h^{-1}\int_{-\infty}^{\infty}\Big(K[\{u-E_{i,j}(\boldsymbol{\gamma})\}/h]\\ &-K\{(u-E_{i,j}^0)/h\}\Big)f^*(u)\,\mathrm{d}u\Big|\\ =\;& \sup_{\boldsymbol{\gamma}\in\widetilde{\Gamma}^*}\Big|\{n(n-1)\}^{-1}\sum_{i\neq j}\int_{-\infty}^{\infty}K(u)[f^*\{hu+E_{i,j}(\boldsymbol{\gamma})\}-f^*(hu+E_{i,j}^0)]\,\mathrm{d}u\Big|\\ \leq\;& \sup_{\boldsymbol{\gamma}\in\widetilde{\Gamma}^*}L_1\{n(n-1)\}^{-1}\sum_{i\neq j}|(\widehat{\mathbf{x}}_i^{(2)}-\widehat{\mathbf{x}}_j^{(2)})^{\mathrm{T}}\boldsymbol{\gamma}|\\ =\;& O_p\big[\lambda_{\min}^{-1}(\boldsymbol{\Sigma})\{q\log(p)/n\}^{1/2}\big],\end{aligned}$$

on $\mathscr{L} \cap \mathscr{N}$ and under Condition (C1), where $E_{i,j}^0 = \widehat{\Delta}_g^{(1)}(Z_i^{(2)}) - \widehat{\Delta}_g^{(1)}(Z_j^{(2)})$ and $E_{i,j}(\boldsymbol{\gamma}) = (\widehat{\mathbf{x}}_i^{(2)} - \widehat{\mathbf{x}}_j^{(2)})^{\mathrm{T}}\boldsymbol{\gamma} + \widehat{\Delta}_g^{(1)}(Z_i^{(2)}) - \widehat{\Delta}_g^{(1)}(Z_j^{(2)})$. Thus, we have that

$$\sup_{\boldsymbol{\gamma}\in\widetilde{\Gamma}^*} |T_{5,1}(\boldsymbol{\gamma})| = O_p\big\{\lambda_{\min}^{-1}(\boldsymbol{\Sigma})\sqrt{q\log p/n}\big(\sqrt{q\log p/n}/h^2 \vee 1\big)\big\}.$$

Then we upper bound $|T_{5,2} - f^*(0)|$, which is further decomposed into $|T_{5,2} - E_{\boldsymbol{\epsilon}^{(2)}}(T_{5,2})| + |E_{\boldsymbol{\epsilon}^{(2)}}(T_{5,2}) - f^*(0)|$. By standard arguments for kernel density estimation, we have that

$$\begin{aligned}
& |E_{\boldsymbol{\epsilon}^{(2)}}(T_{5,2}) - f^*(0)| \\
\leq\ & |f^*(E_{i,j}^0) - f^*(0)| + \Big|\int_{-\infty}^{+\infty} f^*(E_{i,j}^0 + hu)K(u)\,\mathrm{d}u - f^*(E_{i,j}^0)\Big| \\
=\ & O_p(a_n) + \Big|\int_{-\infty}^{+\infty} hu\{(f^*)'(\mu_{i,j}(u)) - (f^*)'(E_{i,j}^0)\}K(u)\,\mathrm{d}u\Big| \\
\leq\ & O_p(a_n) + L_3h^2\int_{-\infty}^{+\infty} u^2K(u)\,\mathrm{d}u = O_p(a_n \vee h^2),
\end{aligned}$$

under Conditions (C1), (C4) and (C5), where $\mu_{i,j}(u)$ lies between $E_{i,j}^0$ and $E_{i,j}^0 + hu$. As for the first term, we note that

$$\begin{aligned}
\mathrm{var}_{\boldsymbol{\epsilon}^{(2)}}(T_{5,2}) &\leq \frac{1}{n!}\sum_{\pi} M_n^{-2}\sum_{i=1}^{M_n} E_{\boldsymbol{\epsilon}^{(2)}}[K^2\{(\epsilon_{\pi(i)}^{(2)} - \epsilon_{\pi(M_n+i)}^{(2)} - E_{i,j}^0)/h\}]/h^2 \\
&\leq M_n^{-1}\kappa_u\int_{-\infty}^{\infty} K^2(u)\,\mathrm{d}u/h = O\{1/(nh)\},
\end{aligned}$$

by Jensen's inequality, so that $|T_{5,2} - E_{\boldsymbol{\epsilon}^{(2)}}(T_{5,2})| = O_p\{1/(nh)^{1/2}\}$ by Chebyshev's inequality.

Therefore, we have

$$\begin{aligned}
|\widehat{f}^*(0) - f^*(0)| &= O_p\big[\big\{\|\boldsymbol{\Theta}\|_1\sqrt{q\log p/n}\big(\sqrt{q\log p/n}/h^2 \vee 1\big) \vee 1/(nh)^{1/2} \vee h^2\big\} \vee a_n\big] \\
&= O_p\Big(\|\boldsymbol{\Theta}\|_1\sqrt{q\log p/n} \vee n^{-2/5} \vee a_n\Big),
\end{aligned}$$

if we let $h \asymp \{\|\boldsymbol{\Theta}\|_1 q\log(p)/n\}^{1/4} \vee n^{-1/5}$, and

$$\Big|\frac{\widehat{f}^*(0)}{f^*(0)} - 1\Big| = O_p\Big(\|\boldsymbol{\Theta}\|_1\sqrt{q\log p/n} \vee n^{-2/5} \vee a_n\Big),$$

because $f^*(0) \geq \kappa_l$ under Condition (C4). Putting the pieces together, we have

$$\begin{aligned}
& \sup_{\boldsymbol{\gamma}\in\widetilde{\Gamma}^*} \Big\|\widehat{\boldsymbol{\Theta}}^{\mathrm{T}}E_{\boldsymbol{\epsilon}^{(2)}}\{\boldsymbol{\nu}^{(2)}(\boldsymbol{\gamma})\} - \frac{\widehat{f}^*(0)}{f^*(0)}\boldsymbol{\gamma}\Big\|_\infty \\
=\ & O_p\Big\{\|\boldsymbol{\Theta}\|_1\sqrt{q\log p/n}\Big(\|\boldsymbol{\Theta}\|_1\sqrt{q\log p/n} \vee n^{-2/5} \vee \sqrt{q}\|\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}\|_1 \\
& \qquad\qquad \vee \|\boldsymbol{\Theta}\|_1^2 L_n\sqrt{q\log p/n} \vee \|\boldsymbol{\Theta}\|_1 a_n\Big)\Big\} \\
=\ & O_p\Big\{\|\boldsymbol{\Theta}\|_1\sqrt{q\log p/n}\Big(n^{-2/5} \vee \sqrt{q}\|\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}\|_1 \vee \|\boldsymbol{\Theta}\|_1^2 L_n\sqrt{q\log p/n} \vee \|\boldsymbol{\Theta}\|_1 a_n\Big)\Big\},
\end{aligned}$$

and

$$\begin{aligned}
&\|\widehat{\boldsymbol{\beta}}^{\mathrm{d},(2)} - \boldsymbol{\beta}_0 + \widehat{\boldsymbol{\Theta}}^{\mathrm{T}}\mathbf{S}^{(2)}(\mathbf{0})/(4\widehat{f}^*(0))\|_\infty \\
= \;& O_p\Big\{\|\boldsymbol{\Theta}\|_1\sqrt{q\log p/n}\Big(n^{-2/5} \vee \sqrt{q}\|\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}\|_1 \vee \|\boldsymbol{\Theta}\|_1^2 L_n\sqrt{q\log p/n} \vee \|\boldsymbol{\Theta}\|_1 a_n\Big) \\
&\qquad \bigvee \|\boldsymbol{\Theta}\|_1^{3/2} L_n^{3/2}\{\log(p)\}^{1/4}(q/n)^{3/4}\Big\} \\
= \;& O_p\Big\{\|\boldsymbol{\Theta}\|_1\sqrt{q\log p/n}\Big(\sqrt{q}\|\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}\|_1 \vee \|\boldsymbol{\Theta}\|_1^2 L_n\sqrt{q\log p/n} \vee \|\boldsymbol{\Theta}\|_1 a_n\Big) \\
&\qquad \bigvee \|\boldsymbol{\Theta}\|_1^{3/2} L_n^{3/2}\{\log(p)\}^{1/4}(q/n)^{3/4}\Big\}.
\end{aligned}$$

This is because, by Theorem 2 and the assumption that $\|\widehat{\boldsymbol{\beta}}^{(2)}\|_0 = O_p(q)$, for any $\varepsilon > 0$, there exist sufficiently large constants $C > 0$ and $C' > 0$ (associated with δ and q') such that $\mathrm{pr}\{\widehat{\boldsymbol{\gamma}}^{(2)} \notin \mathcal{C}(\delta_n, q') \cap \widetilde{\Gamma}^*\} < \varepsilon$.

Because $\|\mathbf{S}^{(2)}(\mathbf{0})\|_\infty = O_p(\sqrt{\log p/n})$ by the proof of Lemma 1, and $|1/\widehat{f}^*(0)| = O_p(1)$, we have

$$\|(\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta})^{\mathrm{T}}\mathbf{S}^{(2)}(\mathbf{0})/(4\widehat{f}^*(0))\|_\infty = O_p(\|\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}\|_1\sqrt{\log p/n}).$$

Moreover, because

$$\Big|\frac{\widehat{f}^*(0)}{f^*(0)} - 1\Big| = O_p\Big(\|\boldsymbol{\Theta}\|_1\sqrt{q\log p/n} \vee n^{-2/5} \vee a_n\Big),$$

we have

$$\begin{aligned}
&\|\boldsymbol{\Theta}^{\mathrm{T}}\mathbf{S}^{(2)}(\mathbf{0})[\{1/(4\widehat{f}^*(0))\} - \{1/(4f^*(0))\}]\|_\infty \\
= \;& O_p\Big\{\Big(\|\boldsymbol{\Theta}\|_1\sqrt{q\log p/n} \vee n^{-2/5} \vee a_n\Big)\|\boldsymbol{\Theta}\|_1\sqrt{\log p/n}\Big\}.
\end{aligned}$$

Thus,

$$\begin{aligned}
&\|\widehat{\boldsymbol{\beta}}^{\mathrm{d},(2)} - \boldsymbol{\beta}_0 + \boldsymbol{\Theta}^{\mathrm{T}}\mathbf{S}^{(2)}(\mathbf{0})/(4f^*(0))\|_\infty \\
= \;& O_p\Big\{\|\boldsymbol{\Theta}\|_1\sqrt{q\log p/n}\Big(\sqrt{q}\|\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}\|_1 \vee \|\boldsymbol{\Theta}\|_1^2 L_n\sqrt{q\log p/n} \vee \|\boldsymbol{\Theta}\|_1 a_n\Big) \\
&\qquad \bigvee \|\boldsymbol{\Theta}\|_1^{3/2} L_n^{3/2}\{\log(p)\}^{1/4}(q/n)^{3/4}\Big\}.
\end{aligned}$$

Then we prove the second part of Lemma 4. Denote $\mathbf{T}_6 = \boldsymbol{\Theta}^{\mathrm{T}}(\mathbf{S}^{(2)}(\mathbf{0}) - \mathbf{S}_0^{(2)})$. We note that

$$\|\mathbf{T}_6\|_\infty \le \|\boldsymbol{\Theta}\|_1\|\mathbf{S}^{(2)}(\mathbf{0}) - \mathbf{S}_0^{(2)}\|_\infty = O_p\Big\{\|\boldsymbol{\Theta}\|_1\Big(r_n a_n \vee a_n^{1/2}\sqrt{\log p/n} \vee \log p/n\Big)\Big\},$$

by using the proof of Lemma 1. It thus remains to deal with $\boldsymbol{\Theta}^{\mathrm{T}}\mathbf{S}_0^{(2)}$, and we claim it is asymptotically equivalent to

$$\mathbf{T}_7 = -\{n(n-1)\}^{-1}\sum_{i\neq j}\boldsymbol{\Theta}^{\mathrm{T}}(\widetilde{\mathbf{x}}_i^{(2)} - \widetilde{\mathbf{x}}_j^{(2)})\mathrm{sign}(\epsilon_i^{(2)} - \epsilon_j^{(2)}),$$

in the sense that

$$\|\mathbf{\Theta}^{\mathrm{T}}\mathbf{S}_0^{(2)} - \mathbf{T}_7\|_\infty = O_p(\|\mathbf{\Theta}\|_1 r_n\sqrt{\log p/n}).$$

To see this, we note that

$$\begin{aligned}\|\mathbf{\Theta}^{\mathrm{T}}\mathbf{S}_0^{(2)} - \mathbf{T}_7\|_\infty &\leq \|\mathbf{\Theta}\|_1\Big\|\{n(n-1)\}^{-1}\sum_{i\neq j}\{\widehat{\mathbf{\Delta}}_{\mathbf{x}}^{(1)}(Z_i^{(2)}) - \widehat{\mathbf{\Delta}}_{\mathbf{x}}^{(1)}(Z_j^{(2)})\}\mathrm{sign}(\epsilon_i^{(2)} - \epsilon_j^{(2)})\Big\|_\infty \\ &= O_p(\|\mathbf{\Theta}\|_1 r_n\sqrt{\log p/n}),\end{aligned}$$

following similar arguments to the proof of Lemma 1 (ii). Finally, let's consider

$$\mathbf{T}_0 = -(4/n)\sum_{i=1}^n \mathbf{\Theta}^{\mathrm{T}}\widetilde{\mathbf{x}}_i^{(2)}\{F(\epsilon_i^{(2)}) - 1/2\}.$$

We first define a truncated version of $\widetilde{X}_{i,k}^{(2)}$, $\widetilde{X}_{i,k,T_n}^{(2)} = \widetilde{X}_{i,k}^{(2)} I(|\widetilde{X}_{i,k}^{(2)}| \leq T_n)$, for some $T_n > 0$. Then, we define in addition

$$\begin{aligned}\mathbf{T}_{7,T_n} &= -\{n(n-1)\}^{-1}\sum_{i\neq j}\mathbf{\Theta}^{\mathrm{T}}(\widetilde{\mathbf{x}}_{i,T_n}^{(2)} - \widetilde{\mathbf{x}}_{j,T_n}^{(2)})\mathrm{sign}(\epsilon_i^{(2)} - \epsilon_j^{(2)}), \\ \mathbf{T}_{0,T_n} &= -(4/n)\sum_{i=1}^n \mathbf{\Theta}^{\mathrm{T}}\{\widetilde{\mathbf{x}}_{i,T_n}^{(2)} - E(\widetilde{\mathbf{x}}_{i,T_n}^{(2)})\}\{F(\epsilon_i^{(2)}) - 1/2\},\end{aligned}$$

where $\widetilde{\mathbf{x}}_{i,T_n}^{(2)} = (\widetilde{X}_{i,1,T_n}^{(2)}, \ldots, \widetilde{X}_{i,p,T_n}^{(2)})$. Therefore,

$$\|\mathbf{T}_7 - \mathbf{T}_0\|_\infty \leq \|\mathbf{T}_7 - \mathbf{T}_{7,T_n}\|_\infty + \|\mathbf{T}_{7,T_n} - \mathbf{T}_{0,T_n}\|_\infty + \|\mathbf{T}_{0,T_n} - \mathbf{T}_0\|_\infty.$$

For $\|\mathbf{T}_{7,T_n} - \mathbf{T}_{0,T_n}\|_\infty$, we note that it is upper bounded by $\|\mathbf{\Theta}\|_1\|\mathbf{\Delta}_{T_n}\|_\infty$, where

$$\begin{aligned}\mathbf{\Delta}_{T_n} = &-\{n(n-1)\}^{-1} \\ &\sum_{i\neq j}\Big[(\widetilde{\mathbf{x}}_{i,T_n}^{(2)} - \widetilde{\mathbf{x}}_{j,T_n}^{(2)})\mathrm{sign}(\epsilon_i^{(2)} - \epsilon_j^{(2)}) - 2\{\widetilde{\mathbf{x}}_{i,T_n}^{(2)} - E(\widetilde{\mathbf{x}}_{i,T_n}^{(2)})\}\{F(\epsilon_i^{(2)}) - 1/2\} \\ &\qquad -2\{\widetilde{\mathbf{x}}_{j,T_n}^{(2)} - E(\widetilde{\mathbf{x}}_{j,T_n}^{(2)})\}\{F(\epsilon_j^{(2)}) - 1/2\}\Big].\end{aligned}$$

Thus, it suffices to upper bound $\mathbf{\Delta}_{T_n}$ componentwise. For the k-th component, Δ_{k,T_n}, we observe that it is a degenerate U-statistic. Therefore, according to Proposition 2.3 (d) in Arcones and Gine (1993),

$$\mathrm{pr}\{(n-1)|\Delta_{k,T_n}| \geq t\} \leq c_4\exp\{-c_5 t/(6T_n)\},$$

for some universal constants c_4, c_5. By the union bound, we have that

$$\mathrm{pr}\{(n-1)\|\mathbf{\Delta}_{T_n}\|_\infty \geq C_0 T_n\log p\} \leq c_4\exp\{-(c_5C_0/6 - 1)\log p\}.$$

By letting $T_n = 2C\sigma L_n$, $\|\mathbf{\Delta}_{T_n}\|_\infty = O_p(\|\mathbf{\Theta}\|_1 L_n\log p/n)$. Moreover, $\|\mathbf{T}_7 - \mathbf{T}_{7,T_n}\|_\infty = 0$, and

$$\|\mathbf{T}_{0,T_n} - \mathbf{T}_0\|_\infty = 4\Big\|\mathbf{\Theta}^{\mathrm{T}}E(\widetilde{\mathbf{x}}_i^{(2)} - \widetilde{\mathbf{x}}_{i,T_n}^{(2)})(1/n)\sum_{i=1}^n\{F(\epsilon_i^{(2)}) - 1/2\}\Big\|_\infty,$$

on $\mathscr{M}_3$. For $\|E(\widetilde{\mathbf{x}}_i^{(2)} - \widetilde{\mathbf{x}}_{i,T_n}^{(2)})\|_\infty$, we observe that

$$\begin{aligned} |E(\widetilde{X}_{i,k}^{(2)} - \widetilde{X}_{i,k,T_n}^{(2)})| &\leq |E\{\widetilde{X}_{i,k}^{(2)} I(|\widetilde{X}_{i,k}^{(2)}| \geq T_n)\}| \\ &\leq 2\sqrt{2}\sigma \exp\{-T_n^2/(4\sigma^2)\} = 2\sqrt{2}\sigma \exp(-C^2 L_n^2), \end{aligned}$$

by Cauchy-Schwarz inequality, $E\{(\widetilde{X}_{i,k}^{(2)})^2\} \leq 4\sigma^2$, and $\mathrm{pr}(|\widetilde{X}_{i,k}^{(2)}| \geq T_n) \leq 2\exp\{-T_n^2/(2\sigma^2)\}$. Therefore,

$$\|\mathbf{T}_{0,T_n} - \mathbf{T}_0\|_\infty = O_p\{\|\boldsymbol{\Theta}\|_1 n^{-1/2} \|E(\widetilde{\mathbf{x}}_i^{(2)} - \widetilde{\mathbf{x}}_{i,T_n}^{(2)})\|_\infty\} = o_p(\|\boldsymbol{\Theta}\|_1 n^{-1}),$$

and

$$\|\mathbf{T}_7 - \mathbf{T}_0\|_\infty = O_p\Big(\|\boldsymbol{\Theta}\|_1 L_n \log p/n\Big).$$

To sum up,

$$\begin{aligned} &\Big\|\boldsymbol{\Theta}^{\mathrm{T}}\mathbf{S}^{(2)}(\mathbf{0}) + (4/n)\sum_{i=1}^n \boldsymbol{\Theta}^{\mathrm{T}}\widetilde{\mathbf{x}}_i^{(2)}\{F(\epsilon_i^{(2)}) - 1/2\}\Big\|_\infty \\ = \;& O_p\Big\{\|\boldsymbol{\Theta}\|_1\Big(r_n a_n \vee a_n^{1/2}\sqrt{\log p/n} \vee L_n \log p/n\Big)\Big\}. \end{aligned}$$

Then the proof is completed. ■

A.4 Proof of Theorem 5

Proof Note that the assumptions $\|\boldsymbol{\Theta}\|_1^3 L_n^3 \{\log(p)\}^{1/2} q^{3/2} n^{-1/2} = o(1)$ and $\|\boldsymbol{\Theta}\|_1^2(\sqrt{q\log p} \vee \log p)a_n = o(1)$ imply that

$$\begin{aligned} \|\boldsymbol{\Theta}\|_1^3 L_n q \log(p) n^{-1/2} &= o(1), \\ \|\boldsymbol{\Theta}\|_1^{3/2} L_n^{3/2} \{\log(p)\}^{1/4} q^{3/4} n^{-1/4} &= o(1), \\ \|\boldsymbol{\Theta}\|_1 L_n \log(p) n^{-1/2} &= o(1), \\ \|\boldsymbol{\Theta}\|_1^2 \sqrt{q\log p} a_n &= o(1), \\ \|\boldsymbol{\Theta}\|_1 \sqrt{\log p} a_n^{1/2} &= o(1). \end{aligned}$$

Therefore, we have $\|\mathbf{R}_{n,1}^{(l)}\|_\infty = o_p(n^{-1/2})$ and $\|\mathbf{R}_{n,2}^{(l)}\|_\infty = o_p(n^{-1/2})$ by assumptions. Recall that $\widehat{\beta}_k^{\mathrm{d}} = \big(\widehat{\beta}_k^{\mathrm{d},(1)} + \widehat{\beta}_k^{\mathrm{d},(2)}\big)/2$. We then have

$$\sqrt{N}\Big[(\widehat{\beta}_k^{\mathrm{d}} - \beta_{0,k}) - N^{-1}\sum_{i=1}^N \boldsymbol{\theta}_k^{\mathrm{T}}\widetilde{\mathbf{x}}_i(F(\epsilon_i) - 1/2)/f^*(0)\Big] = o_p(1),$$

by Lemma 4, where $N = 2n$. For $\widehat{\boldsymbol{\theta}}_k^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}^{(l)}\widehat{\boldsymbol{\theta}}_k$ with $\widehat{\boldsymbol{\Sigma}}^{(l)} = n^{-1}\sum_{i=1}^n(\widehat{\mathbf{x}}_i^{(l)} - \overline{\widehat{\mathbf{x}}}^{(l)})(\widehat{\mathbf{x}}_i^{(l)} - \overline{\widehat{\mathbf{x}}}^{(l)})^{\mathrm{T}}$ and $\overline{\widehat{\mathbf{x}}}^{(l)} = n^{-1}\sum_{i=1}^n \widehat{\mathbf{x}}_i^{(l)}$, we note that

$$\|\widehat{\boldsymbol{\Theta}}^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}^{(l)}\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{\Theta}\|_\infty \leq \|\widehat{\boldsymbol{\Sigma}}^{(l)} - \boldsymbol{\Sigma}\|_\infty \|\widehat{\boldsymbol{\Theta}}\|_1^2 + \|\widehat{\boldsymbol{\Theta}}\boldsymbol{\Sigma}\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{\Theta}\|_\infty.$$

On the one hand, because $\widehat{\mathbf{\Sigma}}^{(l)} = n^{-2}\sum_{i,j=1}^{n}(\widehat{\mathbf{x}}_i^{(l)} - \widehat{\mathbf{x}}_j^{(l)})(\widehat{\mathbf{x}}_i^{(l)} - \widehat{\mathbf{x}}_j^{(l)})^{\mathrm{T}}/2 = n^{-2}\sum_{i\neq j}(\widehat{\mathbf{x}}_i^{(l)} - \widehat{\mathbf{x}}_j^{(l)})(\widehat{\mathbf{x}}_i^{(l)} - \widehat{\mathbf{x}}_j^{(l)})^{\mathrm{T}}/2$, it follows from the proof of Lemma 4 that $\|\widehat{\mathbf{\Sigma}}^{(l)} - \mathbf{\Sigma}\|_\infty = O_p\{(r_n \vee 1)\sqrt{\log p/n} + r_n^2\} = O_p(\sqrt{\log p/n})$. Besides, $\|\widehat{\mathbf{\Theta}}\|_1^2 = O(\|\mathbf{\Theta}\|_1^2 \vee \|\widehat{\mathbf{\Theta}} - \mathbf{\Theta}\|_1^2) = O_p(\|\mathbf{\Theta}\|_1^2)$ by assumption. Therefore,

$$\|\widehat{\mathbf{\Sigma}}^{(l)} - \mathbf{\Sigma}\|_\infty \|\widehat{\mathbf{\Theta}}\|_1^2 = O_p(\|\mathbf{\Theta}\|_1^2\sqrt{\log p/n}).$$

On the other hand, we have that

$$\begin{aligned}\|\widehat{\mathbf{\Theta}}^{\mathrm{T}}\mathbf{\Sigma}\widehat{\mathbf{\Theta}} - \mathbf{\Theta}^{\mathrm{T}}\mathbf{\Sigma}\mathbf{\Theta}\|_\infty &\leq \|(\widehat{\mathbf{\Theta}} - \mathbf{\Theta})^{\mathrm{T}}\mathbf{\Sigma}(\widehat{\mathbf{\Theta}} - \mathbf{\Theta})\|_\infty + 2\|\mathbf{\Theta}^{\mathrm{T}}\mathbf{\Sigma}(\widehat{\mathbf{\Theta}} - \mathbf{\Theta})\|_\infty \\ &= O(\|\widehat{\mathbf{\Theta}} - \mathbf{\Theta}\|_1^2) + O(\|\widehat{\mathbf{\Theta}} - \mathbf{\Theta}\|_1).\end{aligned}$$

Therefore,

$$\|\widehat{\mathbf{\Theta}}^{\mathrm{T}}\widehat{\mathbf{\Sigma}}^{(l)}\widehat{\mathbf{\Theta}} - \mathbf{\Theta}^{\mathrm{T}}\mathbf{\Sigma}\mathbf{\Theta}\|_\infty = O_p(\|\widehat{\mathbf{\Theta}} - \mathbf{\Theta}\|_1 \vee \|\mathbf{\Theta}\|_1^2\sqrt{\log p/n}).$$

Furthermore, by the proof of Lemma 4 and the assumptions, we have

$$\begin{aligned}|\widehat{\omega}_k^2 - \omega_k^2| &= O_p\Big\{\|\widehat{\mathbf{\Theta}} - \mathbf{\Theta}\|_1 \vee \|\mathbf{\Theta}\|_1^2\sqrt{\log p/n} \\ &\qquad \vee\|\mathbf{\Theta}\|_1\Big(\|\mathbf{\Theta}\|_1\sqrt{q\log p/n} \vee n^{-2/5} \vee a_n\Big)\Big\} \\ &= O_p\Big\{\|\widehat{\mathbf{\Theta}} - \mathbf{\Theta}\|_1 \vee \|\mathbf{\Theta}\|_1\Big(\|\mathbf{\Theta}\|_1\sqrt{q\log p/n} \vee n^{-2/5} \vee a_n\Big)\Big\} = o_p(1), \\ \Big|\frac{\widehat{\omega}_k}{\omega_k} - 1\Big| &= \left|\frac{\widehat{\omega}_k^2 - \omega_k^2}{\omega_k\left(\sqrt{\widehat{\omega}_k^2 - \omega_k^2 + \omega_k^2} + \omega_k\right)}\right| = O_p\left(\left|\frac{\widehat{\omega}_k^2 - \omega_k^2}{\omega_k^2}\right|\right) \\ &= O_p\Big(\|\widehat{\mathbf{\Theta}} - \mathbf{\Theta}\|_1 \vee \|\mathbf{\Theta}\|_1^2\sqrt{\log p/n} \vee \|\mathbf{\Theta}\|_1\sqrt{q\log p/n} \vee n^{-2/5} \vee a_n\Big) = o_p(1).\end{aligned}$$

Hence we have $\widehat{\omega}_k/\omega_k \xrightarrow{p} 1$, where $\omega_k^2 = \Theta_{k,k}/[12\{f^*(0)\}^2]$ and

$$\widehat{\omega}_k^2 = \widehat{\boldsymbol{\theta}}_k^{\mathrm{T}}\Big(\frac{\widehat{\mathbf{\Sigma}}^{(1)} + \widehat{\mathbf{\Sigma}}^{(2)}}{2}\Big)\widehat{\boldsymbol{\theta}}_k/[12\{\widehat{f}^*(0)\}^2].$$

By Lindeberg-Feller CLT for triangular arrays (Resnick, 2019, Exercise 9.9.1),

$$N^{-1/2}\sum_{i=1}^{N}\boldsymbol{\theta}_k^{\mathrm{T}}\widetilde{\mathbf{x}}_i\{F(\epsilon_i) - 1/2\}/\{\omega_k f^*(0)\} \xrightarrow{d} \mathcal{N}(0,1),$$

since $E[\boldsymbol{\theta}_k^{\mathrm{T}}\widetilde{\mathbf{x}}_i\{F(\epsilon_i) - 1/2\}/\{N^{1/2}\omega_k f^*(0)\}] = 0$,

$$\sum_{i=1}^{N}E[\boldsymbol{\theta}_k^{\mathrm{T}}\widetilde{\mathbf{x}}_i\{F(\epsilon_i) - 1/2\}/\{N^{1/2}\omega_k f^*(0)\}]^2 = 1,$$

and for any $\varepsilon > 0$,

$$\begin{aligned}
&\sum_{i=1}^{N} E\Big([\boldsymbol{\theta}_k^{\mathrm{T}} \widetilde{\mathbf{x}}_i \{F(\epsilon_i) - 1/2\} / \{N^{1/2} \omega_k f^*(0)\}]^2 \\
&\qquad\qquad \times I[|\boldsymbol{\theta}_k^{\mathrm{T}} \widetilde{\mathbf{x}}_i \{F(\epsilon_i) - 1/2\} / \{N^{1/2} \omega_k f^*(0)\}| > \varepsilon] \Big) \\
= \;& N E\Big([\boldsymbol{\theta}_k^{\mathrm{T}} \widetilde{\mathbf{x}}_1 \{F(\epsilon_1) - 1/2\} / \{N^{1/2} \omega_k f^*(0)\}]^2 \\
&\qquad\qquad \times I[|\boldsymbol{\theta}_k^{\mathrm{T}} \widetilde{\mathbf{x}}_1 \{F(\epsilon_1) - 1/2\} / \{N^{1/2} \omega_k f^*(0)\}| > \varepsilon] \Big) \\
\leq \;& E[|\boldsymbol{\theta}_k^{\mathrm{T}} \widetilde{\mathbf{x}}_1 \{F(\epsilon_1) - 1/2\} / \{\omega_k f^*(0)\}|^3] / (\varepsilon N^{1/2}) \to 0.
\end{aligned}$$

This is because $E|\boldsymbol{\theta}_k^{\mathrm{T}} / \|\boldsymbol{\theta}_k\|_1 \widetilde{\mathbf{x}}_1|^3 = O(1)$ and $\Theta_{k,k} \geq 1/E(\widetilde{X}_{i,k}^2) \geq 1/(4\sigma^2)$ by sub-Gaussianity of $\widetilde{\mathbf{x}}_1$ (see, for exmaple, Vershynin, 2018, Proposition 2.5.2), $f^*(0) \geq \kappa_l$ and $\|\boldsymbol{\Theta}\|_1^3 N^{-1/2} = o(1)$. It thus implies by Slutsky's theorem that

$$\sqrt{N}(\widehat{\beta}_k^{\mathrm{d}} - \beta_{0,k}) / \widehat{\omega}_k \xrightarrow{d} \mathcal{N}(0,1),$$

since $f^*(0)/\widehat{f}^*(0) \xrightarrow{p} 1$, and then the proof is finished. ■

A.5 Proof of Theorem 7

Proof The proof will follow similar arguments to Chernozhukov et al. (2013) and Zhang and Cheng (2017). To begin, we apply Chernozhukov et al. (2013, Corollary 2.1) to derive the following result under the assumption that $\|\boldsymbol{\Theta}\|_1^6 L_n^{14} / n \leq M n^{-\iota}$ for some $M, \iota > 0$:

$$\rho := \sup_{t \in \mathbb{R}} \Big| \mathrm{pr}\Big(\max_{k \in \mathcal{G}} \Big| \sum_{i=1}^{N} \xi_{i,k} \Big| / \sqrt{N} \leq t \Big) - \mathrm{pr}\Big(\max_{k \in \mathcal{G}} \Big| \sum_{i=1}^{N} \widetilde{\xi}_{i,k} \Big| / \sqrt{N} \leq t \Big) \Big| \leq C'' N^{-C'}, \tag{7}$$

for any $\mathcal{G} \subseteq \{1, \ldots, p\}$ and some constants $C', C'' > 0$, where $\xi_{i,k} = \boldsymbol{\theta}_k^{\mathrm{T}} \widetilde{\mathbf{x}}_i \{F(\epsilon_i) - 1/2\} / f^*(0)$ and $\widetilde{\boldsymbol{\xi}}_i = (\widetilde{\xi}_{i,1}, \ldots, \widetilde{\xi}_{i,p})^{\mathrm{T}}$, $i = 1, \ldots, N$ is a sequence of zero-mean independent Gaussian vectors with $E(\widetilde{\boldsymbol{\xi}}_i \widetilde{\boldsymbol{\xi}}_i^{\mathrm{T}}) = \boldsymbol{\Theta} / [12\{f^*(0)\}^2]$. Because

$$\begin{aligned}
\max_{k \in \mathcal{G}} \Big| \sum_{i=1}^{N} \xi_{i,k} \Big| \Big/ \sqrt{N} &= \max_{k \in \mathcal{G}} \max \Big(\sum_{i=1}^{N} \xi_{i,k}, -\sum_{i=1}^{N} \xi_{i,k} \Big) \Big/ \sqrt{N}, \\
\max_{k \in \mathcal{G}} \Big| \sum_{i=1}^{N} \widetilde{\xi}_{i,k} \Big| \Big/ \sqrt{N} &= \max_{k \in \mathcal{G}} \max \Big(\sum_{i=1}^{N} \widetilde{\xi}_{i,k}, -\sum_{i=1}^{N} \widetilde{\xi}_{i,k} \Big) \Big/ \sqrt{N},
\end{aligned}$$

it suffices to verify Condition (E.1) of Chernozhukov et al. (2013), that is, there exist some constants $C_1, C_2 > 0$ and a sequence of constants B_n such that

$$C_1 \leq E(\xi_{i,k}^2) \leq C_2, \quad \max_{t=1,2} E(|\xi_{i,k}|^{2+t} / B_n^t) + E\{\exp(|\xi_{i,k}| / B_n)\} \leq 4,$$

uniformly over $k \in \{1, \ldots, p\}$. We have that $E(\xi_{i,k}^2) = \Theta_{k,k} / [12\{f^*(0)\}^2]$ is bounded away from zero and infinity, uniformly over $k \in \{1, \ldots, p\}$, because $1/(4\sigma^2) \leq \min_{k \in \{1, \ldots, p\}} \Theta_{k,k} \leq$

$\max_{k\in\{1,\ldots,p\}}\Theta_{k,k}<+\infty$ and $\kappa_l\le f^*(0)\le\kappa_u$. By the independence between $\widetilde{\mathbf{x}}_i$ and ϵ_i, we have, for $B_n=B\|\boldsymbol{\Theta}\|_1^3$ with sufficiently large $B>0$, that

$$\begin{aligned}\max_{t=1,2}E(|\xi_{i,k}|^{2+t}/B_n^t) &= \max_{t=1,2}\frac{E|\boldsymbol{\theta}_k^{\mathrm{T}}\widetilde{\mathbf{x}}_i|^{2+t}E[|\{F(\epsilon_i)-1/2\}/f^*(0)|^{2+t}]}{B_n^t}\\ &\le \max_{t=1,2}(t+2)\Gamma(t/2+1)\Big(\frac{\|\boldsymbol{\Theta}\|_1\sigma}{\sqrt{2}\kappa_l}\Big)^{t+2}\Big/B_n^t<1,\end{aligned}$$

and

$$\begin{aligned}E\{\exp(|\xi_{i,k}|/B_n)\} &= 1+\sum_{t=1}^{\infty}\frac{E(|\xi_{i,k}|^t)}{B_n^t t!}=1+\sum_{t=1}^{\infty}\frac{E|\boldsymbol{\theta}_k^{\mathrm{T}}\widetilde{\mathbf{x}}_i|^tE[|\{F(\epsilon_i)-1/2\}/f^*(0)|^t]}{B_n^t t!}\\ &\le 1+2\sum_{t=1}^{\infty}\Big(\frac{\|\boldsymbol{\Theta}\|_1\sigma}{\sqrt{2}B_n\kappa_l}\Big)^t\le\frac{2}{1-\frac{\|\boldsymbol{\Theta}\|_1\sigma}{\sqrt{2}B_n\kappa_l}}<3,\end{aligned}$$

uniformly over $k\in\{1,\ldots,p\}$, where we use the fact $f^*(0)\ge\kappa_l$, $E|\boldsymbol{\theta}_k^{\mathrm{T}}/\|\boldsymbol{\theta}_k\|_1\widetilde{\mathbf{x}}_i|^t\le(2\sigma^2)^{t/2}t\Gamma(t/2)$ under Condition (C2) (see, for exmaple, Vershynin, 2018, Proposition 2.5.2) and $t\Gamma(t/2)\le 2t!$ for all $t\ge 1$. Therefore, for $B_n=B\|\boldsymbol{\Theta}\|_1^3$ with sufficiently large $B>0$, we have $\max_{t=1,2}E(|\xi_{i,k}|^{2+t}/B_n^t)+E\{\exp(|\xi_{i,k}|/B_n)\}\le 4$. Thus, (7) is proved, with constants $C',C''>0$ depending only on C_1, C_2, M, and ι.

Then, without loss of generality, we set $\mathcal{G}=\{1,\ldots,p\}$. Define

$$T_{\mathcal{G}}^*=\max_{k\in\mathcal{G}}\sqrt{N}|\widehat{\beta}_k^{\mathrm{d}}-\beta_{0,k}|,\quad T_{1,\mathcal{G}}=\max_{k\in\mathcal{G}}\Big|\sum_{i=1}^N\xi_{i,k}\Big|\Big/\sqrt{N},\quad T_{0,\mathcal{G}}=\max_{k\in\mathcal{G}}\Big|\sum_{i=1}^N\widetilde{\xi}_{i,k}\Big|\Big/\sqrt{N}.$$

By Lemma 4, we have

$$\begin{aligned}&\Big\|\sqrt{N}(\widehat{\boldsymbol{\beta}}^{\mathrm{d}}-\boldsymbol{\beta}_0)-\sum_{i=1}^N\boldsymbol{\xi}_i/\sqrt{N}\Big\|_\infty\\ =\;& O_p\Big[\|\boldsymbol{\Theta}\|_1\sqrt{q\log p}\Big(\sqrt{q}\|\widehat{\boldsymbol{\Theta}}-\boldsymbol{\Theta}\|_1\vee\|\boldsymbol{\Theta}\|_1^2L_n\sqrt{q\log p/n}\vee\|\boldsymbol{\Theta}\|_1a_n\Big)\\ &\bigvee\|\boldsymbol{\Theta}\|_1^{3/2}L_n^{3/2}\{\log(p)\}^{1/4}q^{3/4}n^{-1/4}\bigvee\|\boldsymbol{\Theta}\|_1\Big(\sqrt{n}r_na_n\vee a_n^{1/2}\sqrt{\log p}\Big)\Big],\end{aligned}$$

where $\boldsymbol{\xi}_i=(\xi_{i,1},\ldots,\xi_{i,p})^{\mathrm{T}}$.

Note that the assumptions $\|\boldsymbol{\Theta}\|_1^3L_n^3\{\log(p)\}^{3/2}q^{3/2}n^{-1/2}=o(1)$, $\|\boldsymbol{\Theta}\|_1^2a_n[\sqrt{q}\log p\vee\{\log(p)\}^2]=o(1)$ and the existence of $\zeta_n\to+\infty$ such that $\zeta_n[\|\boldsymbol{\Theta}\|_1q\log p\vee\{\log(p)\}^2]\|\widehat{\boldsymbol{\Theta}}-\boldsymbol{\Theta}\|_1=o_p(1)$, imply that

$$\begin{aligned}\|\boldsymbol{\Theta}\|_1^{3/2}L_n^{3/2}\{\log(p)\}^{3/4}q^{3/4}n^{-1/4} &= o(1),\\ \|\boldsymbol{\Theta}\|_1^3L_nq\{\log(p)\}^{3/2}n^{-1/2} &= o(1),\\ \|\boldsymbol{\Theta}\|_1\sqrt{a_n}\log p &= o(1),\\ \|\boldsymbol{\Theta}\|_1^2a_n\sqrt{q}\log p &= o(1),\\ \zeta_n\|\boldsymbol{\Theta}\|_1q\log p\|\widehat{\boldsymbol{\Theta}}-\boldsymbol{\Theta}\|_1 &= o_p(1).\end{aligned}$$

Then we can always choose $\zeta_1, \zeta_2 > 0$ such that $\zeta_1\sqrt{1 \vee \log(p/\zeta_1)} = o(1)$, $\zeta_2 = o(1)$ and

$$\mathrm{pr}(|T^*_{\mathcal{G}} - T_{1,\mathcal{G}}| > \zeta_1) < \zeta_2.$$

For example, we can choose ζ_1 such that $\zeta_1 = o(p^{-1})$ and

$$\begin{aligned}\zeta_1^2 &= \|\boldsymbol{\Theta}\|_1\sqrt{q}\Big(\|\boldsymbol{\Theta}\|_1^2 L_n\sqrt{q\log p/n} \vee \|\boldsymbol{\Theta}\|_1 a_n\Big) \bigvee (\zeta_n\sqrt{\log p})^{-2} \\ &\quad \bigvee \|\boldsymbol{\Theta}\|_1^{3/2} L_n^{3/2}\{\log(p)\}^{-1/4} q^{3/4} n^{-1/4} \bigvee \|\boldsymbol{\Theta}\|_1\Big(\sqrt{n/\log p} r_n a_n \vee a_n^{1/2}\Big),\end{aligned}$$

so that $\zeta_1\sqrt{1 \vee \log(p/\zeta_1)} = o(1)$ and $|T^*_{\mathcal{G}} - T_{1,\mathcal{G}}| \le \|\sqrt{N}(\widehat{\boldsymbol{\beta}}^{\mathrm{d}} - \boldsymbol{\beta}_0) - \sum_{i=1}^N \boldsymbol{\xi}_i/\sqrt{N}\|_\infty = o_p(\zeta_1)$.

Recall that

$$c_{\mathcal{G}}(\alpha) = \inf\Big(t \in \mathbb{R} : \mathrm{pr}\Big[W_{\mathcal{G}} \le t \;\Big|\; \{\mathbf{x}_i, Y_i, Z_i\}_{i=1}^N\Big] \ge 1 - \alpha\Big),$$

where

$$W_{\mathcal{G}} = \max_{k\in\mathcal{G}}\Big|\sum_{l=1}^2\sum_{i=1}^n \widehat{\boldsymbol{\theta}}_k^{\mathrm{T}}(\widehat{\mathbf{x}}_i^{(l)} - \overline{\widehat{\mathbf{x}}}^{(l)})u_i^{(l)}\Big| \Big/ [\sqrt{12N}\{\widehat{f}^*(0)\}],$$

$\overline{\widehat{\mathbf{x}}}^{(l)} = n^{-1}\sum_{i=1}^n \widehat{\mathbf{x}}_i^{(l)}$, and $u_1^{(1)}, \ldots, u_n^{(1)}, u_1^{(2)}, \ldots, u_n^{(2)}$ are i.i.d. $\mathcal{N}(0,1)$ random variables independent of $\{\mathbf{x}_i, Y_i, Z_i\}_{i=1}^N$. Let $c_{0,\mathcal{G}}(\alpha) = \inf\{t \in \mathbb{R} : \mathrm{pr}(T_{0,\mathcal{G}} \le t) \ge 1 - \alpha\}$. Because $C_1 \le E(\widetilde{\xi}_{i,k}^2) = E(\xi_{i,k}^2) \le C_2$, and

$$\begin{aligned}&\max_{k\in\mathcal{G}}\Big|\sum_{l=1}^2\sum_{i=1}^n \widehat{\boldsymbol{\theta}}_k^{\mathrm{T}}(\widehat{\mathbf{x}}_i^{(l)} - \overline{\widehat{\mathbf{x}}}^{(l)})u_i^{(l)}\Big| \\ = \;& \max_{k\in\mathcal{G}}\max\Big\{\sum_{l=1}^2\sum_{i=1}^n \widehat{\boldsymbol{\theta}}_k^{\mathrm{T}}(\widehat{\mathbf{x}}_i^{(l)} - \overline{\widehat{\mathbf{x}}}^{(l)})u_i^{(l)}, -\sum_{l=1}^2\sum_{i=1}^n \widehat{\boldsymbol{\theta}}_k^{\mathrm{T}}(\widehat{\mathbf{x}}_i^{(l)} - \overline{\widehat{\mathbf{x}}}^{(l)})u_i^{(l)}\Big\},\end{aligned}$$

we have, for some constant $C_3 > 0$ depending only on C_1 and C_2, that

$$\begin{aligned}&\sup_{t\in\mathbb{R}}\Big|\mathrm{pr}\Big[W_{\mathcal{G}} \le t \;\Big|\; \{\mathbf{x}_i, Y_i, Z_i\}_{i=1}^N\Big] - \mathrm{pr}(T_{0,\mathcal{G}} \le t)\Big| \\ \le \;& C_3\Delta^{1/3}\{1 \vee \log(p/\Delta)\}^{2/3},\end{aligned}$$

by Chernozhukov et al. (2013, Lemma 3.1) or Chernozhukov et al. (2015, Theorem 2), where

$$\Delta = \Big\|\widehat{\boldsymbol{\Theta}}^{\mathrm{T}}\Big(\frac{\widehat{\boldsymbol{\Sigma}}^{(1)} + \widehat{\boldsymbol{\Sigma}}^{(2)}}{2}\Big)\widehat{\boldsymbol{\Theta}}/[12\{\widehat{f}^*(0)\}^2] - \boldsymbol{\Theta}/[12\{f^*(0)\}^2]\Big\|_\infty.$$

By the proofs of Lemma 4 and Theorem 5, we have

$$\Delta = O_p\Big(\|\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}\|_1 \vee \|\boldsymbol{\Theta}\|_1^2\sqrt{\log p/n} \vee \|\boldsymbol{\Theta}\|_1\sqrt{q\log p/n} \vee n^{-2/5} \vee a_n\Big),$$

under the assumption that $\max_{k\in\{1,\ldots,p\}} \Theta_{k,k} < +\infty$.

Moreover, on the event $\{\Delta \leq \zeta_3\}$, we have

$$\sup_{t\in\mathbb{R}} \Big|\mathrm{pr}\Big[W_{\mathcal{G}} \leq t \;\Big|\; \{\mathbf{x}_i, Y_i, Z_i\}_{i=1}^N\Big] - \mathrm{pr}(T_{0,\mathcal{G}} \leq t)\Big| \leq \pi(\zeta_3),$$

where $\pi(\zeta_3) = C_3\zeta_3^{1/3}\{1 \vee \log(p/\zeta_3)\}^{2/3}$. Then, for any $\alpha \in (0,1)$, we have

$$\begin{aligned} & \mathrm{pr}\Big[W_{\mathcal{G}} \leq c_{0,\mathcal{G}}(\alpha - \pi(\zeta_3)) \;\Big|\; \{\mathbf{x}_i, Y_i, Z_i\}_{i=1}^N\Big] \\ \geq\;& \mathrm{pr}\Big\{T_{0,\mathcal{G}} \leq c_{0,\mathcal{G}}(\alpha - \pi(\zeta_3))\Big\} - \pi(\zeta_3) \\ =\;& 1 - \alpha + \pi(\zeta_3) - \pi(\zeta_3) = 1 - \alpha, \end{aligned}$$

because the distribution of $T_{0,\mathcal{G}}$ has no point masses, and hence

$$\mathrm{pr}\{c_{\mathcal{G}}(\alpha) \leq c_{0,\mathcal{G}}(\alpha - \pi(\zeta_3))\} \geq 1 - \mathrm{pr}(\Delta > \zeta_3),$$

by the definition of $c_{\mathcal{G}}(\alpha)$ and because the distribution of $W_{\mathcal{G}}$ has no point masses, either. Similarly, we can derive that

$$\mathrm{pr}\{c_{0,\mathcal{G}}(\alpha + \pi(\zeta_3)) \leq c_{\mathcal{G}}(\alpha)\} \geq 1 - \mathrm{pr}(\Delta > \zeta_3).$$

Now, we have that

$$\begin{aligned} & |\mathrm{pr}(T_{\mathcal{G}}^* > c_{\mathcal{G}}(\alpha)) - \mathrm{pr}(T_{1,\mathcal{G}} > c_{0,\mathcal{G}}(\alpha))| \\ \leq\;& \mathrm{pr}(\{T_{\mathcal{G}}^* > c_{\mathcal{G}}(\alpha)\} \ominus \{T_{1,\mathcal{G}} > c_{0,\mathcal{G}}(\alpha)\}, |T_{\mathcal{G}}^* - T_{1,\mathcal{G}}| \leq \zeta_1) + \zeta_2 \\ \leq\;& \mathrm{pr}\Big[\{T_{\mathcal{G}}^* > c_{\mathcal{G}}(\alpha)\} \ominus \{T_{1,\mathcal{G}} > c_{0,\mathcal{G}}(\alpha)\}, |T_{\mathcal{G}}^* - T_{1,\mathcal{G}}| \leq \zeta_1, \\ & \qquad\qquad c_{0,\mathcal{G}}(\alpha + \pi(\zeta_3)) \leq c_{\mathcal{G}}(\alpha) \leq c_{0,\mathcal{G}}(\alpha - \pi(\zeta_3))\Big] + \zeta_2 + 2\mathrm{pr}(\Delta > \zeta_3) \\ \leq\;& \mathrm{pr}\{c_{0,\mathcal{G}}(\alpha + \pi(\zeta_3)) - \zeta_1 \leq T_{1,\mathcal{G}} \leq c_{0,\mathcal{G}}(\alpha - \pi(\zeta_3)) + \zeta_1\} + \zeta_2 + 2\mathrm{pr}(\Delta > \zeta_3) \\ \leq\;& \mathrm{pr}\{c_{0,\mathcal{G}}(\alpha + \pi(\zeta_3)) - \zeta_1 \leq T_{0,\mathcal{G}} \leq c_{0,\mathcal{G}}(\alpha - \pi(\zeta_3)) + \zeta_1\} + \zeta_2 + 2\mathrm{pr}(\Delta > \zeta_3) + 2\rho \\ \leq\;& 2\pi(\zeta_3) + \zeta_2 + 2\mathrm{pr}(\Delta > \zeta_3) + 4C_4\zeta_1\sqrt{1 \vee \log(p/\zeta_1)} + 2\rho, \end{aligned}$$

for some constant $C_4 > 0$ depending only on C_1 and C_2, where $A \ominus B = (A \cap B^c) \cup (B \cap A^c)$, the first two inequalities hold because $\mathrm{pr}(|T_{\mathcal{G}}^* - T_{1,\mathcal{G}}| > \zeta_1) < \zeta_2$ and $\mathrm{pr}\{c_{0,\mathcal{G}}(\alpha + \pi(\zeta_3)) \leq c_{\mathcal{G}}(\alpha) \leq c_{0,\mathcal{G}}(\alpha - \pi(\zeta_3))\} \geq 1 - 2\mathrm{pr}(\Delta > \zeta_3)$, respectively, and the second-to-last inequality follows from the definition of ρ. The last inequality is because the anti-concentration inequality (see, for exmaple, Chernozhukov et al., 2013, Lemma 2.1 or Chernozhukov et al., 2015, Corollary 1), that is,

$$\begin{aligned} & \sup_{z\in\mathbb{R}} \mathrm{pr}\Big(\Big|\max_k \sum_{i=1}^N \widetilde{\xi}_{i,k}/\sqrt{N} - z\Big| \leq \zeta_1\Big) \\ =\;& \sup_{z\in\mathbb{R}} \Big|\mathrm{pr}\Big(\max_k \sum_{i=1}^N \widetilde{\xi}_{i,k}/\sqrt{N} \leq z + \zeta_1\Big) - \mathrm{pr}\Big(\max_k \sum_{i=1}^N \widetilde{\xi}_{i,k}/\sqrt{N} \leq z - \zeta_1\Big)\Big| \\ \leq\;& C_4\zeta_1\sqrt{1 \vee \log(p/\zeta_1)}. \end{aligned}$$

To obtain the last inequality, we also use the fact that the distribution of $T_{0,\mathcal{G}}$ has no point masses, and

$$\max_{k\in\mathcal{G}}\Big|\sum_{i=1}^{N}\widetilde{\xi}_{i,k}\Big|\Big/\sqrt{N}=\max_{k\in\mathcal{G}}\max\Big(\sum_{i=1}^{N}\widetilde{\xi}_{i,k},-\sum_{i=1}^{N}\widetilde{\xi}_{i,k}\Big)\Big/\sqrt{N}.$$

Finally, we note that

$$\begin{aligned}\sup_{\alpha\in(0,1)}|\mathrm{pr}(T^*_{\mathcal{G}}>c_{\mathcal{G}}(\alpha))-\alpha| &\le \sup_{\alpha\in(0,1)}|\mathrm{pr}(T^*_{\mathcal{G}}>c_{\mathcal{G}}(\alpha))-\mathrm{pr}(T_{1,\mathcal{G}}>c_{0,\mathcal{G}}(\alpha))|\\ &\quad+\sup_{\alpha\in(0,1)}|\mathrm{pr}(T_{1,\mathcal{G}}>c_{0,\mathcal{G}}(\alpha))-\mathrm{pr}(T_{0,\mathcal{G}}>c_{0,\mathcal{G}}(\alpha))|\\ &\le 2\pi(\zeta_3)+\zeta_2+2\mathrm{pr}(\Delta>\zeta_3)+C_4\zeta_1\sqrt{1\vee\log(p/\zeta_1)}+3\rho,\end{aligned}$$

because the distribution of $T_{0,\mathcal{G}}$ has no point masses and $\mathrm{pr}(T_{0,\mathcal{G}}>c_{0,\mathcal{G}}(\alpha))=\alpha$. By the previous discussion, $\zeta_1\sqrt{1\vee\log(p/\zeta_1)}=o(1)$, $\zeta_2=o(1)$ and $\rho\le C''N^{-C'}=o(1)$, we can conclude that

$$\sup_{\alpha\in(0,1)}|\mathrm{pr}(T^*_{\mathcal{G}}>c_{\mathcal{G}}(\alpha))-\alpha|=o(1),$$

if we choose ζ_3 such that

$$\begin{aligned}\zeta_3^2 &= \zeta_n^{-2}\{\log(p)\}^{-4}\vee a_n\{\log(p)\}^{-2}\vee\|\mathbf{\Theta}\|_1 q^{1/2}\{\log(p)\}^{-3/2}n^{-1/2}\\ &\quad\vee\{\log(p)\}^{-2}n^{-2/5}\vee\|\mathbf{\Theta}\|_1^2\{\log(p)\}^{-3/2}n^{-1/2}.\end{aligned}$$

This is because under the assumptions $\|\mathbf{\Theta}\|_1^3 qL_n^3\{\log(p)\}^{3/2}q^{1/2}n^{-1/2}=o(1)$, $\|\mathbf{\Theta}\|_1^2 a_n[\sqrt{q}\log p\vee\{\log(p)\}^2]=o(1)$, and the existence of $\zeta_n\to+\infty$ such that $\zeta_n[\|\mathbf{\Theta}\|_1 q\log p\vee\{\log(p)\}^2]\|\widehat{\mathbf{\Theta}}-\mathbf{\Theta}\|_1=o_p(1)$, we have

$$\begin{aligned}\|\mathbf{\Theta}\|_1^2\{\log(p)\}^{5/2}n^{-1/2} &= o(1),\\ \|\mathbf{\Theta}\|_1\{\log(p)\}^{5/2}q^{1/2}n^{-1/2} &= o(1),\\ \{\log(p)\}^2n^{-2/5} &= o(1),\\ a_n\{\log(p)\}^2 &= o(1),\\ \zeta_n\{\log(p)\}^2\|\widehat{\mathbf{\Theta}}-\mathbf{\Theta}\|_1 &= o_p(1),\end{aligned}$$

and then $\pi(\zeta_3)=o(1)$ and $\Delta=o_p(\zeta_3)$.

Next, we can follow a similar argument to prove

$$\sup_{\alpha\in(0,1)}|\mathrm{pr}(\overline{T}^*_{\mathcal{G}}>\overline{c}_{\mathcal{G}}(\alpha))-\alpha|=o(1),$$

where $\overline{T}^*_{\mathcal{G}}=\max_{k\in\mathcal{G}}\sqrt{N}|\widehat{\beta}^{\mathrm{d}}_k-\beta_{0,k}|/\widehat{\omega}_k$,

$$\widehat{\omega}_k^2=\widehat{\boldsymbol{\theta}}_k^{\mathrm{T}}\Big(\frac{\widehat{\mathbf{\Sigma}}^{(1)}+\widehat{\mathbf{\Sigma}}^{(2)}}{2}\Big)\widehat{\boldsymbol{\theta}}_k/[12\{\widehat{f}^*(0)\}^2],$$

and $\widehat{\mathbf{\Sigma}}^{(l)} = n^{-1}\sum_{i=1}^{n}(\widehat{\mathbf{x}}_i^{(l)} - \overline{\widehat{\mathbf{x}}}^{(l)})(\widehat{\mathbf{x}}_i^{(l)} - \overline{\widehat{\mathbf{x}}}^{(l)})^{\mathrm{T}}$. We define accordingly

$$\overline{T}_{1,\mathcal{G}} = \max_{k\in\mathcal{G}}\Big|\sum_{i=1}^{N}\xi_{i,k}/\omega_k\Big|\Big/\sqrt{N}, \quad \overline{T}_{0,\mathcal{G}} = \max_{k\in\mathcal{G}}\Big|\sum_{i=1}^{N}\widetilde{\xi}_{i,k}/\omega_k\Big|\Big/\sqrt{N},$$

with $\omega_k^2 = \Theta_{k,k}/[12\{f^*(0)\}^2]$, and $\overline{c}_{0,\mathcal{G}}(\alpha) = \inf\{t\in\mathbb{R} : \mathrm{pr}(\overline{T}_{0,\mathcal{G}} \le t) \ge 1-\alpha\}$. Then we can similarly derive: (i) Under the assumption that $\|\mathbf{\Theta}\|_1^6 L_n^{14}/n \le Mn^{-\iota}$ for some $M, \iota > 0$, we have

$$\overline{\rho} := \sup_{t\in\mathbb{R}}\Big|\mathrm{pr}\Big(\max_{k\in\mathcal{G}}\Big|\sum_{i=1}^{N}\xi_{i,k}/\omega_k\Big|/\sqrt{N} \le t\Big) - \mathrm{pr}\Big(\max_{k\in\mathcal{G}}\Big|\sum_{i=1}^{N}\widetilde{\xi}_{i,k}/\omega_k\Big|/\sqrt{N} \le t\Big)\Big| \le \overline{C}''N^{-\overline{C}'}.$$

where $\overline{C}', \overline{C}'' > 0$ depend only on $\overline{C}_1 = \overline{C}_2 = 1$, M, and ι. This is because $E\{(\xi_{i,k}/\omega_k)^2\} = 1$, and for $\overline{B}_n = \overline{B}\|\mathbf{\Theta}\|_1^3$ with sufficiently large $\overline{B} > 0$,

$$\begin{aligned}\max_{t=1,2} E(|\xi_{i,k}/\omega_k|^{2+t}/\overline{B}_n^t) &= \max_{t=1,2}\frac{E|\boldsymbol{\theta}_k^{\mathrm{T}}\widetilde{\mathbf{x}}_i|^{2+t}E[|\sqrt{12/\Theta_{k,k}}\{F(\epsilon_i)-1/2\}|^{2+t}]}{\overline{B}_n^t}\\ &\le \max_{t=1,2}(t+2)\Gamma(t/2+1)(2\sqrt{6}\|\mathbf{\Theta}\|_1\sigma^2)^{t+2}\Big/\overline{B}_n^t < 1,\end{aligned}$$

and

$$\begin{aligned}E\{\exp(|\xi_{i,k}/\omega_k|/\overline{B}_n)\} &= 1+\sum_{t=1}^{\infty}\frac{E(|\xi_{i,k}/\omega_k|^t)}{\overline{B}_n^t t!}\\ &= 1+\sum_{t=1}^{\infty}\frac{E|\boldsymbol{\theta}_k^{\mathrm{T}}\widetilde{\mathbf{x}}_i|^t E[|\sqrt{12/\Theta_{k,k}}\{F(\epsilon_i)-1/2\}|^t]}{\overline{B}_n^t t!}\\ &\le 1+2\sum_{t=1}^{\infty}\Big(\frac{2\sqrt{6}\|\mathbf{\Theta}\|_1\sigma^2}{\overline{B}_n}\Big)^t \le \frac{2}{1-\frac{2\sqrt{6}\|\mathbf{\Theta}\|_1\sigma^2}{\overline{B}_n}} < 3,\end{aligned}$$

uniformly over $k\in\{1,\ldots,p\}$.

(ii) Under the assumptions, there exist some $\overline{\zeta}_1, \overline{\zeta}_2 > 0$ such that $\overline{\zeta}_1\{1\vee\log(p/\overline{\zeta}_1)\}^{1/2} = o(1)$, $\overline{\zeta}_2 = o(1)$ and

$$\mathrm{pr}(|\overline{T}_{\mathcal{G}}^* - \overline{T}_{1,\mathcal{G}}| > \overline{\zeta}_1) < \overline{\zeta}_2.$$

This is because

$$\begin{aligned}&\max_{k\in\{1,\ldots,p\}}\Big|\sqrt{N}(\widehat{\beta}_k^{\mathrm{d}} - \beta_{0,k})/\widehat{\omega}_k - \sum_{i=1}^{N}\xi_{i,k}/(\sqrt{N}\widehat{\omega}_k)\Big|\\ =\ & O_p\Big[\|\mathbf{\Theta}\|_1\sqrt{q\log p}\Big(\sqrt{q}\|\widehat{\mathbf{\Theta}}-\mathbf{\Theta}\|_1 \vee \|\mathbf{\Theta}\|_1^2 L_n\sqrt{q\log p/n}\vee\|\mathbf{\Theta}\|_1 a_n\Big)\\ &\bigvee\|\mathbf{\Theta}\|_1^{3/2}L_n^{3/2}\{\log(p)\}^{1/4}q^{3/4}n^{-1/4}\bigvee\|\mathbf{\Theta}\|_1\Big(\sqrt{n}r_n a_n \vee a_n^{1/2}\sqrt{\log p}\Big)\Big],\end{aligned}$$

and

$$\begin{aligned}&\max_{k\in\{1,\ldots,p\}}\Big|\sum_{i=1}^{N}\xi_{i,k}/\sqrt{N}(1/\widehat{\omega}_k-1/\omega_k)\Big|\\ =\;& O_p\Big\{\|\boldsymbol{\Theta}\|_1\sqrt{\log p}\Big(\|\widehat{\boldsymbol{\Theta}}-\boldsymbol{\Theta}\|_1\vee\|\boldsymbol{\Theta}\|_1^2\sqrt{\log p/n}\vee\|\boldsymbol{\Theta}\|_1\sqrt{q\log p/n}\vee n^{-2/5}\vee a_n\Big)\Big\},\end{aligned}$$

by the proofs of Lemma 4 and Theorem 5, the sub-Gaussianity of $\boldsymbol{\theta}_k^{\mathrm{T}}/\|\boldsymbol{\theta}_k\|_1\widetilde{\mathbf{x}}_i\{F(\epsilon_i)-1/2\}/f^*(0)$, and the boundedness of $\max_{k\in\{1,\ldots,p\}}1/\omega_k$.

(iii)

$$\mathrm{pr}\{\overline{c}_{0,\mathcal{G}}(\alpha+\overline{\pi}(\overline{\zeta}_3))\le\overline{c}_{\mathcal{G}}(\alpha)\le\overline{c}_{0,\mathcal{G}}(\alpha-\overline{\pi}(\overline{\zeta}_3))\}\ge 1-2\mathrm{pr}(\overline{\Delta}>\overline{\zeta}_3),$$

where $\overline{\pi}(\overline{\zeta}_3)=\overline{C}_3\overline{\zeta}_3^{1/3}\{1\vee\log(p/\overline{\zeta}_3)\}^{2/3}$ for some $\overline{C}_3>0$ depending only on $\overline{C}_1=\overline{C}_2=1$,

$$\overline{\Delta}=\max_{k,k'}\left|\frac{\widehat{\boldsymbol{\theta}}_k^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{\theta}}_{k'}}{\sqrt{\widehat{\boldsymbol{\theta}}_k^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{\theta}}_k}\sqrt{\widehat{\boldsymbol{\theta}}_{k'}^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{\theta}}_{k'}}}-\frac{\Theta_{k,k'}}{\sqrt{\Theta_{k,k}}\sqrt{\Theta_{k',k'}}}\right|,$$

and $\widehat{\boldsymbol{\Sigma}}:=(\widehat{\boldsymbol{\Sigma}}^{(1)}+\widehat{\boldsymbol{\Sigma}}^{(2)})/2$. By the proof of Theorem 5, we have

$$\overline{\Delta}=O_p(\|\widehat{\boldsymbol{\Theta}}-\boldsymbol{\Theta}\|_1\vee\|\boldsymbol{\Theta}\|_1^2\sqrt{\log p/n}).$$

Under the assumptions, there exist some $\overline{\zeta}_3>0$ such that $\overline{\pi}(\overline{\zeta}_3)=o(1)$ and $\overline{\Delta}=o_p(\overline{\zeta}_3)$.

(iv)

$$\sup_{z\in\mathbb{R}}\mathrm{pr}\Big(\Big|\max_k\sum_{i=1}^{N}\widetilde{\xi}_{i,k}/\omega_k/\sqrt{N}-z\Big|\le\overline{\zeta}_1\Big)\le\overline{C}_4\overline{\zeta}_1\{1\vee\log(p/\overline{\zeta}_1)\}^{1/2},$$

for some $\overline{C}_4>0$ depending only on $\overline{C}_1=\overline{C}_2=1$.

Finally, following similar arguments to $\sup_{\alpha\in(0,1)}|\mathrm{pr}(\overline{T}_{\mathcal{G}}^*>\overline{c}_{\mathcal{G}}(\alpha))-\alpha|$, we can derive that

$$\begin{aligned}&\sup_{\alpha\in(0,1)}|\mathrm{pr}(\overline{T}_{\mathcal{G}}^*>\overline{c}_{\mathcal{G}}(\alpha))-\alpha|\\ \le\;& 2\overline{\pi}(\overline{\zeta}_3)+\overline{\zeta}_2+2\mathrm{pr}(\overline{\Delta}>\overline{\zeta}_3)+\overline{C}_4\overline{\zeta}_1\{1\vee\log(p/\overline{\zeta}_1)\}^{1/2}+3\overline{\rho}=o(1).\end{aligned}$$

Note that the condition $\max_{k\in\{1,\ldots,p\}}\Theta_{k,k}<+\infty$ is not required for this part of the proof. This completes the proof. ■

A.6 The Assumptions on Random Errors

In general, the assumptions on random errors in Condition (C4) is not restrictive and can be satisfied by a wide range of density functions. Below, we provide two examples for the density of $\epsilon_i-\epsilon_j$.

(i) Suppose ϵ_i follows a normal distribution with mean 0 and variance σ^2. Then, the density function of $\epsilon_i - \epsilon_j$ is $f^*(t) = (4\pi\sigma^2)^{-1/2}\exp\{-t^2/(4\sigma^2)\}$. We have $\inf_{t\in[-\delta_0,\delta_0]} f^*(t) = (4\pi\sigma^2)^{-1/2}\exp\{-\delta_0^2/(4\sigma^2)\} =: \kappa_l$ and $\sup_{t\in\mathbb{R}} f^*(t) \le (4\pi\sigma^2)^{-1/2} =: \kappa_u$. The first derivative of $f^*(t)$ satisfies $\sup_{t\in\mathbb{R}}|f^{*\prime}(t)| = (2\sqrt{2\pi\sigma^2})^{-1}\exp(-1/2)$. Hence, $|f^*(t_1) - f^*(t_2)| \le L_1|t_1 - t_2|$ holds with $L_1 = (2\sqrt{2\pi}\sigma^2)^{-1}\exp(-1/2)$.

(ii) Now, suppose ϵ_i follows a t distribution with ν degrees of freedom, location parameter μ, and scale parameter σ. The density function of ϵ_i is

$$f(t) := f(t;\mu,\sigma,\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sigma\sqrt{\pi\nu}}\left(1 + \frac{1}{\nu}\left(\frac{t-\mu}{\sigma}\right)^2\right)^{-(\nu+1)/2},$$

and the density function of $\epsilon_i - \epsilon_j$ is $f^*(t) = \int_{-\infty}^{\infty} f(x-t)f(x)\,\mathrm{d}x$. By the symmetry and unimodality of the t density around μ, for any $0 \le t \le \delta_0$,

$$\begin{aligned}
f^*(t) &= \int_{-\infty}^{\infty} f(x-t)f(x)\,dx \\
&\ge \int_{-\infty}^{\mu} f(x-t)f(x)\,\mathrm{d}x + \int_{\mu+t}^{\infty} f(x-t)f(x)\,\mathrm{d}x \\
&\ge \int_{-\infty}^{\mu} f^2(x-t)\,\mathrm{d}x + \int_{\mu+t}^{\infty} f^2(x)\,\mathrm{d}x \\
&\ge \int_{-\infty}^{\mu} f^2(x-\delta_0)\,\mathrm{d}x + \int_{\mu+\delta_0}^{\infty} f^2(x)\,\mathrm{d}x \\
&\ge 2\int_{-\infty}^{\mu-\delta_0} f^2(x)\,\mathrm{d}x.
\end{aligned}$$

Since $f^*(t)$ is symmetric about zero, it follows that

$$\begin{aligned}
\inf_{t\in[-\delta_0,\delta_0]} f^*(t) &\ge 2\int_{-\infty}^{\mu-\delta_0} f^2(x)\,dx \\
&= 2\frac{\Gamma^2\left(\frac{\nu+1}{2}\right)\Gamma\left(\nu+\frac{1}{2}\right)}{\Gamma^2\left(\frac{\nu}{2}\right)\Gamma(\nu+1)\sigma\sqrt{\pi\nu}}\int_{-\infty}^{\mu-\delta_0} f\left(x;\mu,\sqrt{\nu/(2\nu+1)}\sigma,2\nu+1\right)\,\mathrm{d}x \\
&=: \kappa_l > 0.
\end{aligned}$$

On the other hand, we have

$$\begin{aligned}
\sup_{t\in\mathbb{R}} f^*(t) &\le \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sigma\sqrt{\pi\nu}}\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x \\
&= \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sigma\sqrt{\pi\nu}} =: \kappa_u.
\end{aligned}$$

Moreover, the first derivative of $f^*(t)$ satisfies

$$\begin{aligned}
\sup_{t\in\mathbb{R}}|f^{*\prime}(t)| &\le \sup_{t\in\mathbb{R}}|f'(t)|\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x \\
&= \frac{\Gamma\left(\frac{\nu+1}{2}\right)(\nu+1)}{\Gamma\left(\frac{\nu}{2}\right)\sigma\sqrt{2\pi\nu(\nu+2)}}\left(1+\frac{1}{\nu+2}\right)^{-(\nu+3)/2} =: L_1.
\end{aligned}$$

Thus, $|f^*(t_1) - f^*(t_2)| \leq L_1|t_1 - t_2|$.

A.7 The Necessity of Data Splitting

Illustration of Remark 3. Consider the following toy example: Assume that Z_i is a continuous random variable, and both $\{\mathbf{x}_i - \mathbf{m}_{\mathbf{x}}(Z_i)\}^{\mathrm{T}}\boldsymbol{\beta}_0$ and ϵ_i are bounded. Define the nonparametric estimators as $\widehat{\mathbf{m}}_{\mathbf{x}}(z) = \mathbf{m}_{\mathbf{x}}(z)$, and

$$\begin{aligned}\widehat{m}_0(z) &= m_0(z) + r_N \sum_{i=1}^{N}\{Y_i - m_0(Z_i)\}I(Z_i = z) \\ &= m_0(z) + r_N \sum_{i=1}^{N}[\{\mathbf{x}_i - \mathbf{m}_{\mathbf{x}}(Z_i)\}^{\mathrm{T}}\boldsymbol{\beta}_0 + \epsilon_i]I(Z_i = z),\end{aligned}$$

so that $\widehat{m}_0(\cdot)$ converges uniformly to $m_0(\cdot)$ at rate r_N. Direct calculation yields

$$\mathbf{S}(\mathbf{0}) = -\{N(N-1)\}^{-1}\sum_{i\neq j}(\widetilde{\mathbf{x}}_i - \widetilde{\mathbf{x}}_j)\mathrm{sign}\big\{(1-r_N)(\epsilon_i - \epsilon_j) - r_N(\widetilde{\mathbf{x}}_i - \widetilde{\mathbf{x}}_j)^{\mathrm{T}}\boldsymbol{\beta}_0\big\}.$$

Consequently,

$$\begin{aligned}E\{\mathbf{S}(\mathbf{0})\} &= -2E\{(\widetilde{\mathbf{x}}_i - \widetilde{\mathbf{x}}_j)[F^*\{r_N/(1-r_N)(\widetilde{\mathbf{x}}_i - \widetilde{\mathbf{x}}_j)^{\mathrm{T}}\boldsymbol{\beta}_0\} - 1/2]\} \\ &\asymp -2r_N/(1-r_N)E\{(\widetilde{\mathbf{x}}_i - \widetilde{\mathbf{x}}_j)(\widetilde{\mathbf{x}}_i - \widetilde{\mathbf{x}}_j)^{\mathrm{T}}\boldsymbol{\beta}_0\} = -4r_N/(1-r_N)\boldsymbol{\Sigma}\boldsymbol{\beta}_0.\end{aligned}$$

Therefore, if $\sqrt{\log p/N} = o(r_N)$ and $\|\boldsymbol{\Sigma}\boldsymbol{\beta}_0\|_\infty$ is bounded away from zero, it follows that $\|\mathbf{S}(\mathbf{0})\|_\infty \asymp_p r_N$. Consequently, using arguments similar to those in the proof of Theorem 2, selecting $\lambda \asymp r_N$ yields an ℓ_2 convergence rate of $O_p(\sqrt{q}r_N)$ for the rank Lasso estimator. This demonstrates that the near-oracle rate $O_p(\sqrt{q\log p/N})$ cannot be guaranteed. In contrast, data-splitting removes the bias induced by overfitting and ensures $E\{\mathbf{S}^{(l)}(\mathbf{0})\} = \mathbf{0}$. Thus, the proposed DS rank Lasso achieves near-oracle rates under mild conditions on nonparametric estimators, even when they are overfitted.

Illustration of Remark 6. Consider the same toy example and the remainder term $\mathbf{R}_{n,2}^{(l)}$ introduced in Lemma 4. The full-sample version of the k-th component of the remainder term is

$$\begin{aligned}R_{N,2,k} = \boldsymbol{\theta}_k^{\mathrm{T}}\Big[&-\{4N(N-1)\}^{-1}\sum_{i\neq j}(\widetilde{\mathbf{x}}_i - \widetilde{\mathbf{x}}_j)\mathrm{sign}\{(1-r_N)(\epsilon_i - \epsilon_j) - r_N(\widetilde{\mathbf{x}}_i - \widetilde{\mathbf{x}}_j)^{\mathrm{T}}\boldsymbol{\beta}_0\} \\ &-N^{-1}\sum_{i=1}^{N}\widetilde{\mathbf{x}}_i\{F(\epsilon_i) - 1/2\}\Big]\Big/ f^*(0).\end{aligned}$$

Through direct calculations, we obtain

$$\begin{aligned}E(R_{N,2,k}) &= -E(\boldsymbol{\theta}_k^{\mathrm{T}}(\widetilde{\mathbf{x}}_i - \widetilde{\mathbf{x}}_j)[F^*\{r_N/(1-r_N)(\widetilde{\mathbf{x}}_i - \widetilde{\mathbf{x}}_j)^{\mathrm{T}}\boldsymbol{\beta}_0\} - 1/2])/\{2f^*(0)\} \\ &\asymp -r_N/(1-r_N)E\{\boldsymbol{\theta}_k^{\mathrm{T}}(\widetilde{\mathbf{x}}_i - \widetilde{\mathbf{x}}_j)(\widetilde{\mathbf{x}}_i - \widetilde{\mathbf{x}}_j)^{\mathrm{T}}\boldsymbol{\beta}_0\}/\{2f^*(0)\} \\ &= -r_N/(1-r_N)\beta_{0,k}/f^*(0).\end{aligned}$$

Error	Method	CP_1	CP_{500}	$ACP_{\mathcal{B}}$	$ACP_{\mathcal{B}^c}$	$ARP_{\mathcal{B}}$	RCP_1	$AIL_{\mathcal{B}}$	$AIL_{\mathcal{B}^c}$
$\mathcal{N}(0, 2^2)$	DSRLasso	0.955	0.955	0.942	0.949	1.000	0.955	0.483	0.476
	$\text{DSRLasso}^{\text{sub}}$	0.960	0.950	0.943	0.949	1.000	0.960	0.483	0.477
	FSORLasso	0.965	0.955	0.940	0.950	1.000	0.965	0.462	0.468
	$\text{FSORLasso}^{\text{sub}}$	0.970	0.955	0.943	0.950	1.000	0.970	0.462	0.468
MN	DSRLasso	0.945	0.950	0.937	0.948	1.000	0.945	0.329	0.326
	$\text{DSRLasso}^{\text{sub}}$	0.955	0.950	0.942	0.948	1.000	0.955	0.330	0.326
	FSORLasso	0.930	0.965	0.937	0.950	1.000	0.930	0.314	0.318
	$\text{FSORLasso}^{\text{sub}}$	0.935	0.965	0.940	0.950	1.000	0.935	0.314	0.318

Table 5: A Comparison of the Proposed Methods and Full-sample Oracle Methods. Empirical Coverage Probabilities, Power and Lengths of Confidence Intervals Based on 200 Simulations.

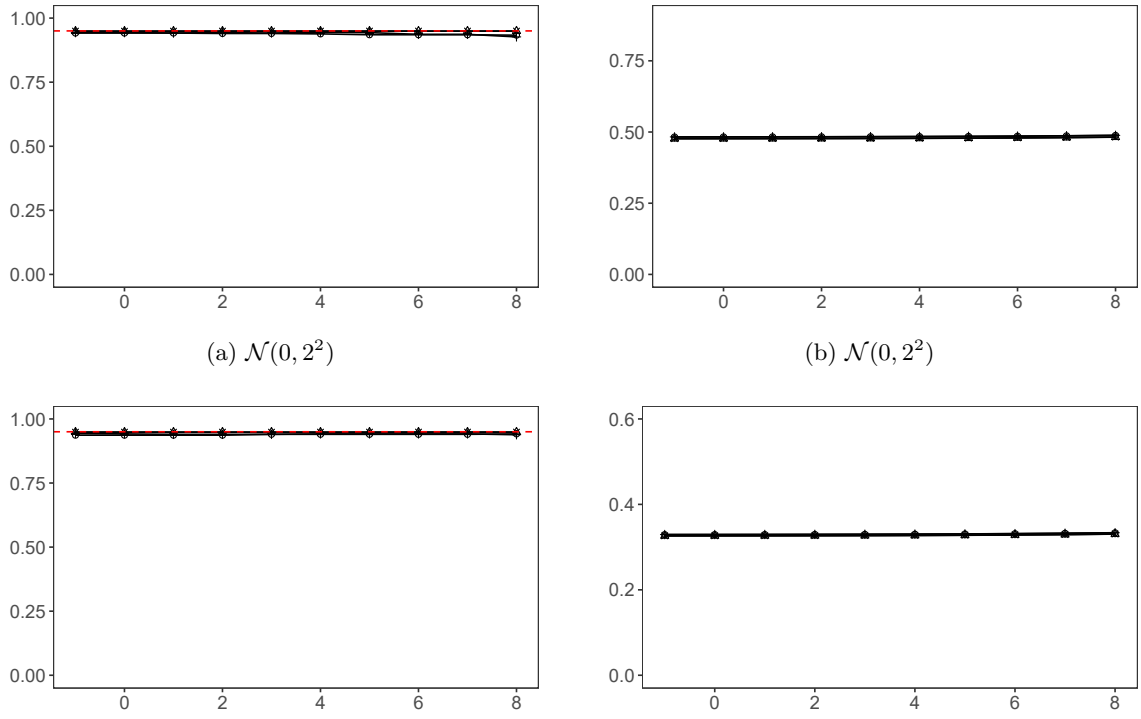
Given that the optimal rate of convergence for nonparametric regression is $r_N = N^{-1/2+\varepsilon}$ for some $\varepsilon > 0$ (Stone, 1982), it follows that $\sqrt{N}R_{N,2,k} \asymp_p N^{\varepsilon} \to \infty$ when $\beta_{0,k} \neq 0$. On the other hand, data splitting eliminates the bias induced by overfitting and ensures $E(\mathbf{R}_{n,2}^{(l)}) = \mathbf{0}$. Consequently, the remainder terms in $\sqrt{N}(\widehat{\beta}_k^{\text{d}} - \beta_{0,k})$ can converge to zero even when overfitting occurs in the nonparametric estimators.

Appendix B. Additional Numerical Results

In this section, we provide additional simulation results to support our proposed method and complement the simulations presented in the main article. The data generation process is the same as that in Example 2 of Section 4.1. Unless otherwise stated, we fix $N = 450$ and consider two error distributions: a light-tailed normal distribution $\epsilon_i \sim \mathcal{N}(0, 2^2)$ and a heavy-tailed mixture normal distribution $\epsilon_i \sim 0.95\mathcal{N}(0, 1) + 0.05\mathcal{N}(0, 10^2)$.

B.1 Comparison with Full-sample Oracle Methods

As established in Theorem 5 and the subsequent discussions in Section 3.1, the proposed debiased data-splitting (DS) rank Lasso estimator does not suffer from asymptotic efficiency loss due to data splitting. Specifically, it achieves the same asymptotic variance as the debiased full-sample (FS) oracle rank Lasso estimator introduced in Section 3.1. Table 5 compares the finite-sample performance of the inference procedure based on the proposed debiased DS rank Lasso estimator with that based on the debiased FS oracle rank Lasso estimator. The evaluation metrics are the same as those in Example 2, and the results are based on 200 simulations. We denote the debiased full-sample oracle rank Lasso estimator and its ANOPE (Fan et al., 2020b) subsampled version as FSORLasso and $\text{FSORLasso}^{\text{sub}}$, respectively. Overall, the finite-sample performance of the two methods is roughly comparable. In particular, the proposed debiased DS rank Lasso yields slightly wider but still comparable confidence intervals relative to the debiased FS oracle rank Lasso. This indicates that data splitting introduces only minor efficiency loss in finite samples when the sample size is moderately large. Such efficiency loss diminishes asymptotically, as supported by our theoretical analysis.



(a) $\mathcal{N}(0, 2^2)$ (b) $\mathcal{N}(0, 2^2)$

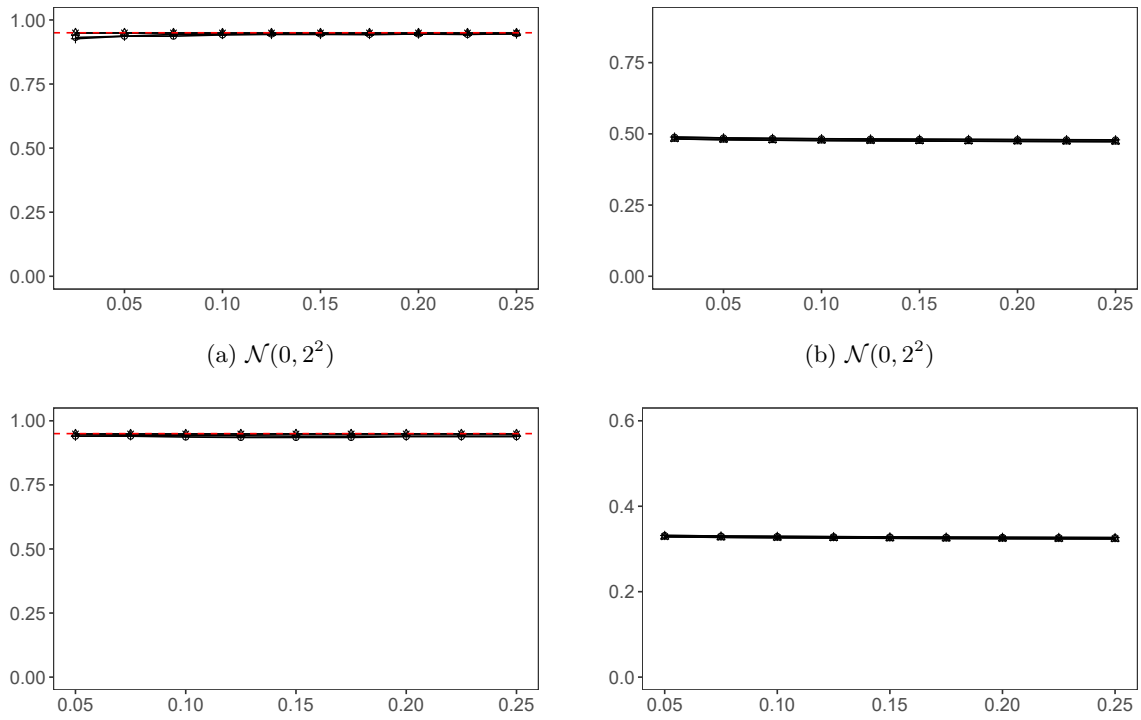
(c) MN (d) MN

Figure 2: Sensitivity Analysis for c with $\alpha_0 = 0.10$ and $c_0 = 0.01$ Fixed. Empirical coverage probabilities (a, c) and interval lengths (b, d) of confidence intervals are shown on the vertical axis against $i \in \{-1, 0, \ldots, 8\}$ (for $c = 1 + 10^{i/8-2}$) on the horizontal axis, based on 200 simulations. Results are displayed for DSRLasso ($\circ$ for $\mathcal{B}$, $\triangle$ for $\mathcal{B}^c$) and DSRLasso$^{\text{sub}}$ ($+$ for $\mathcal{B}$, $\times$ for $\mathcal{B}^c$). A red dashed line indicating the nominal 95% level is overlaid in (a) and (c).

B.2 Sensitivity Analysis for Constants Used in Tuning Parameter Selection

In this section, we conduct a sensitivity analysis of the constants c, α_0 and c_0 used in the selection of the tuning parameter given in (3). We evaluate the empirical coverage probabilities and lengths of confidence intervals based on 200 simulations, as shown in Figures 2–4. Specifically, in Figure 2, we fix $\alpha_0 = 0.10$ and $c_0 = 0.01$, and vary $c = 1 + 10^{i/8-2}$ for $i \in \{-1, 0, \ldots, 8\}$. In Figure 3, we fix $c = 1.01$ and $c_0 = 0.01$, and vary $\alpha_0 \in \{0.025, 0.050, \ldots, 0.250\}$. In Figure 4, we fix $\alpha_0 = 0.10$ and $c = 1.01$, and vary $c_0 = 20^{i/8-1}/5$ for $i \in \{-1, 0, \ldots, 8\}$.

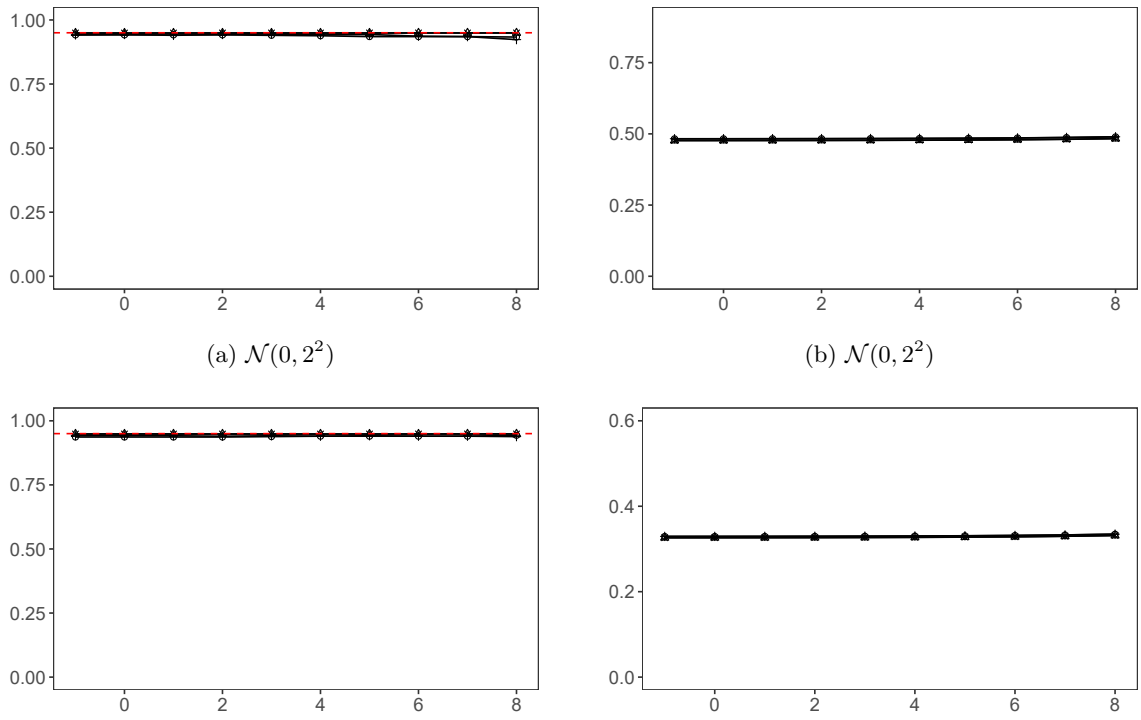
Overall, the performance of the proposed methods, in terms of both coverage probability and interval length, is quite robust to the choice of these constants. The results suggest that it is preferable to choose α_0 not too small, and c and c_0 not too large.



(a) $\mathcal{N}(0, 2^2)$ (b) $\mathcal{N}(0, 2^2)$

(c) MN (d) MN

Figure 3: Sensitivity Analysis for α_0 with $c = 1.01$ and $c_0 = 0.01$ Fixed. Empirical coverage probabilities (a, c) and interval lengths (b, d) of confidence intervals are shown on the vertical axis against $\alpha_0 \in \{0.025, 0.050, \ldots, 0.250\}$ on the horizontal axis, based on 200 simulations. Results are displayed for DSRLasso ($\circ$ for $\mathcal{B}$, $\triangle$ for $\mathcal{B}^c$) and DSRLasso$^{\text{sub}}$ ($+$ for $\mathcal{B}$, $\times$ for $\mathcal{B}^c$). A red dashed line indicating the nominal 95% level is overlaid in (a) and (c).



(a) $\mathcal{N}(0, 2^2)$ (b) $\mathcal{N}(0, 2^2)$

(c) MN (d) MN

Figure 4: Sensitivity Analysis for c_0 with $c = 1.01$ and $\alpha_0 = 0.10$ Fixed. Empirical coverage probabilities (a, c) and interval lengths (b, d) of confidence intervals are shown on the vertical axis against $i \in \{-1, 0, \ldots, 8\}$ (for $c_0 = 20^{i/8-1}/5$) on the horizontal axis, based on 200 simulations. Results are displayed for DSRLasso ($\circ$ for $\mathcal{B}$, $\triangle$ for $\mathcal{B}^c$) and DSRLasso$^{\text{sub}}$ ($+$ for $\mathcal{B}$, $\times$ for $\mathcal{B}^c$). A red dashed line indicating the nominal 95% level is overlaid in (a) and (c).

Error	$g_0(z)$	Method	CP_1	CP_{500}	$\text{ACP}_{\mathcal{B}}$	$\text{ACP}_{\mathcal{B}^c}$	$\text{ARP}_{\mathcal{B}}$	RCP_1	$\text{AIL}_{\mathcal{B}}$	$\text{AIL}_{\mathcal{B}^c}$
$\mathcal{N}(0,2^2)$	(i)	TSLasso	0.925	0.945	0.923	0.948	1.000	0.925	0.386	0.390
		DSRLasso	0.960	0.965	0.942	0.950	1.000	0.960	0.482	0.476
		$\text{DSRLasso}^{\text{sub}}$	0.955	0.965	0.938	0.950	1.000	0.955	0.483	0.477
	(ii)	TSLasso	0.920	0.935	0.925	0.949	1.000	0.920	0.385	0.389
		DSRLasso	0.940	0.940	0.938	0.949	1.000	0.940	0.483	0.477
		$\text{DSRLasso}^{\text{sub}}$	0.940	0.935	0.940	0.949	1.000	0.940	0.483	0.477
	(iii)	TSLasso	0.920	0.945	0.918	0.949	1.000	0.920	0.384	0.388
		DSRLasso	0.960	0.950	0.940	0.949	1.000	0.960	0.483	0.477
		$\text{DSRLasso}^{\text{sub}}$	0.965	0.960	0.942	0.949	1.000	0.965	0.483	0.477
MN	(i)	TSLasso	0.910	0.950	0.872	0.949	1.000	0.910	0.466	0.472
		DSRLasso	0.955	0.955	0.945	0.949	1.000	0.955	0.327	0.324
		$\text{DSRLasso}^{\text{sub}}$	0.955	0.950	0.943	0.949	1.000	0.955	0.327	0.324
	(ii)	TSLasso	0.905	0.955	0.880	0.949	1.000	0.905	0.464	0.470
		DSRLasso	0.965	0.950	0.948	0.949	1.000	0.965	0.326	0.323
		$\text{DSRLasso}^{\text{sub}}$	0.955	0.950	0.943	0.949	1.000	0.955	0.326	0.323
	(iii)	TSLasso	0.905	0.955	0.872	0.949	1.000	0.905	0.465	0.471
		DSRLasso	0.950	0.960	0.943	0.950	1.000	0.950	0.330	0.327
		$\text{DSRLasso}^{\text{sub}}$	0.950	0.945	0.942	0.950	1.000	0.950	0.330	0.327

Table 6: Simulation Results for Example 2 with Different Forms of Nuisance Functions $g_0(z)$. Empirical Coverage Probabilities, Power and Lengths of Confidence Intervals Based on 200 Simulations.

B.3 Different Forms of Nuisance Functions

In this section, we examine the sensitivity of the proposed inference methods to different forms of the nuisance function $g_0(z)$. Specifically, we consider the following additional forms:

(i) $g_0(z) = \frac{(z/2)^{10}(1-z/2)^4+(z/2)^4(1-z/2)^{10}}{B(11,5)}$, where $B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}\,\mathrm{d}x$ is the beta function;

(ii) $g_0(z) = \exp\{\sin(2\pi z)\} + (z-1)^2$; and

(iii) $g_0(z) = z\{1+\sin(2\pi z)\}$.

The results based on 200 simulations are summarized in Table 6. The evaluation metrics remain the same as those used in Example 2. Overall, the performance of the proposed inference methods is robust across the different functional forms of $g_0(z)$.

B.4 Different Nonparametric Estimators

To investigate the sensitivity of the proposed methods to the choice of nonparametric estimators, we first compare the simulation results for Example 2 using kernel-based and B-spline-based estimators for the nonparametric functions $m_0(\cdot)$ and $m_{\mathbf{x}}(\cdot)$, as summarized in Table 7. Here, DSRLasso(Kernel) and $\text{DSRLasso}^{\text{sub}}$(Kernel) denote the methods using kernel estimation, while DSRLasso(Spline) and $\text{DSRLasso}^{\text{sub}}$(Spline) denote those using

Error	Method	CP_1	CP_{500}	$ACP_{\mathcal{B}}$	$ACP_{\mathcal{B}^c}$	$ARP_{\mathcal{B}}$	RCP_1	$AIL_{\mathcal{B}}$	$AIL_{\mathcal{B}^c}$
$\mathcal{N}(0, 2^2)$	DSRLasso(Kernel)	0.955	0.955	0.942	0.949	1.000	0.955	0.483	0.476
	DSRLasso^{sub}(Kernel)	0.960	0.950	0.943	0.949	1.000	0.960	0.483	0.477
	DSRLasso(Spline)	0.945	0.955	0.938	0.952	1.000	0.945	0.531	0.533
	DSRLasso^{sub}(Spline)	0.950	0.955	0.940	0.952	1.000	0.950	0.532	0.534
MN	DSRLasso(Kernel)	0.945	0.950	0.937	0.948	1.000	0.945	0.329	0.326
	DSRLasso^{sub}(Kernel)	0.955	0.950	0.942	0.948	1.000	0.955	0.330	0.326
	DSRLasso(Spline)	0.960	0.930	0.943	0.954	1.000	0.960	0.386	0.387
	DSRLasso^{sub}(Spline)	0.960	0.935	0.943	0.954	1.000	0.960	0.387	0.387

Table 7: A Comparison of Kernel-Based and Spline-Based Estimators for Nonparametric Functions. Empirical Coverage Probabilities, Power and Lengths of Confidence Intervals Based on 200 Simulations.

B-splines. We note that, under certain regularity conditions (Newey, 1997), uniform convergence rates have been established for B-splines and, more generally, for series estimators. Hence, Condition (C1) can also be satisfied by such nonparametric estimators under appropriate regularity assumptions. The results indicate that the spline-based methods yield coverage probabilities comparable to those of the kernel-based methods, though with slightly wider interval lengths.

In addition, as discussed in Section 4.1, even when the moment condition for $\widetilde{Y}_i$ is not satisfied, such as when the error follows a $t(2)$ or Cauchy distribution, the performance of the proposed methods remains reasonable. This indicates that even mild violations of Condition (C1) do not substantially impair the methods, since this condition is only required in the first-step regression and not in the second-step rank regression. To further illustrate this point, we also compare the results from Example 2 using local linear mean regression (Fan and Gijbels, 1996; the method used in the main article) and local linear median regression (Koenker, 2005) for estimating $m_0(\cdot)$, as presented in Table 8. Here, DSRLasso(Mean) and DSRLasso^{sub}(Mean) refer to the methods using local mean regression, while DSRLasso(Median) and DSRLasso^{sub}(Median) denote those using local median regression. It is worth noting that, under certain regularity conditions (Guerre and Sabbah, 2012), uniform convergence rates for local median regression have been established. In our simulation settings, Condition (C1) holds even for $t(2)$ and Cauchy errors when local median regression is employed. We observe that the interval lengths for DSRLasso(Mean) and DSRLasso^{sub}(Mean) are slightly smaller under normal, mixture-normal, and $t(2)$ distributions, but larger under the Cauchy distribution. Nevertheless, the coverage probabilities across all methods remain roughly comparable.

In summary, although the interval lengths of the confidence intervals may vary depending on the choice of nonparametric estimator, the coverage probabilities of the proposed methods are quite robust to this choice. Moreover, while Condition (C1) is important for theoretical justification, the methods perform well even when it is mildly violated.

Error	Method	CP_1	CP_{500}	$ACP_{\mathcal{B}}$	$ACP_{\mathcal{B}^c}$	$ARP_{\mathcal{B}}$	RCP_1	$AIL_{\mathcal{B}}$	$AIL_{\mathcal{B}^c}$
$\mathcal{N}(0,2^2)$	DSRLasso(Mean)	0.955	0.955	0.942	0.949	1.000	0.955	0.483	0.476
	DSRLasso$^{\text{sub}}$(Mean)	0.960	0.950	0.943	0.949	1.000	0.960	0.483	0.477
	DSRLasso(Median)	0.950	0.950	0.938	0.948	1.000	0.950	0.518	0.522
	DSRLasso$^{\text{sub}}$(Median)	0.940	0.950	0.942	0.947	1.000	0.940	0.518	0.522
MN	DSRLasso(Mean)	0.945	0.950	0.937	0.948	1.000	0.945	0.329	0.326
	DSRLasso$^{\text{sub}}$(Mean)	0.955	0.950	0.942	0.948	1.000	0.955	0.330	0.326
	DSRLasso(Median)	0.950	0.955	0.942	0.951	1.000	0.950	0.367	0.363
	DSRLasso$^{\text{sub}}$(Median)	0.955	0.960	0.935	0.951	1.000	0.955	0.367	0.363
$t(2)$	DSRLasso(Mean)	0.950	0.960	0.940	0.949	1.000	0.950	0.555	0.560
	DSRLasso$^{\text{sub}}$(Mean)	0.965	0.960	0.943	0.949	1.000	0.965	0.556	0.562
	DSRLasso(Median)	0.950	0.955	0.942	0.951	1.000	0.950	0.561	0.566
	DSRLasso$^{\text{sub}}$(Median)	0.945	0.960	0.935	0.951	1.000	0.945	0.562	0.568
Cauchy	DSRLasso(Mean)	0.950	0.945	0.937	0.951	0.992	0.940	1.104	1.116
	DSRLasso$^{\text{sub}}$(Mean)	0.945	0.935	0.933	0.952	0.987	0.930	1.137	1.150
	DSRLasso(Median)	0.965	0.940	0.948	0.950	1.000	0.965	0.610	0.616
	DSRLasso$^{\text{sub}}$(Median)	0.965	0.940	0.948	0.950	1.000	0.965	0.611	0.617

Table 8: A Comparison of Local Mean Regression and Local Median Regression for Nonparametric Functions. Empirical Coverage Probabilities, Power and Lengths of Confidence Intervals Based on 200 Simulations.

B.5 Comparison with the Median-based Method

For comparison, we constructed a debiased data-splitting (DS) Lasso-penalized median estimator, following ideas similar to our proposed debiased DS rank Lasso estimator, as follows:

Using the partial prediction residuals $\{\widehat{Y}_i^{(1)}, \widehat{\mathbf{x}}_i^{(1)}\}_{i=1}^n$ and $\{\widehat{Y}_i^{(2)}, \widehat{\mathbf{x}}_i^{(2)}\}_{i=1}^n$ obtained in Section 2.1, we define

$$
\begin{aligned}
(\widehat{\boldsymbol{\beta}}_{\text{med}}^{(1)}, \widehat{\beta}_{\text{med},0}^{(1)}) &= \underset{\boldsymbol{\beta}\in\mathbb{R}^p, \beta_0\in\mathbb{R}}{\arg\min} \Big\{ n^{-1} \sum_{i=1}^n |\widehat{Y}_i^{(2)} - \boldsymbol{\beta}^{\mathrm{T}} \widehat{\mathbf{x}}_i^{(2)} - \beta_0| + \lambda^{(1)} \|\boldsymbol{\beta}\|_1 \Big\}, \\
(\widehat{\boldsymbol{\beta}}_{\text{med}}^{(2)}, \widehat{\beta}_{\text{med},0}^{(2)}) &= \underset{\boldsymbol{\beta}\in\mathbb{R}^p, \beta_0\in\mathbb{R}}{\arg\min} \Big\{ n^{-1} \sum_{i=1}^n |\widehat{Y}_i^{(1)} - \boldsymbol{\beta}^{\mathrm{T}} \widehat{\mathbf{x}}_i^{(1)} - \beta_0| + \lambda^{(2)} \|\boldsymbol{\beta}\|_1 \Big\},
\end{aligned}
$$

where $\lambda^{(1)}$ and $\lambda^{(2)}$ are tuning parameters selected via cross-validation. The debiased estimator is then defined as

$$
\widehat{\boldsymbol{\beta}}_{\text{med}}^{\text{d},(l)} = \widehat{\boldsymbol{\beta}}_{\text{med}}^{(l)} + \widehat{\boldsymbol{\Theta}}^{\mathrm{T}} n^{-1} \sum_{i=1}^n \widehat{\mathbf{x}}_i^{(l)} \text{sign}\{\widehat{Y}_i^{(l)} - (\widehat{\boldsymbol{\beta}}_{\text{med}}^{(l)})^{\mathrm{T}} \widehat{\mathbf{x}}_i^{(l)} - \widehat{\beta}_{\text{med},0}^{(l)}\} / (2\widehat{\kappa}_{\text{med}}),
$$

for each $l \in \{1, 2\}$. Here, $\widehat{\boldsymbol{\Theta}}$ is the estimator of $\boldsymbol{\Theta}$ as in Section 3, and $\widehat{\kappa}_{\text{med}}$ is the estimator of $\kappa_{\text{med}} = f\{\text{med}(\epsilon_i)\}$, with $f(\cdot)$ and $\text{med}(\epsilon_i)$ denoting the probability density function and the population median of ϵ_i, respectively. Specifically, we define $\widehat{\kappa}_{\text{med}}^{(l)} = n^{-1} \sum_{i=1}^n K(\widehat{\epsilon}_{\text{med},i}^{(l)}/h)/h$, where $K(\cdot)$ is a kernel function, $h > 0$ is the bandwidth, and

$\widehat{\epsilon}^{(l)}_{\text{med},i} = \widehat{Y}^{(l)}_i - (\widehat{\mathbf{x}}^{(l)}_i)^{\mathrm{T}}\widehat{\boldsymbol{\beta}}^{(l)}_{\text{med}} - \widehat{\beta}^{(l)}_{\text{med},0}$. The final estimator for κ_{med} is $\widehat{\kappa}_{\text{med}} = (\widehat{\kappa}^{(1)}_{\text{med}} + \widehat{\kappa}^{(2)}_{\text{med}})/2$. The final debiased DS Lasso-penalized median estimator is

$$\widehat{\boldsymbol{\beta}}^{\text{d}}_{\text{med}} = \frac{\widehat{\boldsymbol{\beta}}^{\text{d},(1)}_{\text{med}} + \widehat{\boldsymbol{\beta}}^{\text{d},(2)}_{\text{med}}}{2}.$$

Following arguments similar to Bradic and Kolar (2017) and the proof of Theorem 5, we can show that, under certain conditions,

$$\sqrt{N}(\widehat{\beta}^{\text{d}}_{\text{med},k} - \beta_{0,k})/\widehat{\omega}_{\text{med},k} \xrightarrow{d} \mathcal{N}(0,1),$$

for each $k \in \{1,\ldots,p\}$, where $\widehat{\beta}^{\text{d}}_{\text{med},k}$ is the k-th component of $\widehat{\boldsymbol{\beta}}^{\text{d}}_{\text{med}}$,

$$\widehat{\omega}^2_{\text{med},k} = \widehat{\boldsymbol{\theta}}_k^{\mathrm{T}}\Big(\frac{\widehat{\boldsymbol{\Sigma}}^{(1)} + \widehat{\boldsymbol{\Sigma}}^{(2)}}{2}\Big)\widehat{\boldsymbol{\theta}}_k \Big/ (4\widehat{\kappa}^2_{\text{med}}),$$

with $\widehat{\boldsymbol{\Sigma}}^{(l)} = n^{-1}\sum_{i=1}^n (\widehat{\mathbf{x}}^{(l)}_i - \overline{\widehat{\mathbf{x}}}^{(l)})(\widehat{\mathbf{x}}^{(l)}_i - \overline{\widehat{\mathbf{x}}}^{(l)})^{\mathrm{T}}$ and $\overline{\widehat{\mathbf{x}}}^{(l)} = n^{-1}\sum_{i=1}^n \widehat{\mathbf{x}}^{(l)}_i$. The $(1-\alpha)$-confidence interval for $\beta_{0,k}$ is thus

$$\Big[\widehat{\beta}^{\text{d}}_{\text{med},k} - \Phi^{-1}(1-\alpha/2)\widehat{\omega}_{\text{med},k}/\sqrt{N}, \quad \widehat{\beta}^{\text{d}}_{\text{med},k} + \Phi^{-1}(1-\alpha/2)\widehat{\omega}_{\text{med},k}/\sqrt{N}\Big].$$

For $\omega^2_{\text{med},k} = \Theta_{k,k}/(4\kappa^2_{\text{med}})$, we have $\widehat{\omega}^2_{\text{med},k}/\omega^2_{\text{med},k}$ probability to 1, and hence the asymptotic relative efficiency (ARE) of β^{d}_k with respect to $\beta^{\text{d}}_{\text{med},k}$ is $3\{f^*(0)\}^2/\kappa^2_{\text{med}}$. When ϵ_i follows a normal distribution, the ARE is 1.5, indicating significant efficiency loss when using the debiased DS Lasso-penalized median estimator compared to the debiased DS rank Lasso estimator.

Table 9 compares the performance of the proposed methods with the median-based method described above (denoted DSMLasso). The evaluation metrics are the same as those in Example 2, and the results are based on 200 simulations. Overall, the confidence intervals for the median-based method are significantly wider than those for the proposed rank-based methods when the random errors follow a normal distribution. As the error distributions become more heavy-tailed, the difference diminishes, and the median-based intervals even become slightly shorter than those of the proposed methods under a Cauchy error distribution.

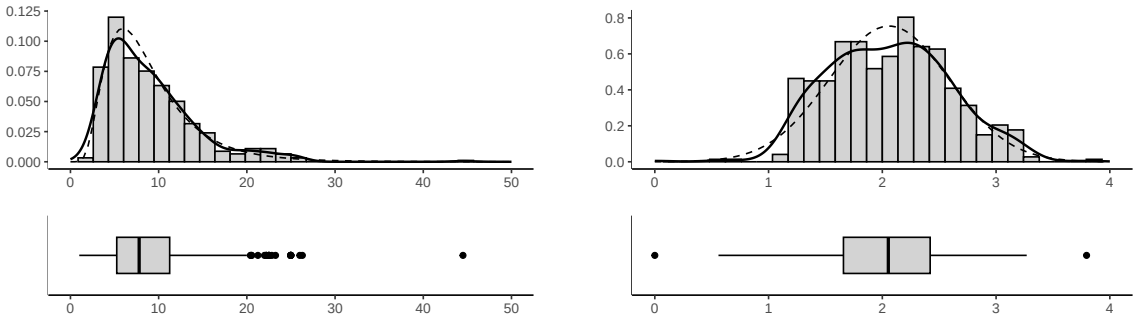
B.6 An Additional Real Example

We apply the proposed methods to analyze workers' wage data from Berndt (1991), which includes 534 workers' wages, years of experience, education, living region, gender, race, occupation, and marital status. This data set has also been studied in Xie and Huang (2009) and Zhu et al. (2019), both of which concern high-dimensional partially linear models. We consider the following model:

$$Y_i = \sum_{k=1}^{14} X_{i,k}\beta_{0,k} + g_0(Z_i) + \epsilon_i, \ i = 1,\ldots,534,$$

Error	Method	CP_1	CP_{500}	$ACP_{\mathcal{B}}$	$ACP_{\mathcal{B}^c}$	$ARP_{\mathcal{B}}$	RCP_1	$AIL_{\mathcal{B}}$	$AIL_{\mathcal{B}^c}$
$\mathcal{N}(0, 2^2)$	DSRLasso	0.955	0.955	0.942	0.949	1.000	0.955	0.483	0.476
	$\text{DSRLasso}^{\text{sub}}$	0.960	0.950	0.943	0.949	1.000	0.960	0.483	0.477
	DSMLasso	0.930	0.955	0.945	0.951	1.000	0.930	0.596	0.588
MN	DSRLasso	0.945	0.950	0.937	0.948	1.000	0.945	0.329	0.326
	$\text{DSRLasso}^{\text{sub}}$	0.955	0.950	0.942	0.948	1.000	0.955	0.330	0.326
	DSMLasso	0.960	0.960	0.943	0.953	1.000	0.960	0.369	0.366
$t(2)$	DSRLasso	0.950	0.960	0.940	0.949	1.000	0.950	0.555	0.560
	$\text{DSRLasso}^{\text{sub}}$	0.965	0.960	0.943	0.949	1.000	0.965	0.556	0.562
	DSMLasso	0.940	0.950	0.947	0.952	1.000	0.940	0.584	0.577
Cauchy	DSRLasso	0.950	0.945	0.937	0.951	0.992	0.940	1.104	1.116
	$\text{DSRLasso}^{\text{sub}}$	0.945	0.935	0.933	0.952	0.987	0.930	1.137	1.150
	DSMLasso	0.950	0.965	0.955	0.962	0.990	0.945	1.067	1.079

Table 9: A Comparison of the Proposed Methods and the Median-based Method. Empirical Coverage Probabilities, Power and Lengths of Confidence Intervals Based on 200 Simulations.



(a) (b)

Figure 5: (a) Histogram of Wages (Y_i) for 534 Workers, Overlaid with a Boxplot, a Kernel Density Curve (Solid Line), and a Log-normal Density Curve (Dashed Line); (b) Histogram of Natural Logarithms of Wages ($\log(Y_i)$), Overlaid with a Boxplot, a Kernel Density Curve (Solid Line), and a Normal Density Curve (Dashed Line).

	Description	TSLasso		DSRLasso	
		$\widehat{\beta}_k^{\mathrm{d}}$(SE)	p-value	$\widehat{\beta}_k^{\mathrm{d}}$(SE)	p-value
X_1	Years of education	0.620(0.100)	< 0.001	0.486(0.088)	< 0.001
X_2	1 = southern region, 0 = other	-0.505(0.403)	0.209	-0.882(0.353)	0.013
X_3	1 = female, 0 = other	-2.002(0.411)	< 0.001	-1.833(0.367)	< 0.001
X_4	1 = union member, 0 = nonmember	1.582(0.479)	0.001	1.767(0.413)	< 0.001
X_5	1 = nonwhite, 0 = other	-0.838(0.554)	0.130	-0.547(0.480)	0.255
X_6	1 = Hispanic, 0 = other	-0.545(0.838)	0.515	-1.089(0.741)	0.142
X_7	1 = management, 0 = other	3.401(0.767)	< 0.001	2.710(0.660)	< 0.001
X_8	1 = sales, 0 = other	-0.588(0.793)	0.458	-0.195(0.681)	0.775
X_9	1 = clerical, 0 = other	0.084(0.630)	0.894	0.620(0.542)	0.253
X_{10}	1 = service, 0 = other	-0.522(0.643)	0.417	-0.265(0.561)	0.637
X_{11}	1 = professional, 0 = other	2.096(0.689)	0.002	1.919(0.599)	0.001
X_{12}	1 = manufacturing, 0 = other	1.171(0.537)	0.029	1.398(0.465)	0.003
X_{13}	1 = construction, 0 = other	0.691(0.930)	0.458	0.922(0.797)	0.247
X_{14}	1 = married, 0 = other	0.026(0.410)	0.949	0.428(0.357)	0.230

Table 10: Estimates of $\beta_{0,k}$ $(k = 1, \ldots, 14)$ for the Real Data Example. The debiased estimators $(\widehat{\beta}_k^{\mathrm{d}})$ are presented alongside their standard errors (in parentheses) and corresponding p-values.

where Y_i denotes the wage of the i-th worker, Z_i represents the years of experience, and ϵ_i $(i = 1, \ldots, 534)$ are independent and identically distributed random errors. The design matrix $\mathbf{X} = (X_{i,k})_{534\times 14}$ contains covariates other than the years of experience, with $X_{i,k}$ being the k-th covariate for the i-th worker.

The histogram of wages for 534 workers and their corresponding natural logarithm transformations are presented in Figures 5 (a) and (b), respectively. Boxplots and kernel density curves are included in both panels. Additionally, a log-normal density curve is overlaid in Figure 5 (a), while a normal density curve is overlaid in Figure 5 (b). Notably, the wage distribution shows a close resemblance to a log-normal distribution, which motivates the application of the proposed debiased DS rank Lasso (DSRLasso) to this data set.

Table 10 presents the debiased DSRLasso estimators of $\beta_{0,k}$ $(k = 1, \ldots, 14)$, along with their standard errors and p-values. For comparison, we also include results from the debiased two-step projection based Lasso (TSLasso; Zhu, 2017; Zhu et al., 2019). The signs of the debiased DSRLasso and TSLasso estimators align exactly. However, debiased DSRLasso yields consistently smaller standard errors. Furthermore, debiased DSRLasso identifies one additional significant covariate (X_2, whether living in the south or not) at the 5% significance level, compared to debiased TSLasso.

To evaluate the performance of debiased DSRLasso in high-dimensional settings, we expand the original design matrix $\mathbf{X}$ into an augmented matrix $\mathbf{X}_{\mathrm{aug}} = (\mathbf{X}, \mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(40)}) \in \mathbb{R}^{534\times 574}$. Each submatrix $\mathbf{X}^{(b)}$ $(b = 1, \ldots, 40)$ is constructed by randomly sampling rows from $\mathbf{X}$, ensuring each row of $\mathbf{X}^{(b)}$ has the same distribution as the original covariates

	Description	TSLasso	DSRLasso
X_1	Years of education	1.000	1.000
X_2	1 = southern region, 0 = other	0.110	0.985
X_3	1 = female, 0 = other	1.000	1.000
X_4	1 = union member, 0 = nonmember	1.000	1.000
X_5	1 = nonwhite, 0 = other	0.015	0.005
X_6	1 = Hispanic, 0 = other	0.000	0.160
X_7	1 = management, 0 = other	1.000	1.000
X_8	1 = sales, 0 = other	0.050	0.195
X_9	1 = clerical, 0 = other	0.000	0.000
X_{10}	1 = service, 0 = other	0.155	0.580
X_{11}	1 = professional, 0 = other	1.000	0.955
X_{12}	1 = manufacturing, 0 = other	0.830	0.880
X_{13}	1 = construction, 0 = other	0.000	0.040
X_{14}	1 = married, 0 = other	0.000	0.010

Table 11: Rejection probabilities of the first 14 variables at the 5% significance level in 200 replications.

Method	TSLasso		DSRLasso	
	Power	Length	Power	Length
NST	1.000	3.812	1.000	3.558
ST	1.000	2.913	1.000	2.713

Table 12: Empirical Powers and Average Interval Lengths of NST-TSLasso, ST-TSLasso, NST-DSRLasso and ST-DSRLasso for Testing the Simultaneous Hypothesis $H_{0,\mathcal{G}_0} : \beta_{0,1} = \cdots = \beta_{0,14} = 0$. "Power" and "Length" denote the empirical powers and average interval lengths, respectively. The results are based on 200 replications.

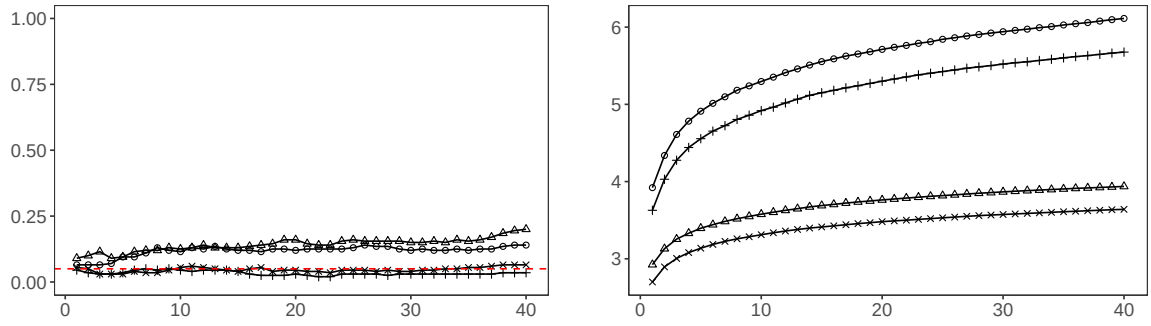
$\{X_{i,k}\}_{k=1}^{14}$. The augmented model is:

$$Y_i = \sum_{k=1}^{574} X_{i,k}\beta_{0,k} + g_0(Z_i) + \epsilon_i,\ i = 1, \ldots, 534,$$

where $\beta_{0,15} = \cdots = \beta_{0,574} = 0$ by construction.

Let p_k denote the p-value for testing the significance of the k-th covariate. The average type I error rate, defined as $\sum_{k=15}^{574} \text{pr}(p_k \leq 0.05)/560$, is estimated based on 200 replications. The resulting values are 0.051 for debiased DSRLasso and 0.052 for debiased TSLasso, demonstrating that both methods effectively control type I errors at the 5% significance level.

To evaluate the relative importance of the first 14 variables, we compare their rejection probabilities at the 5% significance level across 200 replications. As shown in Table 11, covariates with smaller p-values in Table 10 generally demonstrate higher relative importance. Notably, the living region covariate (X_2) is identified as highly important by debiased DSR-



(a) (b)

Figure 6: Empirical Type I Error Rates (a) and Average Interval Lengths (b) for Testing the Simultaneous Hypothesis $H_{0,\mathcal{G}_B} : \beta_{0,15} = \cdots = \beta_{0,14\times(B+1)} = 0$ Across $B \in \{1, \ldots, 40\}$. The results compare four methods: NST-TSLasso ($\circ$), ST-TSLasso ($\triangle$), NST-DSRLasso ($+$), and ST-DSRLasso ($\times$). A red dashed line indicating the significance level $\alpha = 0.05$ is overlaid in (a). All results are based on 200 replications.

Lasso but not by debiased TSLasso, which echoes our prior results in the low-dimensional setting.

Finally, we apply the proposed debiased DSRLasso-based simultaneous testing procedure to test $H_{0,\mathcal{G}} : \beta_{0,k} = 0$ for all $k \in \mathcal{G}$. Here, $\mathcal{G}$ is either $\mathcal{G}_0 = \{1, \ldots, 14\}$ or $\mathcal{G}_B = \{15, \ldots, 14 \times (B + 1)\}$ for $B \in \{1, \ldots, 40\}$. In the case of $\mathcal{G}_0$, the null hypothesis holds, while for $\mathcal{G}_B$ the alternative hypothesis holds. We use the debiased TSLasso-based simultaneous testing procedure from (Zhu et al., 2019) as a benchmark for comparison. Both non-studentized (NST) and studentized (ST) test statistics are considered, leading to four methods: NST-TSLasso, ST-TSLasso, NST-DSRLasso, and ST-DSRLasso. The empirical type I error rates for testing $H_{0,\mathcal{G}_0}$, along with the average lengths of the corresponding simultaneous confidence intervals, are presented in Table 12. These results are based on 200 replications. On the other hand, empirical powers for testing $H_{0,\mathcal{G}_0}$, along with average interval lengths, are shown in Figure 6. The results indicate that NST-type methods generally achieve better control of type I errors for testing $H_{0,\mathcal{G}_B}$ compared to ST-type methods, although with wider interval lengths. Moreover, all four methods achieve an empirical power of one when testing $H_{0,\mathcal{G}_0}$, and our proposed simultaneous testing procedure provides better type I error control and narrower interval lengths than the debiased TSLasso-based procedure.

References

Miguel A. Arcones and Evarist Gine. Limit theorems for U-processes. *The Annals of Probability*, 21(3):1494–1542, 1993.

Alexandre Belloni, Victor Chernozhukov, and Christian Hansen. Inference on treatment effects after selection among high-dimensional controls. *The Review of Economic Studies*, 81(2):608–650, 2014.

Ernst R. Berndt. *The Practice of Econometrics: Classic and Contemporary*. Addison-Wesley Publishing Company, 1991.

Jelena Bradic and Mladen Kolar. Uniform inference for high-dimensional quantile regression: linear functionals and regression rank scores. *arXiv preprint arXiv:1702.06209*, 2017.

Leheng Cai, Xu Guo, Heng Lian, and Liping Zhu. Statistical inference for high-dimensional convoluted rank regression. *Journal of the American Statistical Association*, 0(0):1–25, 2025.

T. Tony Cai and Zijian Guo. Confidence intervals for high-dimensional linear regression: Minimax rates and adaptivity. *The Annals of Statistics*, 45(2):615–646, 2017.

Jinyuan Chang, Song Xi Chen, Cheng Yong Tang, and Tong Tong Wu. High-dimensional empirical likelihood inference. *Biometrika*, 108(1):127–147, 2021.

Patrick Chène. Inhibiting the p53–mdm2 interaction: an important target for cancer therapy. *Nature Reviews Cancer*, 3(2):102–109, 2003.

Victor Chernozhukov, Denis Chetverikov, and Kengo Kato. Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *The Annals of Statistics*, 41(6):2786–2819, 2013.

Victor Chernozhukov, Denis Chetverikov, and Kengo Kato. Comparison and anti-concentration bounds for maxima of Gaussian random vectors. *Probability Theory and Related Fields*, 162(1):47–70, 2015.

Ruben Dezeure, Peter Bühlmann, and Cun-Hui Zhang. High-dimensional simultaneous inference with the bootstrap. *Test*, 26(4):685–719, 2017.

Robert F. Engle, C. W. J. Granger, John Rice, and Andrew Weiss. Semiparametric estimates of the relation between weather and electricity sales. *Journal of the American Statistical Association*, 81(394):310–320, 1986.

Jianqing Fan and Irene Gijbels. *Local Polynomial Modelling and Its Applications*. Chapman & Hall, 1996.

Jianqing Fan, Runze Li, Cun-Hui Zhang, and Hui Zou. *Statistical Foundations of Data Science*. CRC Press, 2020a.

Jianqing Fan, Cong Ma, and Kaizheng Wang. Comment on "a tuning-free robust and efficient approach to high-dimensional regression". *Journal of the American Statistical Association*, 115(532):1720–1725, 2020b.

Long Feng, Changliang Zou, Zhaojun Wang, and Bin Chen. Rank-based score tests for high-dimensional regression coefficients. *Electronic Journal of Statistics*, 7:2131–2149, 2013.

Hui Gong, Yixuan Li, Wen Yang, Zishan Xie, Jiahao Hu, Peihang Li, Rou Xu, Yifan Li, Tianyu Tao, Riqing Li, Shuguang Liu, Yefeng Zhu, Libing Song, and Lishan Fang. Prodh2-mediated metabolism in the bone microenvironment promotes breast cancer metastasis. *Cancer Research*, 85(21):4198–4211, 2025.

Karolina Gronkowska and Agnieszka Robaszkiewicz. Genetic dysregulation of ep300 in cancers in light of cancer epigenome control–targeting of p300-proficient and -deficient cancers. *Molecular Therapy: Oncology*, 32(4):200871, 2024.

Emmanuel Guerre and Camille Sabbah. Uniform bias study and bahadur representation for local polynomial estimators of the conditional quantile function. *Econometric Theory*, 28 (1):87–129, 2012.

Thomas P. Hettmansperger and Joseph W. McKean. *Robust Nonparametric Statistical Methods*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. CRC Press, 2010.

Jean Honorio and Tommi Jaakkola. Tight bounds for the expected risk of linear classifiers and PAC-Bayes finite-sample guarantees. In *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics*, volume 33 of *Proceedings of Machine Learning Research*, pages 384–392. PMLR, 2014.

Shangshang Hu, Muzi Ding, Jinwei Lou, Jian Qin, Yuhan Chen, Zixuan Liu, Yue Li, Junjie Nie, Mu Xu, Huiling Sun, Xinliang Gu, Tao Xu, Shukui Wang, and Yuqin Pan. Col10a1+ fibroblasts promote colorectal cancer metastasis and m2 macrophage polarization with pan-cancer relevance. *Journal of Experimental & Clinical Cancer Research*, 44(1):243, 2025.

Adel Javanmard and Andrea Montanari. Confidence intervals and hypothesis testing for high-dimensional regression. *Journal of Machine Learning Research*, 15(1):2869–2909, 2014.

Adel Javanmard and Andrea Montanari. Debiasing the lasso: Optimal sample size for Gaussian designs. *The Annals of Statistics*, 46(6A):2593 – 2622, 2018.

Roger Koenker. *Quantile Regression*. Econometric Society Monographs. Cambridge University Press, 2005.

Chrysoula Komini, Irini Theohari, Andromachi Lambrianidou, Lydia Nakopoulou, and Theoni Trangas. Papola contributes to cyclin d1 mrna alternative polyadenylation and promotes breast cancer cell proliferation. *Journal of Cell Science*, 134(7):jcs252304, 2021.

M. Ledoux and M. Talagrand. *Probability in Banach Spaces: Isoperimetry and Processes*. A Series of Modern Surveys in Mathematics Series. Springer, 1991.

Qi Li and Jeffrey Scott Racine. *Nonparametric Econometrics: Theory and Practice*. Princeton University Press, 2007.

Elias Masry. Multivariate local polynomial regression for time series: Uniform strong consistency and rates. *Journal of Time Series Analysis*, 17(6):571–599, 1996.

Pascal Massart. About the constants in Talagrand's concentration inequalities for empirical processes. *The Annals of Probability*, 28(2):863–884, 2000.

Léa Montégut, Carlos López-Otín, and Guido Kroemer. Aging and cancer. *Molecular Cancer*, 23(1):106, May 2024.

Pavel Moudry, Katarina Chroma, Sladana Bursac, Sinisa Volarevic, and Jiri Bartek. Rna-interference screen for p53 regulators unveils a role of wdr75 in ribosome biogenesis. *Cell Death & Differentiation*, 29(3):687–696, 2022.

Whitney K. Newey. Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics*, 79(1):147–168, 1997.

Matey Neykov, Yang Ning, Jun S Liu, and Han Liu. A unified theory of confidence regions and testing for high-dimensional estimating equations. *Statistical Science*, 33(3):427–443, 2018.

Yang Ning and Han Liu. A general theory of hypothesis tests and confidence regions for sparse high dimensional models. *The Annals of Statistics*, 45(1):158–195, 2017.

Andrew G. Renehan, Margaret Tyson, Matthias Egger, Richard F. Heller, and Marcel Zwahlen. Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies. *The Lancet*, 371(9612):569–578, 2008.

Sidney Resnick. *A Probability Path*. Birkhäuser Boston, 2019.

Charles J. Stone. Optimal global rates of convergence for nonparametric regression. *The Annals of Statistics*, 10(4):1040–1053, 1982.

Sara van de Geer, Peter Bühlmann, Ya'acov Ritov, and Ruben Dezeure. On asymptotically optimal confidence regions and tests for high-dimensional models. *The Annals of Statistics*, 42(3):1166–1202, 2014.

Roman Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2018.

Martin J. Wainwright. *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2019.

Han Wang, Ziling Zhou, Hanyi Zhong, Shoutang Wang, Kunwei Shen, Renhong Huang, Ruo Wang, and Zheng Wang. A pyrimidine metabolism-related gene signature for prognosis prediction and immune microenvironment description of breast cancer. *Journal of Translational Medicine*, 23(1):698, 2025.

Jun Wang, Xuehong Zhang, Andrew H. Beck, Laura C. Collins, Wendy Y. Chen, Rulla M. Tamimi, Aditi Hazra, Myles Brown, Bernard Rosner, and Susan E. Hankinson. Alcohol consumption and risk of breast cancer by tumor receptor expression. *Hormones and Cancer*, 6(5):237–246, 2015.

Lan Wang, Bo Peng, Jelena Bradic, Runze Li, and Yunan Wu. A tuning-free robust and efficient approach to high-dimensional regression. *Journal of the American Statistical Association*, 115(532):1700–1714, 2020.

Huiliang Xie and Jian Huang. SCAD-penalized regression in high-dimensional partially linear models. *The Annals of Statistics*, 37(2):673–696, 2009.

Limeng Yang, Vanessa A. N. Kraft, Susanne Pfeiffer, Juliane Merl-Pham, Xuanwen Bao, Yu An, Stefanie M. Hauck, and Joel A. Schick. Nonsense-mediated decay factor smg7 sensitizes cells to tnfα-induced apoptosis via cyld tumor suppressor and the noncoding oncogene pvt1. *Molecular Oncology*, 14(10):2420–2435, 2020.

Jung Eun Yoo, Kyungdo Han, Dong Wook Shin, Dahye Kim, Bong-seong Kim, Sohyun Chun, Keun Hye Jeon, Wonyoung Jung, Jinsung Park, Jin Ho Park, Kui Son Choi, and Joo Sung Kim. Association between changes in alcohol consumption and cancer risk. *JAMA Network Open*, 5(8):e2228544–e2228544, 2022.

Nikolaos Zacharakis, Harshini Chinnasamy, Mary Black, Hui Xu, Yong-Chen Lu, Zhili Zheng, Anna Pasetto, Michelle Langhan, Thomas Shelton, Todd Prickett, Jared Gartner, Li Jia, Katarzyna Trebska-McGowan, Robert P. Somerville, Paul F. Robbins, Steven A. Rosenberg, Stephanie L. Goff, and Steven A. Feldman. Immune recognition of somatic mutations leading to complete durable regression in metastatic breast cancer. *Nature Medicine*, 24(6):724–730, 2018.

Cun-Hui Zhang and Stephanie S. Zhang. Confidence intervals for low dimensional parameters in high dimensional linear models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 76(1):217–242, 2014.

Huijun Zhang, An Wang, Yulong Tan, Shaohua Wang, Qinyun Ma, Xiaofeng Chen, and Zelai He. Ncbp1 promotes the development of lung adenocarcinoma through up-regulation of cul4b. *Journal of Cellular and Molecular Medicine*, 23(10):6965–6977, 2019.

Xianyang Zhang and Guang Cheng. Simultaneous inference for high-dimensional linear models. *Journal of the American Statistical Association*, 112(518):757–768, 2017.

Huanyu Zhao, Ruoyu Dang, Yipan Zhu, Baijian Qu, Yasra Sayyed, Ying Wen, Xicheng Liu, Jianping Lin, and Luyuan Li. Hub genes associated with immune cell infiltration in breast cancer, identified through bioinformatic analyses of multiple datasets. *Cancer Biology & Medicine*, 19(9):1352–1374, 2022. doi: 10.20892/j.issn.2095-3941.2021.0586.

Yue Zhou, Tongjia Zhang, Shujie Wang, Ruixiang Yang, Zitao Jiao, Kejia Lu, Hui Li, Wei Jiang, and Xiaowei Zhang. Targeting of hbp1/timp3 axis as a novel strategy against breast cancer. *Pharmacological Research*, 194:106846, 2023.

Yinchu Zhu and Jelena Bradic. Linear hypothesis testing in dense high-dimensional linear models. *Journal of the American Statistical Association*, 113(524):1583–1600, 2018a.

Yinchu Zhu and Jelena Bradic. Significance testing in non-sparse high-dimensional linear models. *Electronic Journal of Statistics*, 12(2):3312–3364, 2018b.

Ying Zhu. Nonasymptotic analysis of semiparametric regression models with high-dimensional parametric coefficients. *The Annals of Statistics*, 45(5):2274–2298, 2017.

Ying Zhu, Zhuqing Yu, and Guang Cheng. High dimensional inference in partially linear models. In *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, volume 89 of *Proceedings of Machine Learning Research*, pages 2760–2769. PMLR, 2019.