mathematics

Article

A Flexibly Conditional Screening Approach via a Nonparametric Quantile Partial Correlation

Xiaochao Xia *,† and **Hao Ming** †

College of Mathematics and Statistics, Chongqing University, Chongqing 401331, China
* Correspondence: xxc@cqu.edu.cn
† These authors contributed equally to this work.

Abstract: Considering the influence of conditional variables is crucial to statistical modeling, ignoring this may lead to misleading results. Recently, Ma, Li and Tsai proposed the quantile partial correlation (QPC)-based screening approach that takes into account conditional variables for ultrahigh dimensional data. In this paper, we propose a nonparametric version of quantile partial correlation (NQPC), which is able to describe the influence of conditional variables on other relevant variables more flexibly and precisely. Specifically, the NQPC firstly removes the effect of conditional variables via fitting two nonparametric additive models, which differs from the conventional partial correlation that fits two parametric models, and secondly computes the QPC of the resulting residuals as NQPC. This measure is very useful in the situation where the conditional variables are highly nonlinearly correlated with both the predictors and response. Then, we employ this NQPC as the screening utility to do variable screening. A variable screening procedure based on NPQC (NQPC-SIS) is proposed. Theoretically, we prove that the NQPC-SIS enjoys the sure screening property that, with probability going to one, the selected subset can recruit all the truly important predictors under mild conditions. Finally, extensive simulations and an empirical application are carried out to demonstrate the usefulness of our proposal.

Keywords: ultrahigh dimensional screening; quantile partial correlation; conditional variables; sure screening property

MSC: 62H30; 62J07

Citation: Xia, X.; Ming, H. A Flexibly Conditional Screening Approach via a Nonparametric Quantile Partial Correlation. *Mathematics* **2022**, *10*, 4638. https://doi.org/10.3390/math10244638

Academic Editors: Andreas Artemiou and Jin-Ting Zhang

Received: 14 November 2022
Accepted: 3 December 2022
Published: 7 December 2022

1. Introduction

Variable screening technique has been demonstrated as a computationally fast and efficient tool in solving many problems in ultrahigh dimensions. For example, in many scientific areas, such as biological genetics, finance and econometrics, we may collect the ultrahigh dimensional data sets (e.g., biomarkers, financial factors, assets and stocks), where the number p_n of predictors extremely exceeds the sample size n. Theoretically, ultrahigh dimension often refers to the dimensionality p_n and sample size n satisfies the relationship: $p_n = O(\exp(n^a))$ for some constant $a > 0$. Variable screening is able to reduce the computational cost, to avoid the instability of algorithms, and to improve the estimation accuracy. These issues exist in the variable selection approaches based on LASSO [1], SCAD [2,3] or MCP [4] for ultrahigh dimensional data. Since the seminal work of [5], which pioneeringly proposed the sure independence screening (SIS) procedure, many variable screening approaches have been consecutively documented over the last fifteen years, including the model-based methods (e.g., [6–11]) and the model-free methods [12–20]. These papers have showed that with probability approaching one, the set of selected predictors contain the set of all truly important predictors.

Most marginal approaches focus only on developing various effective and robust measures to characterize the marginal association between the response and individual predictor. Whereas, these methods do not take into consideration the influence of conditional variables or confounding factors on the response. A simple application of SIS is relatively rough since SIS may perform poorly when predictors are highly correlated with each other. Some predictors that are weakly relevant or irrelevant, but jointly correlated to the response, may be excluded in the final model after applying marginal screening methods. This will result in a high false positive rate (FPR). To surmount this weakness, an iterated screening algorithm or a penalization-based variable selection is usually offered as a refined follow-up step (e.g., [5,10]).

Conditional variable screening can be viewed as an important extension of the marginal screening. It accounts for conditional information when calculating the marginal screening utility. There is relatively less work in the literature. To name a few, Ref. [21] proposed a conditional SIS (CIS) procedure to improve the performance of SIS because some correlated conditional variables may increase the chance of boosting the rank of the marginally weak predictor and that of reducing the number of false negatives. The paper [22] proposed a confounder-adjusted screening method for high dimensional censoring data, in which the additional environmental confounders are regarded as conditional variables. The researchers in [23] studied the variable screening by incorporating within-subject correlation for ultrahigh dimensional longitudinal data, where they used some baseline variables as conditional variables. Ref. [24] proposed a conditional distance correlation-based screening via kernel smoothing method, while [25] further presented a screening procedure based on conditional distance correlation, which is similar to [24] in methodology, but differs in theory. Additionally, Ref. [11] developed a conditional quantile correlation-based screening approach using the B-spline smoothing technique. However, in [11,24,25], among others, the conditional variable they considered is only univariate. Further, Ref. [21] focuses on the generalized linear models, but cannot handle heavy-tailed data. For this regard, we aim to develop a screener that behaves more robustly to outliers and heavy-tailed data, and simultaneously considers more than one conditional variable. On the choice of conditional variables, one can achieve that through some prior knowledge such as published research work or the experience of experts from relevant subjects. When no prior knowledge is available, one can apply some marginal screening approaches, such as the SIS or its robust variants, to select several top-ranked predictors as conditional variables.

On the other hand, to the best of our knowledge, several works have considered multiple conditional variables based on distinct partial correlations. For instance, Ref. [26] proposed a thresholded partial correlation approach to select significant variables in linear regression models. Additionally, Ref. [17] presented a screening procedure on the basis of the quantile partial correlation in [27], and they referred to the procedure as QPC-SIS. More recently, Ref. [28] proposed a copula partial correlation-based screening approach. It is worth noting that the partial correlation used in both [17,28] removes the effect of conditional variables on the response and each predictor through fitting two parametric models with a linear structure. However, this manner may be ineffective, especially when the conditional variables have a nonlinear influence on the response nonlinear. This motivates us to work out a flexible way to control the impact of conditional variables. Meanwhile, we also take into account the issue of the robustness to outlying or heavy-tail response in this paper.

This paper contributes a robust and flexible conditional variable screening procedure via a partial correlation coefficient, which is a non-trivial extension of [17]. First of all, in order to precisely control conditional variables, we propose a nonparametric definition of QPC, which extends that of [17] and allows for more flexibility. Specifically, we first fit two nonparametric additive models to remove the effect of conditional variables on the response and an individual predictor, where we use the B-spline smoothing technique to estimate the nonparametric functions. This can be viewed as a nonparametric adjustment for controlling conditional variables. By that, we can obtain two residuals, on which a quantile correlation can be calculated to formulate a nonparametric QPC. Second, we use this quantity as the screening utility in variable screening. This procedure can be implemented rapidly. We refer to this procedure as the nonparametric quantile partial correlation-based screening, denoted as NQPC-SIS. Third, theoretically, we establish the sure screening property for NQPC-SIS under some mild conditions. Compared to [17], our approach is more flexible and our theory on the sure screening property is more difficult to derive. Moreover, our screening idea can be easily transferred to some existing screening methods that use some popular partial correlation.

The remainder of the paper is organized as follows. In Section 2, the NQPC-SIS is introduced. The technical conditions needed are listed and asymptotic properties are established in Section 3. Section 4 provides an iterative algorithm for a further refinement. Numerical studies and empirical analysis of real data set are carried out in Section 5. Concluding remarks are given in Section 6. All the proofs of the main results are relegated to the Appendix A.

2. Methodology

2.1. A Preliminary

In this section, we formally introduce the NQPC-SIS procedure. To begin with, we give some background on the quantile correlation (QC) introduced in [27]. Let X and Y be two random variables, and EX be the expectation of X. The definition of QC is formulated as

$$\mathrm{qcor}_\tau(Y,X) = \frac{\mathrm{E}[\psi_\tau(Y - Q_{\tau,Y})(X - \mathrm{E}(X))]}{\sqrt{\mathrm{var}(I(Y - Q_{\tau,Y} > 0))\mathrm{var}(X)}}, \tag{1}$$

where $Q_{\tau,Y}$ is the τth quantile of Y, and $\psi_\tau(u) = \tau - I(u < 0)$ for some quantile level $\tau \in (0,1)$, here $I(\cdot)$ denotes an indicator function. This correlation takes on a value between -1 and 1 and is asymmetric with respect to Y and X compared with the conventional correlation coefficient. The QC shares the merits: the property of monotone invariance for Y as well as the robustness of Y, due to the use of the quantile rather than the mean in the definition. Thus, QC affects little in the presence of outliers in Y. Besides, as shown in [27], $\mathrm{qcor}_\tau(Y,X)$ is closely related to the quantile regression. If we denote by $(a^*_{0\tau}, a^*_{1\tau})$ the minimizer of $\mathrm{E}\{\rho_\tau(Y - a_{0\tau} - a_{1\tau}X)\}$ with respect to $a_{0\tau}$ and $a_{1\tau}$, where $\rho_\tau(u) = u[\tau - I(u < 0)]$. Then, it follows that $\mathrm{qcor}_\tau(Y,X) = \varphi(a^*_{1\tau})$, where $\varphi(\cdot)$ is a continuous and increasing function, and $\varphi(a^*_{1\tau}) = 0$ if and only if $a^*_{1\tau} = 0$.

When QC is used as a marginal screening utility for variable screening, the screening results obtained may be misleading when the predictors are highly correlated. To overcome this problem, Ref. [17] proposed the screening based on quantile partial correlation (QPC) to reduce the effect from conditional predictors. For the sake of presentation, write $\mathbf{X}_{-j} = (X_k, k \neq j)^T$ for $j = 1, \ldots, p_n$. The QPC in [17] is defined as

$$\begin{aligned}\mathrm{qpcor}_\tau(Y, X_j|\mathbf{X}_{-j}) &= \frac{\mathrm{cov}(\psi_\tau(Y - \mathbf{X}^T_{-j}\boldsymbol{\alpha}^0_j), X_j - \mathbf{X}^T_{-j}\boldsymbol{\theta}^0_j)}{\sqrt{\mathrm{var}(\psi_\tau(Y - \mathbf{X}^T_{-j}\boldsymbol{\alpha}^0_j))\mathrm{var}(X_j - \mathbf{X}^T_{-j}\boldsymbol{\theta}^0_j)}} \\ &= \frac{\mathrm{E}\{\psi_\tau(Y - \mathbf{X}^T_{-j}\boldsymbol{\alpha}^0_j)(X_j - \mathbf{X}^T_{-j}\boldsymbol{\theta}^0_j)\}}{\sqrt{\tau(1-\tau)\sigma^2_j}},\end{aligned} \tag{2}$$

where $\sigma^2_j = \mathrm{var}(X_j - \mathbf{X}^T_{-j}\boldsymbol{\theta}^0_j)$, $\boldsymbol{\alpha}^0_j = \mathrm{argmin}_{\boldsymbol{\alpha}_j}\mathrm{E}\{\rho_\tau(Y - \mathbf{X}^T_{-j}\boldsymbol{\alpha}_j)\}$ and $\boldsymbol{\theta}^0_j = \mathrm{argmin}_{\boldsymbol{\theta}_j}\mathrm{E}\{(X_j - \mathbf{X}^T_{-j}\boldsymbol{\theta}_j)^2\}$. When applying the QPC to variable screening, we must estimate two quantities $\boldsymbol{\alpha}^0_j$ and $\boldsymbol{\theta}^0_j$ in advance. However, for ultrahigh dimensional data, the dimensionality of $\mathbf{X}_{-j}$ is $p_n - 1$, which can still be much bigger than the sample size n. In this situation, it is difficult to obtain the estimators of $\boldsymbol{\alpha}^0_j$ and $\boldsymbol{\theta}^0_j$. On the other hand, it is usually believed that the useful conditional variables are relatively less. Thus, it is reasonable to consider a small subset of $\{k : k \neq j, k = 1 \ldots, p_n\}$, denoted by $\mathcal{S}_j$ in practice. Here, $\mathcal{S}_j$ is said to be conditional set with a size smaller than n and it can be specified as the set of previously selected variables and the variables related to the jth predictor, if there is no prior knowledge on it. As a result, Ref. [17] suggested using the following measure to perform variable screening:

$$\mathrm{qpcor}_\tau(Y, X_j|\mathbf{X}_{\mathcal{S}_j}) = \frac{\mathrm{E}\{\psi_\tau(Y - \mathbf{X}^T_{\mathcal{S}_j}\boldsymbol{\alpha}^0_j)(X_j - \mathbf{X}^T_{\mathcal{S}_j}\boldsymbol{\theta}^0_j)\}}{\sqrt{\tau(1-\tau)\sigma^2_j}}, \tag{3}$$

where $\sigma_j^2 = \mathrm{var}(X_j - \mathbf{X}_{\mathcal{S}_j}^T \boldsymbol{\theta}_j^0)$, $\boldsymbol{\alpha}_j^0 = \mathrm{argmin}_{\boldsymbol{\alpha}_j} \mathrm{E}\{\rho_\tau(Y - \mathbf{X}_{\mathcal{S}_j}^T \boldsymbol{\alpha}_j)\}$ and $\boldsymbol{\theta}_j^0 = \mathrm{argmin}_{\boldsymbol{\theta}_j} \mathrm{E}\{(X_j - \mathbf{X}_{\mathcal{S}_j}^T \boldsymbol{\theta}_j)^2\}$, in which $\mathbf{X}_{\mathcal{S}_j} = (X_k, k \in \mathcal{S}_j)^T$.

From the definition, one can see that the QPC is just the QC between Y and X_j after removing the confounding effects of conditional variables $\mathbf{X}_{\mathcal{S}_j}$. Typically, it is through fitting two parametric regression models: one is to fit a linear quantile regression of Y on $\mathbf{X}_{\mathcal{S}_j}$, and another is on a multivariate linear regression of X_j on $\mathbf{X}_{\mathcal{S}_j}$. Afterwards, the QPC computes the QC of two residuals that are obtained from these two regression fittings. However, in real applications, the parametric models used to dispel the confounding effects may not be adequate, especially when a nonlinear dependence structure between the response and the predictions is present, which is quite common in high-dimensional data. This motivates us to consider a more flexible and efficient approach to control the influence of the confounding/conditional variables.

2.2. Proposed Method: NQPC-SIS

Without loss of generality, we assume that the predictors $\{X_j, 1 \leq j \leq p\}$ are standardized and the response Y satisfies τ-qauntile centered, i.e., $Q_{\tau,Y} = 0$, which is similar to the treatment where the response is centered by mean. Then, we consider the quantile additive model as

$$Y = m_1(X_1) + m_2(X_2) + \cdots + m_p(X_p) + \varepsilon,$$

where the error term satisfies $P(\varepsilon < 0|\mathbf{X}) = \tau$. This means that the conditional τ-quantile of Y given $\mathbf{X}$ is $Q_{\tau,Y|\mathbf{X}} = m_1(X_1) + m_2(X_2) + \cdots + m_p(X_p)$. We denote by $\mathcal{M}_* = \{j : m_j(X_j) \neq 0, 1 \leq j \leq p\}$ the active set, which indicates the set of indices associated with the nonzero coefficients in the true model and is often assumed to be sparse.

Let $|\mathcal{S}_j|$ be the cardinality of a set $\mathcal{S}_j$, and m_{jk} and g_{jk}, $k \in \mathcal{S}_j$, be ℓ_2-smoothing functions satisfying some conditions. For the identification, we require that $\int m_{jk}(x)\mathrm{d}x = 0$ and $\mathrm{E}\{g_{jk}(X_k)\} = 0$ for all j, k. Set $m_j(\mathbf{X}_{\mathcal{S}_j}) = \sum_{k \in \mathcal{S}_j} m_{jk}(X_k)$ and $g_j(\mathbf{X}_{\mathcal{S}_j}) = \sum_{k \in \mathcal{S}_j} g_{jk}(X_k)$. A nonparametric version of QPC (denoted as NQPC) is formulated as

$$\varrho_\tau(Y, X_j|\mathbf{X}_{\mathcal{S}_j}) = \frac{\mathrm{E}\{\psi_\tau(Y - m_j^0(\mathbf{X}_{\mathcal{S}_j}))(X_j - g_j^0(\mathbf{X}_{\mathcal{S}_j}))\}}{\sqrt{\tau(1-\tau)\sigma_{j,0}^2}}, \quad (4)$$

where $\sigma_{j,0}^2 = \mathrm{var}(X_j - g_j^0(\mathbf{X}_{\mathcal{S}_j}))$, $m_j^0 = \mathrm{argmin}_{m_j} \mathrm{E}\{\rho_\tau(Y - m_j(\mathbf{X}_{\mathcal{S}_j}))\}$ and $g_j^0 = \mathrm{argmin}_{g_j} \mathrm{E}\{(X_j - g_j(\mathbf{X}_{\mathcal{S}_j}))^2\}$. Suppose we have a dataset: $\{(Y_i, \mathbf{X}_i), i = 1, \cdots, n\}$ consisting of n independent copies of $(Y, \mathbf{X})$, where the dimensionality of $\mathbf{X}_i$ is p_n. Let $\mathbf{X}_{i,\mathcal{S}_j}$ be the sub-vector of $\mathbf{X}_i$ indexed by $\mathcal{S}_j$. Then, a sample estimate for NQPC can be given as

$$\widetilde{\varrho}_\tau(Y, X_j|\mathbf{X}_{\mathcal{S}_j}) = \frac{n^{-1}\sum_{i=1}^n \psi_\tau(Y_i - \widetilde{m}_j(\mathbf{X}_{i,\mathcal{S}_j}))(X_{ij} - \widetilde{g}_j(\mathbf{X}_{i,\mathcal{S}_j}))}{\sqrt{\tau(1-\tau)\widetilde{\sigma}_j^2}}, \quad (5)$$

where $\widetilde{\sigma}_j^2 = n^{-1}\sum_{i=1}^n (X_{ij} - \widetilde{g}_j(\mathbf{X}_{i,\mathcal{S}_j}))^2$, $\widetilde{m}_j = \mathrm{argmin}_{m_j} \frac{1}{n}\sum_{i=1}^n \rho_\tau(Y_i - m_j(\mathbf{X}_{i,\mathcal{S}_j}))$ and $\widetilde{g}_j = \mathrm{argmin}_{g_j} \frac{1}{n}\sum_{i=1}^n (X_{ij} - g_j(\mathbf{X}_{i,\mathcal{S}_j}))^2$. Since m_{jk} and g_{jk} are unknown nonparametric functions, so $\widetilde{m}_j$ and $\widetilde{g}_j$ cannot be used, rendering $\widetilde{\varrho}_\tau(Y, X_j|\mathbf{X}_{\mathcal{S}_j})$ inapplicable. In what follows, we estimate each of m_{jk}s and g_{jk}s by making use of nonparametric B-spline approximation.

To proceed, we denote $\{B_k(\cdot), k = 1, \cdots, L_n\}$ with $\|B_k\|_\infty \leq 1$ by a sequence of normalized and centered B-spline basis functions, where L_n is the number of basis functions. Then, according to the theory of B-spline approximation ([29]), for a generic smoothing function m, there exists a vector $\boldsymbol{\gamma} \in \mathrm{R}^{L_n}$ such that $m(x) \approx \mathbf{B}(x)^T\boldsymbol{\gamma}$, where $\mathbf{B}(\cdot) = (B_1(\cdot), \cdots, B_{L_n}(\cdot))^T$. Therefore, there exist vectors $\boldsymbol{\alpha}_{jk} \in \mathrm{R}^{L_n}$ and $\boldsymbol{\theta}_{jk} \in \mathrm{R}^{L_n}$ such that $m_{jk}(X_k) \approx \mathbf{B}(X_k)^T\boldsymbol{\alpha}_{jk}$ and $g_{jk}(X_k) \approx \mathbf{B}(X_k)^T\boldsymbol{\theta}_{jk}$. Since $\int m_{jk}(x)\mathrm{d}x = 0$ and $\mathrm{E}\{g_{jk}(X_k)\} = 0$, it naturally implies that $\mathrm{E}\{\mathbf{B}(X_k)\} = 0$ for $k \in \mathcal{S}_j$. Write $\boldsymbol{\alpha}_j = (\{\boldsymbol{\alpha}_{jk}^T, k \in \mathcal{S}_j\})^T$, $\boldsymbol{\theta}_j = (\{\boldsymbol{\theta}_{jk}^T, k \in \mathcal{S}_j\})^T$

and $\mathbf{B}_j = (\{\mathbf{B}(X_k)^T, k \in \mathcal{S}_j\})^T$. Denote by $\widehat{m}_j(\mathbf{X}_{i,\mathcal{S}_j}) = \mathbf{B}_{ij}^T\widehat{\boldsymbol{\alpha}}_j$, $\widehat{g}_j(\mathbf{X}_{i,\mathcal{S}_j}) = \mathbf{B}_{ij}^T\widehat{\boldsymbol{\theta}}_j$ and $\widehat{\sigma}_j^2 = n^{-1}\sum_{i=1}^n (X_{ij} - \widehat{g}_j(\mathbf{X}_{i,\mathcal{S}_j}))^2$, where

$$\widehat{\boldsymbol{\alpha}}_j = \mathrm{argmin}_{\boldsymbol{\alpha}_j} \frac{1}{n}\sum_{i=1}^n \rho_\tau(Y_i - \mathbf{B}_{ij}^T\boldsymbol{\alpha}_j)$$

and

$$\widehat{\boldsymbol{\theta}}_j = \mathrm{argmin}_{\boldsymbol{\theta}_j} \frac{1}{n}\sum_{i=1}^n (X_{ij} - \mathbf{B}_{ij}^T\boldsymbol{\theta}_j)^2,$$

where $\mathbf{B}_{ij}$ indicates $\mathbf{B}_j$ within $\mathbf{B}(X_k)$ being replaced by $\mathbf{B}(X_{ik})$ for $i = 1, \cdots, n$ and $k \in \mathcal{S}_j$. Then, it follows that a feasible sample estimate for NQPC is given by

$$\widehat{\varrho}_\tau(Y, X_j|\mathbf{X}_{\mathcal{S}_j}) = \frac{n^{-1}\sum_{i=1}^n \psi_\tau(Y_i - \widehat{m}_j(\mathbf{X}_{i,\mathcal{S}_j}))(X_{ij} - \widehat{g}_j(\mathbf{X}_{i,\mathcal{S}_j}))}{\sqrt{\tau(1-\tau)\widehat{\sigma}_j^2}}. \tag{6}$$

Next, we employ the above NQPC estimator as a screening utility in variable screening. To this end, we denote $\widehat{\mathcal{M}}_{\nu_n}$ to be the selected active set via the screening procedure such that the maximal absolute sample NQPC of the selected variables in $\widehat{\mathcal{M}}_{\nu_n}$ are greater than a user-specified threshold value ν_n. In other words, we can select an active set of variables by

$$\widehat{\mathcal{M}}_{\nu_n} = \{j : |\widehat{\varrho}_\tau(Y, X_j|\mathbf{X}_{\mathcal{S}_j})| \geq \nu_n \text{ for } 1 \leq j \leq p\}. \tag{7}$$

We name this procedure as the NQPC-based variable screening, abbreviated as NQPC-SIS. In the next section, we will provide some theoretical justification for this approach.

3. Theoretical Properties

To state our theoretical results, we first make some notations. Let $r_n = \max_{1\leq j\leq p}|\mathcal{S}_j|$. Throughout the rest of the paper, for any matrix $\mathbf{A}$, we use $\|\mathbf{A}\| = \sqrt{\lambda_{\max}(\mathbf{A}^T\mathbf{A})}$, $\|\mathbf{A}\|_\infty = \max_{i,j}|A_{ij}|$, and $\lambda_{\min}(\mathbf{A})$ and $\lambda_{\max}(\mathbf{A})$ to stand for the operator norm, the infinity norm as well as the minimum and maximum eigenvalues for a symmetric matrix $\mathbf{A}$, respectively. In addition, for any vector $\mathbf{a}$, $\|\mathbf{a}\| = \sqrt{\sum_i a_i^2}$ means the Euclidean norm.

Denote $u_j = |\varrho_\tau(Y, X_j|\mathbf{X}_{\mathcal{S}_j})|$ and $\widehat{u}_j = |\widehat{\varrho}_\tau(Y, X_j|\mathbf{X}_{\mathcal{S}_j})|$, where $\varrho_\tau(Y, X_j|\mathbf{X}_{\mathcal{S}_j})$ is given in Equation (4) and $\widehat{\varrho}_\tau(Y, X_j|\mathbf{X}_{\mathcal{S}_j})$ is given in Equation (7). Further, we also denote $u_j^* = |\varrho_\tau^*(Y, X_j|\mathbf{X}_{\mathcal{S}_j})|$, where

$$\varrho_\tau^*(Y, X_j|\mathbf{X}_{\mathcal{S}_j}) = \frac{\mathrm{E}\{\psi_\tau(Y - \mathbf{B}_j^T\boldsymbol{\alpha}_j^0)(X_j - \mathbf{B}_j^T\boldsymbol{\theta}_j^0)\}}{\sqrt{\tau(1-\tau)\sigma_j^2}}, \tag{8}$$

where $\sigma_j^2 = \mathrm{var}(X_j - \mathbf{B}_j^T\boldsymbol{\theta}_j^0)$, $\boldsymbol{\alpha}_j^0 = \mathrm{argmin}_{\boldsymbol{\alpha}_j}\mathrm{E}\{\rho_\tau(Y - \mathbf{B}_j^T\boldsymbol{\alpha}_j)\}$ and $\boldsymbol{\theta}_j^0 = \mathrm{argmin}_{\boldsymbol{\theta}_j}\mathrm{E}\{(X_j - \mathbf{B}_j^T\boldsymbol{\theta}_j)^2\}$. Before we establish the uniform convergence of $\widehat{u}_j$ to u_j, we first investigate the bound of the gap between u_j and u_j^*, which is helpful to understand the marginal signal level after applying B-spline approximation to the population utility. We need the following conditions:

(B1) We assume that $\mathrm{E}\{X_j|\mathbf{X}_{\mathcal{S}_j}\} = g_j^0(\mathbf{X}_{\mathcal{S}_j}) = \sum_{k\in\mathcal{S}_j} g_{jk}^0(X_k)$ and $\mathcal{X}_k$ denotes the support of covariate X_k. There exist some positive constants C_g and C_m such that for any $k \in \mathcal{S}_j$,

$$\max_{1\leq j\leq p}\sup_{x\in\mathcal{X}_k}|g_{jk}^0(x) - \mathbf{B}_j(x)^T\boldsymbol{\theta}_{jk}^0| \leq C_g L_n^{-d},$$

$$\max_{1\leq j\leq p}\sup_{x\in\mathcal{X}_k}|m_{jk}^0(x) - \mathbf{B}_j(x)^T\boldsymbol{\alpha}_{jk}^0| \leq C_m L_n^{-d},$$

where d is defined in condition (C1) below.

(B2) There exist some positive constants $c_{\sigma,\min}, c_{\sigma,\max}, \tilde{c}_{\sigma,\min}, \tilde{c}_{\sigma,\max}$ such that

$$0 < c_{\sigma,\min} \leq \max_{1\leq j\leq p} \sigma_j^2 \leq c_{\sigma,\max} < \infty,$$
$$0 < \tilde{c}_{\sigma,\min} \leq \max_{1\leq j\leq p} \sigma_{j,0}^2 \leq \tilde{c}_{\sigma,\max} < \infty,$$

where σ_j^2 and $\sigma_{j,0}^2$ are given in (4) and (8), respectively.

(B3) In a neighborhood of $\mathbf{B}_j^T \boldsymbol{\alpha}_j^0$, the conditional density of Y given $(X_j, \mathbf{X}_{\mathcal{S}_j})$, $f_{Y|(X_j,\mathbf{X}_{\mathcal{S}_j})}(y)$, is bounded on the support of $(X_j, \mathbf{X}_{\mathcal{S}_j})$ and uniformly in j.

(B4) $\min_{j\in\mathcal{M}_*} u_j \geq C_0 r_n n^{-\kappa}$ for some $C_0 > 0$ and $0 < \kappa < 1/2$.

Condition (B1) is imposed on the approximation error condition for nonparametric function in B-spline smoothing literature (e.g., [11,30,31]). Condition (B2) requires variances σ_j^2 and $\sigma_{j,0}^2$ to be uniformly bounded. Condition (B3) implies that there exists a finite constant $\bar{c}_f > 0$ such that for a small $\epsilon > 0$, $\sup_{|y-\mathbf{B}_j^T\boldsymbol{\alpha}_j^0|<\epsilon} f_{Y|(X_j,\mathbf{X}_{\mathcal{S}_j})}(y) \leq \bar{c}_f$ holds uniformly. Condition (B4) guarantees that the marginal signal of active components in model $\mathcal{M}_*$ does not vanish. These conditions are similar to those in [17].

Proposition 1. *Under conditions (B1)–(B3), there exists a positive constant M_{1*} such that*

$$u_j - u_j^* \leq M_{1*} r_n L_n^{-d},$$

In addition, if condition (B4) further holds, then

$$\min_{j\in\mathcal{M}_*} u_j^* \geq C_0 \xi r_n n^{-\kappa},$$

provided that $L_n^{-d} \leq C_0(1-\xi)n^{-\kappa}/M_1$ for some $\xi \in (0,1)$.

To establish the sure screening property, we make the following assumptions: .

(C1) $\{m_{kj}\}$ and $\{g_{kj}\}$ belong to a class of functions $\mathcal{F}$, whose rth derivatives $m_{kj}^{(r)}$ and $g_{kj}^{(r)}$ exist and are Lipschitz of order α,

$$\mathcal{F} = \{b(\cdot) : |b^{(r)}(s) - b^{(r)}(t)| \leq K|s-t|^{\alpha}\}, \text{ for } s,t \in [a,b]$$

for some positive constant K, where $[a,b]$ is the support of X_k, r is a non-negative integer and $\alpha \in (0,1]$ such that $d = r + \alpha > 0.5$.

(C2) The joint density of $\mathbf{X}$, $f_{\mathbf{X}}$ is bounded by two positive numbers b_{1f} and b_{2f} satisfying $b_{1f} \leq f_{\mathbf{X}} \leq b_{2f}$. The density of X_j, f_{X_j} is bounded away from zero and infinity uniformly in j, that is, there exist two positive constants c_{1f} and c_{2f} such that $c_{1f} \leq f_{X_j}(x) \leq c_{2f}$.

(C3) There exist two positive constants K_1 and K_2, such that $P(X_j > x|\mathbf{X}_{-j}) \leq K_1 \exp(-K_2^{-1}x)$ for every j.

(C4) The conditional density of Y given $\mathbf{X} = \mathbf{x}$, $f_{Y|\mathbf{X}=\mathbf{x}}(y)$, satisfies the Lipschitz condition of first order and $c_{3f} \leq f_{Y|\mathbf{X}=\mathbf{x}}(y) \leq c_{4f}$ for some positive constants c_{3f} and c_{4f} for any y in a neighborhood of $\mathbf{B}_j^T\boldsymbol{\alpha}_j^0$ for $1 \leq j \leq p$.

(C5) There exist some positive constants M_1 and M_2 such that $\sup_{i,j} |\mathbf{B}_{ij}^T\boldsymbol{\alpha}_j^0| \leq M_1 < \infty$, $\sup_{i,j} |\mathbf{B}_{ij}^T\boldsymbol{\theta}_j^0| \leq M_2 < \infty$. Furthermore, assume that $\min_{1\leq j\leq p} \sigma_j^2 \geq M_3 > 0$ for some constant M_3.

(C6) There exists some constant $\xi \in (0,1)$ such that $L_n^{-d} \leq C_0(1-\xi)n^{-\kappa}/M_{1*}$.

Condition (C1) is a smoothness assumption on $\{m_{kj}\}$ and $\{g_{kj}\}$ in nonparametric B-spline-related literature ([7,32]). Condition (C3) is a moment constraint on each of the predictors. Conditions (C2), (C4) and (C5) are similar to those imposed in [17]. Condition (C6) is assumed to ensure the marginal signal level of truly active variables not too weak after B-spline approximation. The above conditions are standard in variable screening literature (e.g., [17,28]).

According to the properties of normalized B-splines and under the conditions (C1) and (C2) (c.f., [33,34]), we can obtain the fact that for each $j = 1, \cdots, p$ and $k = 1, \ldots, L_n$, there exist positive constants C_1, C_2 and C_3 independent of j, k such that

$$C_1 L_n^{-1} \leq \lambda_{\min}(\mathrm{E}\{\mathbf{B}(X_j)\mathbf{B}(X_j)^T\}) \leq \lambda_{\max}(\mathrm{E}\{\mathbf{B}(X_j)\mathbf{B}(X_j)^T\}) \leq C_2 L_n^{-1}, \tag{9}$$

and

$$\mathrm{E}\{B_k^2(X_j)\} \leq C_3 L_n^{-1}. \tag{10}$$

The following lemma bounds the eigenvalues of the B-spline basis matrix from below and from above. This result extends Lemma 3 of [32] from a fixed dimension to a diverging dimension, which may be crucial to the independent interest of some readers.

Lemma 1. *Suppose that conditions (C1) and (C2) hold, then we have*

$$C_1 \left(\frac{1-\delta_0}{2}\right)^{|\mathcal{S}_j|-1} L_n^{-1} \leq \lambda_{\min}(\mathrm{E}\{\mathbf{B}_j \mathbf{B}_j^T\}) \leq \lambda_{\max}(\mathrm{E}\{\mathbf{B}_j \mathbf{B}_j^T\}) \leq C_2 |\mathcal{S}_j| L_n^{-1},$$

where $\delta_0 = (1 - b_{1f}^2 b_{2f}^{-2} \zeta)^{1/2}$ *for some constant* $0 < \zeta < 1$.

This result reveals that r_n plays an important role in bounding the eigenvalues of the B-spline basis matrix. When r_n goes to infinity rapidly, the minimum eigenvalue of the basis matrix will degrade to zero very quickly at an exponential rate. However, if the following result holds, then the divergence rate of r_n cannot achieve a polynomial order of n, but can be of an order of $\log n$.

Theorem 1. *Suppose that conditions (B1)–(B5) and (C1)–(C5) hold and assume that* $a_0^{-2r_n} L_n / n^{1-2\kappa} = o(1)$ *and* $a_0^{-2r_n} r_n^3 L_n n^{-\kappa} = o(1)$ *are satisfied.*

(i) *For any* $C > 0$, *then there exist some positive constants* c_6^*, c_{14}^* *such that, for* $0 < \kappa < 1/2$ *and sufficiently large* n,

$$\begin{aligned} &P\Big(\max_{1 \leq j \leq p_n} |\widehat{u}_j - u_j^*| \geq C r_n n^{-\kappa}\Big) \\ &\leq p_n\{7 \exp\big(-c_6^* a_0^{2r_n} r_n^2 n^{1-4\kappa}\big) + [116(r_n L_n)^2 + 60 r_n L_n + 10] \exp(-c_{14}^* a_0^{2r_n} L_n^{-3} n^{1-2\kappa})\}, \end{aligned}$$

where $a_0 = (1-\delta_0)/2$ *and* δ_0 *is given in Lemma 1.*

(ii) *In addition, if condition (C6) is further satisfied, by choosing* $\nu_n = \widetilde{C}_0 r_n n^{-\kappa}$ *with* $\widetilde{C}_0 \leq C_0 \xi/2$, *we have*

$$\begin{aligned} P\Big(\mathcal{M}_* \subset \widehat{\mathcal{M}}_{\nu_n}\Big) \quad &\geq 1 - s_n\{7 \exp\big(-c_6^* a_0^{2r_n} r_n^2 n^{1-4\kappa}\big) \\ &+ [116(r_n L_n)^2 + 60 r_n L_n + 10] \exp(-c_{14}^* a_0^{2r_n} L_n^{-3} n^{1-2\kappa})\} \end{aligned}$$

for sufficiently large n, *where* $s_n = |\mathcal{M}_*|$.

The above establishes the sure screening property that all the relevant variables can be recruited with probability going to one in the final model. The probability bound in the property is free of p_n, but depends on r_n and the number of basis functions L_n. Though this ensures that NQPC-SIS retains all important predictors with high probability, the noisy variables can be included by NQPC-SIS. Ideally, this can be realized by the choice of ν_n, according to Theorem 1 and by setting $\max_{j \notin \mathcal{M}_*} \left|\varrho^*_\tau(Y, X_j|\mathbf{X}_{\mathcal{S}_j})\right| = o(r_n n^{-\kappa})$, to achieve the selection consistency, i.e.,

$$P(\mathcal{M}_* = \widehat{\mathcal{M}}_{\nu_n}) \to 1$$

when n is sufficiently large. This property can also be achieved by Theorem 1 and by assuming that $\varrho^*_\tau(Y, X_j|\mathbf{X}_{\mathcal{S}_j}) = 0$ for $j \notin \mathcal{M}_*$. However, this would be too restrictive to check in practice. Similar to [17], we may assume that $\sum_{j=1}^{p} u_j^* = O(n^{\varsigma})$ for some $\varsigma > 0$ to control the false selection rate. With this condition, we can obtain the following property to control the size of the selected model.

Theorem 2. *Under the conditions of Theorem 1 and by choosing $\nu_n = \widetilde{C}_0 r_n n^{-\kappa}$ with $\widetilde{C}_0 \le C_0\xi/2$ and if $\sum_{j=1}^{p} u_j^* = O(n^{\varsigma})$ for some $\varsigma > 0$, then for some positive constant C_*, there exist some constants $\tilde{c}_6^*, \tilde{c}_{14}^*$ such that*

$$\begin{aligned} P(|\widehat{\mathcal{M}}_{\nu_n}| \le C_* r_n^{-1} n^{\kappa+\varsigma}) \quad &\ge 1 - p_n\{7\exp\left(-\tilde{c}_6^* a_0^{2r_n} r_n^2 n^{1-4\kappa}\right) \\ &\quad + [116(r_n L_n)^2 + 60 r_n L_n + 10]\exp(-\tilde{c}_{14}^* a_0^{2r_n} L_n^{-3} n^{1-2\kappa})\} \end{aligned}$$

for sufficiently large n.

This theorem reveals that after an application of the NQPC-SIS, the dimensionality can be reduced from an exponential order to a polynomial size of n at the same time retaining all the important predictors with probability approaching one.

4. Algorithm for NQPC-SIS

To make the NQPS-SIS practically applicable, for each X_J, we need to specify the conditional set $\mathcal{S}_j$. We note that a sequential test was developed to identify $\mathcal{S}_j$ in [17] via an application of the Fisher's Z-transformation [35] and partial correlation. In this section, we provide a two-stage procedure based on nonparametric additive quantile regression model, which can be viewed as a complementary to [17].

To reduce the computational burden, we first apply the quantile-adaptive model-free feature screening (Qa-SIS) proposed by [13] to select a subset from $\{X_j, 1 \le j \le p_n\}$, denoted by $\widehat{\mathcal{M}}_{\text{Qa-SIS}}$ with $|\widehat{\mathcal{M}}_{\text{Qa-SIS}}| = \lfloor 0.5 n L_n^{-1}/\log(n L_n^{-1}) \rfloor + 1$, where L_n is the number of basis functions used in Qa-SIS and $\lfloor a \rfloor$ denotes the largest integer not exceeding a. Second, for each X_j, if $X_j \in \widehat{\mathcal{M}}_{\text{Qa-SIS}}$, we set $\mathcal{C}_j = \{X_k | X_k \in \widehat{\mathcal{M}}_{\text{Qa-SIS}}, k \ne j\}$, otherwise $\mathcal{C}_j = \{X_k | X_k \in \widehat{\mathcal{M}}_{\text{Qa-SIS}}, k \ne |\widehat{\mathcal{M}}_{\text{Qa-SIS}}|\}$. Thus, $|\mathcal{C}_j| = \lfloor 0.5 n L_n^{-1}/\log(n L_n^{-1}) \rfloor$. Third, we carry out a variable selection with SCAD penalty [2] based on additive quantile regression model for data set $\{(X_{ij}, X_{i\mathcal{C}_j}), i = 1, \cdots, n\}$ and then a small reduced subset is obtained, denoted by $\mathcal{C}_j^v$. Such a two-stage procedure can help to find the conditional subset for the jth variable and will be incorporated in the following algorithm. With a slight abuse of notation, we use d_n to denote the screening threshold parameter of the NQPC-SIS, in other words, for the NQPC-SIS, we select d_n covariates that correspond to the first d_n largest NQPCs.

Algorithm 1 has the same spirit as the QPCS algorithm of [17], who demonstrated empirically that the QPCS algorithm outperforms their QTCS and QFR algorithms. In the implementation, we choose $d_n^* = \lfloor 0.5 n L_n^{-1}/\log(n L_n^{-1}) \rfloor$ and $d_n = \lfloor n/\log n \rfloor$, which does not exclude other choice. According to our limited simulation experience, this choice works satisfactorily. The values of d_n^* and r_n we take on cannot be too large, due to the use of B-spline basis approximations. Theoretically, we need to specify d_n^* such that $d_n^* \le r_n$, while it is sufficient to require $L_n d_n^* < n$ practically.

Algorithm 1 **The implementation of NQPC-SIS.**

1: Given d_n, we set a pre-specified number $d_n^* \leq d_n$ and an initial set $\mathcal{A}^{(0)} = \emptyset$.
2: For $k = 1, \ldots, d_n^*$,
(2a) update $\mathcal{S}_j = \mathcal{A}^{(k-1)} \cup \mathcal{C}_j^v$;
(2b) update $\mathcal{A}^{(k)} = \mathcal{A}^{(k-1)} \cup \{j^*\}$, where the variable index j^* is defined by

$$j^* = \mathrm{argmax}_{j \notin \mathcal{A}^{(k-1)}} |\widehat{\rho}_\tau(Y, X_j | X_{\mathcal{S}_j})|.$$

3: For $k = d_n^* + 1, \ldots, d_n$,
(3a) update $\mathcal{S}_j = \mathcal{A}^{(d_n^*)} \cup \mathcal{C}_j^v$;
(3b) update $\mathcal{A}^{(k)} = \mathcal{A}^{(k-1)} \cup \{j^*\}$, where the variable index j^* is such that

$$j^* = \mathrm{argmax}_{j \notin \mathcal{A}^{(k-1)}} |\widehat{\rho}_\tau(Y, X_j | X_{\mathcal{S}_j})|.$$

4: Repeat Step 3 until $k \geq d_n$. The final selected set is denoted as $\hat{\mathcal{M}}$.

5. Numerical Studies

5.1. Simulations

In this subsection, we conduct some simulation studies to examine the finite sample performance of the proposed NQPC-SIS. In order to evaluate the performance, we employ three criteria: the minimum model size (MMS), i.e., the smallest number of covariates that contain all the active variables, its robust standard deviation (RSD), and the proportion of all the active variables selected ($\mathcal{P}$) with the screening threshold parameter being specified as $d_n = \lfloor n/\log n \rfloor$. Throughout this subsection, we adopt the following simulation settings: the sample size $n = 200$, the number of basis $L_n = \lfloor n^{1/5} \rfloor + 1$, and the dimensionality $p_n = 1000$. We simulate the random error ε from two distributions: $N(0,1)$ and $t(3)$, respectively. Three quantile levels $\tau = 0.2, 0.5, 0.8$ are considered in all situations. For each simulation scenario, all the results are obtained over $N = 200$ replications.

Example 1. *Let $\mathbf{X} = (X_1, \cdots, X_{p_n})^T$ be a p_n-dimensional random vector having a multivariate normal distribution with mean zero and covariance matrix $\mathbf{\Sigma} = (\sigma_{jk})_{1 \leq j,k \leq p_n}$, where $\sigma_{jj} = 1$ and $\sigma_{j,k} = \rho, j \neq k$ except that $\sigma_{i4} = \sigma_{4j} = \sqrt{\rho}$. Generate the response as:*

$$Y = \beta X_1 + \beta X_2 + \beta X_3 - 3\beta\sqrt{\rho}X_4 + \varepsilon.$$

It is easily observed that the marginal Pearson's correlation between X_4 and Y is zero. We take $\rho = 0.5, 0.8$ and set $\beta = 2.5(1 + |\tau - 0.5|)$ to incorporate the quantile information.

Example 2. *We follow the simulation model of [17] and generate the response as*

$$Y = \beta X_1 + \beta X_2 + \beta X_3 - 3\beta\sqrt{\rho}X_4 - 0.25\beta X_5 + \varepsilon,$$

where β, ρ, and $\mathbf{X}$ are defined as in Example 1 except that $\sigma_{i5} = \sigma_{5j} = 0$ such that X_5 is uncorrelated with $X_j, j \neq 5$.

Table 1. Simulation results for Example 1 when n = 200.

ε	ρ	Method	s_n	$\tau = 0.2$ MMS(RSD)	$\mathcal{P}$	$\tau = 0.5$ MMS(RSD)	$\mathcal{P}$	$\tau = 0.8$ MMS(RSD)	$\mathcal{P}$
$N(0,1)$	0.5	SIS	4	455.5(319.3)	0	437(330.3)	0	434.5(372)	0
		NIS	4	451(456.5)	0.025	506(421.3)	0	486.5(390.5)	0
		Qa-SIS	4	466(392.5)	0.02	466.5(375.5)	0.01	490.5(382.3)	0.01
		QPC-SIS	4	4(0)	1	4(0)	1	4(0)	1
		NQPC-SIS	4	4(0)	0.995	4(0)	1	4(0)	1
	0.8	SIS	4	444.5(141.3)	0	458(161.8)	0	452.5(188)	0
		NIS	4	489.5(274.5)	0	518.5(274)	0	511(285.8)	0
		Qa-SIS	4	522(372.3)	0.01	510.5(358)	0	560.5(292.8)	0
		QPC-SIS	4	5(2)	0.99	4(1)	1	5(2)	0.98
		NQPC-SIS	4	6(3)	0.96	4(2)	0.99	6(3)	0.96
$t(3)$	0.5	SIS	4	434.5(352.8)	0.005	475(343.3)	0	472.5(366)	0
		NIS	4	492.5(347.5)	0.01	501.5(415)	0	555.5(352.3)	0
		Qa-SIS	4	510.5(390.3)	0.005	481(463.3)	0.015	541.5(460.3)	0.01
		QPC-SIS	4	4(0)	1	4(0)	1	4(0)	1
		NQPC-SIS	4	4(0)	0.995	4(0)	1	4(0)	1
	0.8	SIS	4	453(135.8)	0	468(200.5)	0	473(283.5)	0
		NIS	4	535.5(288.3)	0	507(253.3)	0	507.5(368.3)	0
		Qa-SIS	4	597.5(329.3)	0	578.5(374)	0	591.5(366.8)	0.005
		QPC-SIS	4	6(3)	0.915	5(2)	0.975	6(2)	0.945
		NQPC-SIS	4	6.5(6)	0.84	5(3)	0.955	6(5.3)	0.855

Table 2. Simulation results for Example 2 when $n = 200$.

ε	ρ	Method	s_n	$\tau = 0.2$ MMS(RSD)	$\mathcal{P}$	$\tau = 0.5$ MMS(RSD)	$\mathcal{P}$	$\tau = 0.8$ MMS(RSD)	$\mathcal{P}$
$N(0,1)$	0.5	SIS	5	439.5(359.5)	0	477(319)	0	427(324.8)	0
		NIS	5	522(362)	0.005	566(429.8)	0	507.5(392)	0
		Qa-SIS	5	542.5(400.3)	0	565.5(351.5)	0	554(340.3)	0
		QPC-SIS	5	5(0)	1	5(0)	1	5(0)	1
		NQPC-SIS	5	5(0)	1	5(0)	1	5(0)	1
	0.8	SIS	5	436(111.3)	0	479.5(232.8)	0	466.5(219.3)	0
		NIS	5	523.5(246.8)	0	556.5(265.8)	0	527(286.3)	0
		Qa-SIS	5	557.5(376.5)	0	604(358.8)	0	542(363.8)	0
		QPC-SIS	5	6(2)	0.97	6(2)	0.97	7(3)	0.93
		NQPC-SIS	5	7(2)	0.9	6(2)	0.945	7(3)	0.9
$t(3)$	0.5	SIS	5	478.5(347)	0	451.5(308.8)	0	483.5(322.8)	0
		NIS	5	535.5(384.3)	0	545.5(317.8)	0.005	508(353)	0
		Qa-SIS	5	597.5(389.5)	0.005	568.5(341.5)	0	593.5(435.8)	0.005
		QPC-SIS	5	5(0)	0.99	5(0)	1	5(0)	0.995
		NQPC-SIS	5	5(1)	0.985	5(0)	0.995	5(1)	0.975
	0.8	SIS	5	468(286.3)	0	477(238.5)	0	466(229)	0
		NIS	5	530.5(324.8)	0	532.5(321.3)	0	525.5(245.5)	0
		Qa-SIS	5	655(391.8)	0	590.5(361.5)	0	591.5(374.8)	0
		QPC-SIS	5	7(21.3)	0.765	7(3)	0.88	8(44.8)	0.74
		NQPC-SIS	5	8(81.8)	0.63	7(7.3)	0.82	11(136.5)	0.57

Table 3. Simulation results for Example 3 when $n = 200$.

ε	τ	Method	s_n	$t=1$		$t=2$	
				MMS(RSD)	$\mathcal{P}$	MMS(RSD)	$\mathcal{P}$
$N(0,1)$	$\tau = 0.2$	Qa-SIS	5	695(461.5)	0.005	799.5(326.5)	0
		QPC-SIS	5	46(165.3)	0.485	298.5(477.5)	0.075
		NQPC-SIS	5	7(13)	0.82	160(345.8)	0.25
	$\tau = 0.5$	SIS	5	492(805)	0.155	1000(1)	0
		NIS	5	533(608)	0.035	798(391)	0.005
		Qa-SIS	5	761.5(460)	0.005	762.5(353)	0
		QPC-SIS	5	8.5(33)	0.75	211(372.8)	0.145
		NQPC-SIS	5	5(0)	0.96	29(103.8)	0.56
	$\tau = 0.8$	Qa-SIS	5	599(508)	0.01	708(348.3)	0
		QPC-SIS	5	47(213)	0.46	393(400)	0.015
		NQPC-SIS	5	6(7)	0.86	156.5(388)	0.23
$t(3)$	$\tau = 0.2$	Qa-SIS	5	689.5(434)	0	794(345)	0
		QPC-SIS	5	85.5(239.5)	0.28	548.5(518.5)	0.025
		NQPC-SIS	5	46.5(181)	0.46	487(551)	0.055
	$\tau = 0.5$	SIS	5	626.5(767.8)	0.1	999(6)	0
		NIS	5	560.5(574.5)	0.025	742(430.3)	0
		Qa-SIS	5	673.5(449.3)	0.005	751.5(358.5)	0
		QPC-SIS	5	21.5(89.3)	0.56	331.5(467.8)	0.06
		NQPC-SIS	5	6(5)	0.855	136.5(388)	0.21
	$\tau = 0.8$	Qa-SIS	5	583(458.5)	0.015	711.5(382.3)	0
		QPC-SIS	5	108.5(303)	0.3	623(448)	0.005
		NQPC-SIS	5	28.5(136.3)	0.52	413(489.5)	0.03

Table 4. Simulation results for Example 4 when $n = 200$.

ε	τ	Method	s_n	$t=1$		$t=2$	
				MMS(RSD)	$\mathcal{P}$	MMS(RSD)	$\mathcal{P}$
$N(0,1)$	$\tau = 0.2$	Qa-SIS	5	735(448.5)	0.005	681.5(388.5)	0
		QPC-SIS	5	647.5(425)	0.005	746.5(329.3)	0
		NQPC-SIS	5	5(1)	0.945	86(334)	0.385
	$\tau = 0.5$	SIS	5	793(268.3)	0	846.5(252.8)	0
		NIS	5	765.5(298.8)	0	896.5(225.8)	0
		Qa-SIS	5	749.5(274.8)	0	818(301.5)	0
		QPC-SIS	5	717.5(326.8)	0	805.5(254.3)	0
		NQPC-SIS	5	5(0)	1	8(60.3)	0.705
	$\tau = 0.8$	Qa-SIS	5	836(274)	0	867.5(243.8)	0
		QPC-SIS	5	798.5(248.3)	0	811.5(249.3)	0
		NQPC-SIS	5	5(1)	0.985	61(355.8)	0.44
$t(3)$	$\tau = 0.2$	Qa-SIS	5	716.5(374.3)	0	703(375.3)	0
		QPC-SIS	5	603(422)	0.01	743.5(295.3)	0
		NQPC-SIS	5	7(22.5)	0.78	317(592.3)	0.15
	$\tau = 0.5$	SIS	5	786.5(261.8)	0	869.5(301.3)	0
		NIS	5	779(285.5)	0	833.5(259.8)	0
		Qa-SIS	5	754.5(324.5)	0	800(255.8)	0
		QPC-SIS	5	755.5(379.5)	0	825(296.3)	0
		NQPC-SIS	5	5(0)	0.99	61.5(302.3)	0.435
	$\tau = 0.8$	Qa-SIS	5	819.5(255.8)	0	869(241.3)	0
		QPC-SIS	5	795(249.3)	0	847(260.3)	0
		NQPC-SIS	5	6(9)	0.835	375(576)	0.14

Example 3. *We simulate the response from the following nonlinear model:*

$$Y = 3g_1(X_1) + 3g_2(X_2) + 3g_3(X_3) + 3g_4(X_4) + 3g_5(X_5) + \varepsilon,$$

where $g_1(x) = 1.5x, g_2(x) = 2x(2x-1), g_3(x) = \sin(2\pi x)/(2\sin(2\pi x)), g_4(x) = \sin(2\pi x)$, $g_5(x) = e^{x-0.5}$. *The covariates* $\mathbf{X} = (X_1, \cdots, X_{p_n})$ *are simulated from a random-effects model* $X_j = \frac{W_j+tU}{1+t}, j = 1, \cdots, p_n$, *where* W_j*s and* U *are iid* $Unif(0,1)$. *We consider two cases of* $t = 1$ *and* $t = 2$, *corresponding to* $\mathrm{corr}(X_j, X_k) = 0.5$ *and* 0.8 *for* $j \neq k$, *respectively.*

Example 4. *We consider the same model as that in Example 3, with exception that* X_2 *and* X_5 *are replaced by* $X_2 = \cos(2\pi X_6) + \epsilon$ *and* $X_5 = (X_1 - 0.5)^2 + \epsilon$, *where* $\epsilon \sim N(0,1)$ *and is independent of* ε, *the error in the model in Example 3.*

Table 5. Simulation results for Examples 1 to 4 when $\rho = 0.9$ and $\tau = 0.5$.

		$n = 200$				$n = 400$			
		$\varepsilon \sim N(0,1)$		$\varepsilon \sim t(3)$		$\varepsilon \sim N(0,1)$		$\varepsilon \sim t(3)$	
	Method	**MMS(RSD)**	$\mathcal{P}$	**MMS(RSD)**	$\mathcal{P}$	**MMS(RSD)**	$\mathcal{P}$	**MMS(RSD)**	$\mathcal{P}$
Example 1	QPC-SIS	6(2)	0.94	9(64)	0.665	5(2)	1	5(2)	0.985
	NQPC-SIS	6(3)	0.88	30.5(174.3)	0.52	5(2)	0.99	6(2)	0.93
Example 2	QPC-SIS	7(2)	0.9	25.5(204.5)	0.525	6(2)	1	7(2)	0.985
	NQPC-SIS	8(9.25)	0.825	44.5(178.8)	0.465	6(2)	0.995	7(2)	0.95
Example 3	QPC-SIS	608.5(471.5)	0	687.5(400)	0.005	253(408.3)	0.175	399.5(521)	0.105
	NQPC-SIS	341(505.8)	0.09	527.5(496.5)	0.025	19(75)	0.705	69.5(247.8)	0.475
Example 4	QPC-SIS	770.5(311.5)	0	802.5(232.8)	0	752.5(312)	0	784(258.3)	0
	NQPC-SIS	301.5(484.3)	0.11	433(554.8)	0.075	11(44)	0.76	38.5(142.8)	0.58

The simulation results of Examples 1–4 are shown in Tables 1–4, respectively. The results in Table 1 show that when the true relation between the response and covariates in the model is linear, the SIS, NIS and Qa-SIS methods fail to work. However, when comparing to those methods, we can see that both QPC-SIS and NQPC-SIS with $\tau = 0.5$ work reasonably well, although the QPC-SIS slightly outperform our NQPC-SIS when $\rho = 0.8$. This is expected because the QPC works for the model with linear relationship between the covariates. A similar observation can be drawn in Table 2 for Example 2, which is also a linear model, albeit the difference that X_5 and $X_j, j \neq 5$ are independent in Example 2. The results in Table 3 indicate that when the relationship between Y and $\mathbf{X}$ is nonlinear and the relationship between covariates is linear, our proposed NQPC-SIS performs best and then followed by QPC-SIS. From Table 4, we can see that when the relationship between Y and $\mathbf{X}$ is nonlinear and there also exists a nonlinear relationship among $\mathbf{X}$, NQPC-SIS works most satisfactorily and is much better than Qa-SIS and QPC-SIS in terms of both MMS and selection rate $\mathcal{P}$.

In addition, the simulation results of QPC-SIS and NQPC-SIS for Examples 1–4 with $\rho = 0.9$ and $\tau = 0.5$ are reported in Table 5. It can be observed from Table 5 that when the sample size increases from 200 to 400, the performance of QPC-SIS and NQPC-SIS are improved by much, although QPC-SIS and NQPC-SIS perform very competitively in Examples 1 and 2, while NQPC-SIS performs significantly better than QPC-SIS in Examples 3 and 4. These evidences indicate the effectiveness and usefulness of our NQPC-SIS.

As suggested by one anonymous reviewer, we add one more simulation to compare our NQPC-SIS with the following two approaches: (a) QC-SIS, which is the screening method based on quantile correlation, but simply ignores the effect of conditional variables on the response, and (b) RFQPC-SIS, which is a procedure very similarly to our NQPC-SIS, yet removes the effect of conditional variables through fitting Random Forest models. We examine the performance of these three approaches under $\tau = 0.5$ and $n = 200$ for Examples 1 to 4, where RFQPC-SIS is a variant of the NQPC method and implemented with *randomForest* in R package "`randomForest`". Note that RFQPC-SIS requires $2(p - |\mathcal{A}^{(k)}|)$

random forest regressions in the k-th iteration, which is highly computationally intensive. Here, we evaluate NQPC-SIS, QC-SIS and RFQPC-SIS using effective model size (EMS) and $\mathcal{P}$, where EMS indicates the average of true variables contained in the first $d_n = \lfloor n/\log(n) \rfloor$ variables selected from 200 replicate experiments. The results are reported in Table 6, showing that our NQPC-SIS still performs the best and is followed by RFQPC-SIS. Moreover, the computational cost of NQPC-SIS is much less than that of RFQPC-SIS.

Table 6. Simulation results for Examples 1 to 4 when $n = 200$ and $\tau = 0.5$.

		$\rho = 0.5$				$\rho = 0.8$			
		$\varepsilon \sim N(0,1)$		$\varepsilon \sim t(3)$		$\varepsilon \sim N(0,1)$		$\varepsilon \sim t(3)$	
	Method	**EMS(SD)**	$\mathcal{P}$	**EMS(RSD)**	$\mathcal{P}$	**EMS(SD)**	$\mathcal{P}$	**EMS(SD)**	$\mathcal{P}$
Example 1	QC-SIS	2.985(0.122)	0	2.925(0.346)	0	2.570(0.969)	0	2.340(1.162)	0
	RFQPC-SIS	3.995(0.071)	0.995	3.975(0.157)	0.975	3.675(0.470)	0.675	3.460(0.500)	0.46
	NQPC-SIS	4(0)	1	4(0)	1	3.985(0.122)	0.985	3.930(0.256)	0.93
Example 2	QC-SIS	3.565(0.646)	0	3.490(0.657)	0	3.355(1.147)	0	3.255(1.280)	0
	RFQPC-SIS	4.745(0.437)	0.745	4.630(0.504)	0.64	4.475(0.609)	0.535	4.215(0.649)	0.335
	NQPC-SIS	5(0)	1	5(0)	1	4.960(0.221)	0.965	4.785(0.447)	0.8
Example 3	QC-SIS	2.470(0.862)	0.03	2.575(0.894)	0	1.505(0.567)	0	1.460(0.592)	0
	RFQPC-SIS	4.945(0.229)	0.945	4.730(0.788)	0.75	4.275(0.808)	0.465	3.535(0.175)	0.175
	NQPC-SIS	4.955(0.208)	0.955	4.785(0.424)	0.79	4.320(0.825)	0.49	3.715(1.109)	0.265
Example 4	QC-SIS	1.775(0.613)	0	1.790(0.720)	0	1.685(0.639)	0	1.690(0.629)	0
	RFQPC-SIS	3.045(0.739)	0.02	3.040(0.788)	0.045	2.455(0.616)	0	2.350(0.528)	0
	NQPC-SIS	5(0)	1	4.980(0.140)	0.98	4.590(0.731)	0.715	4.205(0.864)	0.465

5.2. An Application to Breast Cancer Data

In this subsection, we apply the proposed NQPC-SIS to breast cancer data with a high lethality rate, which is reported by [36]. The data consists of 19,672 gene expression and 2,149 CGH measurements from 89 cancer patient samples, which is available at https://github.com/bnaras/PMA/blob/master/data/breastdata.rda (accessed on 18 June 2021). Our interest here is to detect the genes that have the most impact on comparative genomic hybridization (CGH) measurements. A similar purpose was achieved in [25,37]. Following [37], we consider the first principal component of 136 CGH measurements as the response Y and the remaining 18,672 gene probes as the explanatory variables $\mathbf{X}$. We implement the two stage procedure for the sake of comparison, where a variable screening method is implemented in the first stage and a predictive regression model is conducted in the second stage. To this end, we select $d_n = \lfloor n/\log(n) \rfloor$ variables in the first stage using one of the screening methods: SIS, NIS, Qa-SIS, QPC-SIS and NQPC-SIS, as mentioned in the simulation study. In the second stage, we randomly select 80% sample data as the training set, and the remaining 20% sample as the test set. Then, we apply one machine learning method, regression tree, to the dimension-reduced data to examine the finite sample performance on the test set. We use the command *M5P* in R package "`RWeka`" for implementing the regression tree method. We use the mean of absolute prediction error (MAPE), defined as

$$\text{MAPE} = \frac{1}{n^{(test)}} \sum_{i=1}^{n^{(test)}} \left| Y_i^{(test)} - \hat{Y}_i^{(test)} \right|,$$

as our evaluation index, where $n^{(test)}$ is the number of observations in the training set and $\hat{Y}_i^{(test)}$ is the predicted value of Y at the observation x_i in the test set. We repeat the above procedure 500 times and report the mean and standard deviation of 500 MAPEs in Table 7. According to the results in Table 7, we can observe that the NQPC-SIS outperforms both the SIS, NIS and Qa-SIS. Typically, our NQPC-SIS produces the lowest prediction error (MAPE) among these methods when $\tau = 0.4$, $\tau = 0.5$ and $\tau = 0.7$. Moreover, we also note that the QPC-SIS performs better than our NQPC-SIS at $\tau = 0.3$ and $\tau = 0.6$, but worse than our method at other three quantile levels. Qa-SIS performs worst among these methods. This evidence supports that the proposed NPQC-SIS in this paper works well for this real data.

Table 7. Prediction results for the real data on the test set, where the standard deviation is given in the parenthesis.

Method	$\tau = 0.3$	$\tau = 0.4$	$\tau = 0.5$	$\tau = 0.6$	$\tau = 0.7$
SIS	-	-	0.8202(0.1362)	-	-
NIS	-	-	0.8254(0.1348)	-	-
Qa-SIS	0.8318(0.1472)	0.8261(0.1446)	0.8375(0.1448)	0.8612(0.1541)	0.8431(0.1512)
QPC-SIS	0.7269(0.1267)	0.7989(0.1458)	0.8347(0.1471)	0.6495(0.1155)	1.0240(0.1732)
NQPC-SIS	0.7629(0.1356)	0.6488(0.1232)	0.6742(0.1285)	0.8156(0.1643)	0.7802(0.1315)

6. Concluding Remarks

In this paper, we proposed a nonparametric quantile partial correlation-based variable screening approach (NQPC-SIS), which can be viewed as an extension of the QPC-SIS proposed in [17] from a parametric framework to the nonparametric situation. Our proposed NQPC-SIS enjoys the sure independence screening property under some mild technical conditions. Furthermore, an algorithm of NQPC-SIS for implementation is provided for users. Extensive numerical experiments including simulations and real-world data analysis are carried out for illustration. The numerical results showed that our NQPC-SIS works fairly well especially when the relationship between variables is highly nonlinear.

Author Contributions: All the authors contributed to formulating the research idea, methodology, theory, algorithm design, result analysis, writing and reviewing the research. Conceptualization, X.X. and H.M.; methodology, X.X.; software, H.M.; validation, H.M.; formal analysis, H.M.; investigation, X.X.; writing—original draft preparation, H.M.; writing—review and editing, X.X.; supervision, X.X. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Fundamental Research Funds for the Central Universities (Grant No. 2021CDJQY-047) and National Natural Science Foundation of China (Grant No. 11801202).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

Appendix A. Technical Proofs

Proof of Proposition 1. First, recalling definitions of u_j and u_j^*, we can make a simple algebra decomposition:

$$\begin{aligned}
&\sqrt{\tau(1-\tau)}[\varrho_\tau(Y, X_j|\mathbf{X}_{\mathcal{S}_j}) - \varrho_\tau^*(Y, X_j|\mathbf{X}_{\mathcal{S}_j})] \\
&= (\sigma_{j,0}\sigma_j)^{-1}(\sigma_j - \sigma_{j,0})\mathrm{E}\{\psi_\tau(Y - m_j^0(\mathbf{X}_{\mathcal{S}_j}))(X_j - g_j^0(\mathbf{X}_{\mathcal{S}_j}))\} \\
&\quad + \sigma_j^{-1}\mathrm{E}\{[\psi_\tau(Y - m_j^0(\mathbf{X}_{\mathcal{S}_j})) - \psi_\tau(Y - \mathbf{B}_j^T\boldsymbol{\alpha}_j^0)](X_j - g_j^0(\mathbf{X}_{\mathcal{S}_j}))\} \\
&\quad - \sigma_j^{-1}\mathrm{E}\{\psi_\tau(Y - \mathbf{B}_j^T\boldsymbol{\alpha}_j^0)[g_j^0(\mathbf{X}_{\mathcal{S}_j}) - \mathbf{B}_j^T\boldsymbol{\theta}_j^0]\} \\
&\triangleq A_1 + A_2 + A_3 \text{ (say)}.
\end{aligned} \tag{A1}$$

Due to condition (B1), we can observe

$$\begin{aligned}
\sigma_j^2 - \sigma_{j,0}^2 &= \mathrm{E}\{(X_j - \mathbf{B}_j^T\boldsymbol{\theta}_j^0)^2\} - \mathrm{E}\{(X_j - g_j^0(\mathbf{X}_{\mathcal{S}_j}))^2\} \\
&= \mathrm{E}\{(g_j^0(\mathbf{X}_{\mathcal{S}_j}) - \mathbf{B}_j^T\boldsymbol{\theta}_j^0)^2\} + 2\mathrm{E}\{(g_j^0(\mathbf{X}_{\mathcal{S}_j}) - \mathbf{B}_j^T\boldsymbol{\theta}_j^0)(X_j - g_j^0(\mathbf{X}_{\mathcal{S}_j}))\} \\
&= \mathrm{E}\{(g_j^0(\mathbf{X}_{\mathcal{S}_j}) - \mathbf{B}_j^T\boldsymbol{\theta}_j^0)^2\},
\end{aligned}$$

where the cross product is zero due to $\mathrm{E}\{X_j - g_j^0(\mathbf{X}_{\mathcal{S}_j})|\mathbf{X}_{\mathcal{S}_j}\} = 0$ by condition (B1). This, in conjunction with condition (B1) and the basic inequality that $\sqrt{a} - \sqrt{b} \le \sqrt{a-b}$ for $a > b > 0$, gives

$$\sigma_j - \sigma_{j,0} \le \left[\mathrm{E}\{\sum_{k\in\mathcal{S}_j}(g_{jk}^0(X_k) - \mathbf{B}_j(X_k)^T\boldsymbol{\theta}_{jk}^0)\}^2\right]^{1/2} \le C_g r_n L_n^{-d}. \tag{A2}$$

Using Cauchy–Schwarz inequality, (A2) and the fact that $|\psi_\tau(u)| \le \max(\tau, 1-\tau) \le 1$, we have

$$\begin{aligned} |A_1| &\le (\sigma_{j,0}\sigma_j)^{-1}(\sigma_j - \sigma_{j,0})\left[\mathrm{E}\{|\psi_\tau(Y - m_j^0(\mathbf{X}_{\mathcal{S}_j}))|^2\}\right]^{1/2}\left[\mathrm{E}\{|(X_j - g_j^0(\mathbf{X}_{\mathcal{S}_j}))|^2\}\right]^{1/2} \\ &\le \sigma_j^{-1}(\sigma_j - \sigma_{j,0}) \le c_{\sigma,\min}^{-1/2} C_g r_n L_n^{-d}. \end{aligned} \tag{A3}$$

For A_2, we note that

$$\mathrm{E}\{[\psi_\tau(Y - m_j^0(\mathbf{X}_{\mathcal{S}_j})) - \psi_\tau(Y - \mathbf{B}_j^T\boldsymbol{\alpha}_j^0)](X_j - g_j^0(\mathbf{X}_{\mathcal{S}_j}))\} \triangleq \mathrm{E}\{(X_j - g_j^0(\mathbf{X}_{\mathcal{S}_j}))A_{21}\},$$

where, by Taylor's expansion,

$$\begin{aligned} A_{21} &= \mathrm{E}\{[\psi_\tau(Y - m_j^0(\mathbf{X}_{\mathcal{S}_j})) - \psi_\tau(Y - \mathbf{B}_j^T\boldsymbol{\alpha}_j^0)]|X_j, \mathbf{X}_{\mathcal{S}_j}\} \\ &= -f_{Y|(X_j,\mathbf{X}_{\mathcal{S}_j})}(y^*)(m_j^0(\mathbf{X}_{\mathcal{S}_j}) - \mathbf{B}_j^T\boldsymbol{\alpha}_j^0), \end{aligned}$$

where y^* is a number between $m_j^0(\mathbf{X}_{\mathcal{S}_j})$ and $\mathbf{B}_j^T\boldsymbol{\alpha}_j^0$. Hence, by condition (B1)–(B3) and Cauchy–Schwarz inequality, we can obtain

$$\begin{aligned} |A_2| &\le \sigma_j^{-1}\mathrm{E}\{|X_j - g_j^0(\mathbf{X}_{\mathcal{S}_j})| \cdot |A_{21}|\} \\ &\le \sigma_j^{-1}\bar{c}_f\mathrm{E}\{|m_j^0(\mathbf{X}_{\mathcal{S}_j}) - \mathbf{B}_j^T\boldsymbol{\alpha}_j^0| \cdot |X_j - g_j^0(\mathbf{X}_{\mathcal{S}_j})|\} \\ &\le \sigma_j^{-1}\bar{c}_f\{\mathrm{E}[|m_j^0(\mathbf{X}_{\mathcal{S}_j}) - \mathbf{B}_j^T\boldsymbol{\alpha}_j^0|^2]\}^{1/2}\{\mathrm{E}[|X_j - g_j^0(\mathbf{X}_{\mathcal{S}_j})|^2]\}^{1/2} \\ &\le \sigma_j^{-1}\sigma_{j,0}\bar{c}_f C_m r_n L_n^{-d} \le c_{\sigma,\min}^{-1/2}\bar{c}_{\sigma,\max}^{1/2}\bar{c}_f C_m r_n L_n^{-d} \end{aligned} \tag{A4}$$

for some constant $\bar{c}_f > 0$.

For A_3, by a similar argument, we can obtain

$$\begin{aligned} |A_3| &\le \sigma_j^{-1}\mathrm{E}\{|\psi_\tau(Y - \mathbf{B}_j^T\boldsymbol{\alpha}_j^0)| \cdot |g_j^0(\mathbf{X}_{\mathcal{S}_j}) - \mathbf{B}_j^T\boldsymbol{\theta}_j^0|\} \\ &\le \sigma_j^{-1}\{\mathrm{E}[|\psi_\tau(Y - \mathbf{B}_j^T\boldsymbol{\alpha}_j^0)|^2]\}^{1/2}\{\mathrm{E}[|g_j^0(\mathbf{X}_{\mathcal{S}_j}) - \mathbf{B}_j^T\boldsymbol{\theta}_j^0|^2]\}^{1/2} \\ &\le \sigma_j^{-1}C_g r_n L_n^{-d} \le c_{\sigma,\min}^{-1/2}C_g r_n L_n^{-d}. \end{aligned} \tag{A5}$$

Therefore, combining (A1) and the results in (A3)–(A5), we have

$$\begin{aligned} &|\varrho_\tau(Y, X_j|\mathbf{X}_{\mathcal{S}_j}) - \varrho_\tau^*(Y, X_j|\mathbf{X}_{\mathcal{S}_j})| \\ &\le [\tau(1-\tau)]^{-1/2}c_{\sigma,\min}^{-1/2}(2C_g + \bar{c}_{\sigma,\max}^{1/2}\bar{c}_f C_m)r_n L_n^{-d}. \end{aligned}$$

Using the basic inequality that $|a| - |b| \le \big||a| - |b|\big| \le |a - b|$, we can immediately conclude

$$u_j - u_j^* \le M_{1*} r_n L_n^{-d},$$

where $M_{1*} = [\tau(1-\tau)]^{-1/2}c_{\sigma,\min}^{-1/2}(2C_g + \bar{c}_{\sigma,\max}^{1/2}\bar{c}_f C_m)$. Thus, we complete the proof. □

Proof of Lemma 1. Without loss of generality, suppose that $\mathcal{S}_j = \{1, 2, \ldots, s\}$. Then, $\mathbf{B}_j = (\mathbf{B}(X_1)^T, \cdots, \mathbf{B}(X_s)^T)^T$. Let $\|\mathbf{a}\| = 1$, where $\mathbf{a} = (\mathbf{a}_1^T, \ldots, \mathbf{a}_s^T)^T$ with $\mathbf{a}_k \in \mathrm{R}^{L_n}$. On one hand, since $(\sum_{i=1}^n x_i)^2 \leq n \sum_{i=1}^n x_i^2$ by Cauchy–Schwarz inequality, we have

$$\mathbf{a}^T \mathrm{E}\{\mathbf{B}_j \mathbf{B}_j^T\}\mathbf{a} = \mathrm{E}\Big\{\Big[\sum_{k=1}^s \mathbf{a}_k^T \mathbf{B}(X_k)\Big]^2\Big\} \leq s \sum_{k=1}^s \mathbf{a}_k^T \mathrm{E}\{\mathbf{B}(X_k)\mathbf{B}(X_k)^T\}\mathbf{a}_k.$$

This together with the right hand side of (9) implies that

$$\lambda_{\max}(\mathrm{E}\{\mathbf{B}_j \mathbf{B}_j^T\}) \leq s\lambda_{\max}(\mathrm{E}\{\mathbf{B}(X_k)\mathbf{B}(X_k)^T\}) \leq C_2 s L_n^{-1}. \tag{A6}$$

On the other hand, an application of Lemma S.1 of [38] leads to

$$\begin{aligned}
\mathbf{a}^T \mathrm{E}\{\mathbf{B}_j \mathbf{B}_j^T\}\mathbf{a} &= \mathrm{E}\Big\{\Big[\sum_{k=1}^s \mathbf{a}_k^T \mathbf{B}(X_k)\Big]^2\Big\} \\
&\geq \Big(\frac{1-\delta_0}{2}\Big)^{s-1}\Big[\sum_{k=1}^s \sqrt{\mathrm{E}\{\mathbf{a}_k^T \mathbf{B}(X_k)\mathbf{B}(X_k)^T \mathbf{a}_k\}}\Big]^2 \\
&\geq \Big(\frac{1-\delta_0}{2}\Big)^{s-1} \lambda_{\min}(\mathrm{E}\{\mathbf{B}(X_k)\mathbf{B}(X_k)^T\})\Big[\sum_{k=1}^s \|\mathbf{a}_k\|\Big]^2,
\end{aligned}$$

where $\delta_0 = (1 - b_{1f}^2 b_{2f}^{-2}\zeta)^{1/2}$ for some positive constant $\zeta > 0$ and the last line uses the fact that $\mathbf{a}^T \mathbf{A}\mathbf{a} \geq \lambda_{\min}(\mathbf{A})$ for any $\|\mathbf{a}\| = 1$. It follows from the result on the left hand side of (9) that

$$\mathbf{a}^T \mathrm{E}\{\mathbf{B}_j \mathbf{B}_j^T\}\mathbf{a} \geq \Big(\frac{1-\delta_0}{2}\Big)^{s-1} C_1 L_n^{-1}\Big[\sum_{k=1}^s \|\mathbf{a}_k\|\Big]^2 \geq \Big(\frac{1-\delta_0}{2}\Big)^{s-1} C_1 L_n^{-1},$$

where the second inequality stems from $(\sum_{i=1}^n |x_i|)^2 \geq \sum_{i=1}^n x_i^2$ and $\|\mathbf{a}\| = 1$. This in turns implies that

$$\lambda_{\min}(\mathrm{E}\{\mathbf{B}_j \mathbf{B}_j^T\}) \geq \Big(\frac{1-\delta_0}{2}\Big)^{s-1} C_1 L_n^{-1}. \tag{A7}$$

Hence, combining (A6) and (A7) completes the proof of Lemma 1. □

Lemma A1. *Suppose that condition (C3) holds, then, for all $r \geq 2$,*

$$\mathrm{E}(|X_j|^r | \mathbf{X}_{-j}) \leq K_1 K_2^r r!$$

holds uniformly in j.

Lemma A1 is the same as Lemma 1 of [11]. From this, it is easily seen that $\mathrm{E}\{|X_j|^2 | \mathbf{X}_{-j}\}$ is finite and bounded by $2K_1 K_2^2$.

Lemma A2 (Bernstein's inequality, Lemma 2.2.11, [39])**.** *For independent random variables $Y_1, \ldots, Y_n$ with mean zero and $\mathrm{E}\{|Y_i|^r\} \leq r! K^{r-2} v_i/2$ for every $r \geq 2$, $i = 1, \ldots, n$ and some constants K, v_i. Then, for $x > 0$, we have*

$$P(|Y_1 + \cdots + Y_n| > x) \leq 2\exp\Big(-\frac{x^2}{2(v + Kx)}\Big),$$

for $v \geq \sum_{i=1}^n v_i$.

Lemma A3. *(Bernstein's inequality, Lemma 2.2.9, [39]) For independent random variables* $Y_1, \ldots, Y_n$ *with mean zero and bounded range* $[-M, M]$*, then*

$$P(|Y_1 + \cdots + Y_n| > x) \leq 2\exp\Big(-\frac{x^2}{2(v + Mx/3)}\Big),$$

for $v \geq \mathrm{var}(Y_i + \cdots + Y_n)$*.*

Lemma A4 (Symmetrization, Lemma 2.3.1, [39])**.** *Let* $Z_1, \ldots, Z_n$ *be independent random variables with values in* $\mathcal{Z}$ *and* $\mathcal{F}$ *is a class of real valued functions on* $\mathcal{Z}$*. Then,*

$$\mathrm{E}\Big\{\sup_{f\in\mathcal{F}} |(\mathbb{P}_n - \mathbb{P})f(Z)|\Big\} \leq 2\mathrm{E}\Big\{\sup_{f\in\mathcal{F}} |\mathbb{P}_n \varepsilon f(Z)|\Big\},$$

where $\varepsilon_1, \ldots, \varepsilon_n$ *is a Rademacher sequence (i.e., independent and identically distributed sequence taking values* ± 1 *with probability* $\frac{1}{2}$*) independent of* $Z_1, \ldots, Z_n$*, and* $\mathbb{P}f(Z) = \mathrm{E}f(Z)$ *and* $\mathbb{P}_n f(Z) = n^{-1}\sum_{i=1}^n f(Z_i)$*.*

Lemma A5 (Contraction theorem, [40])**.** *Let* $z_1, \ldots, z_n$ *be nonrandom elements of some space* $\mathcal{Z}$ *and let* $\mathcal{F}$ *be a class of real valued functions on* $\mathcal{Z}$*. Denote by* $\varepsilon_1, \ldots, \varepsilon_n$ *a Rademacher sequence. Consider Lipschitz functions* $g_i : \mathrm{R} \mapsto \mathrm{R}$*, that is*

$$|g_i(s_1) - g_i(s_2)| \leq |s_1 - s_2|, \forall s_1, s_2 \in \mathrm{R}.$$

Then, for any function $f_1 : \mathcal{Z} \mapsto \mathrm{R}$*, we have*

$$\mathrm{E}\Big\{\sup_{f\in\mathcal{F}} |\mathbb{P}_n \varepsilon(g(f) - g(f_1))|\Big\} \leq 2\mathrm{E}\Big\{\sup_{f\in\mathcal{F}} |\mathbb{P}_n \varepsilon(f - f_1)|\Big\}.$$

Lemma A6 (Concentration theorem, [41])**.** *Let* $Z_1, \ldots, Z_n$ *be independent random variables with values in* $\mathcal{Z}$ *and let* $g \in \mathcal{G}$*, a class of real valued functions on* $\mathcal{Z}$*. We assume that for some positive constants* l_i, g *and* $u_{i,g}$*,* $l_{i,g} \leq g(Z_i) \leq u_{i,g} \forall g \in \mathcal{G}$*. Define* $D^2 = \sup_{g\in\mathcal{G}} \sum_{i=1}^n (u_{i,g} - l_{i,g})^2/n$*, and* $\mathrm{U} = \sup_{g\in\mathcal{G}} \big|(\mathbb{P}_n - \mathbb{P})g(Z)\big|$*, then for any* $t > 0$*,*

$$P(\mathrm{U} \geq \mathrm{EU} + t) \leq \exp\Big(-\frac{nt^2}{2D^2}\Big).$$

Next, we need several lemmas to establish the consistency inequalities for $\widehat{\boldsymbol{\theta}}_j$ and $\widehat{\boldsymbol{\alpha}}_j$. Write $\mathbf{D}_{nj} = \frac{1}{n}\sum_{i=1}^n \mathbf{B}_{ij}\mathbf{B}_{ij}^T$, $\mathbf{D}_j = \mathrm{E}\{\mathbf{B}_{ij}\mathbf{B}_{ij}^T\} = \mathrm{E}\{\mathbf{B}_j\mathbf{B}_j^T\}$, $\mathbf{E}_{nj} = \frac{1}{n}\sum_{i=1}^n \mathbf{B}_{ij}X_{ij}$ and $\mathbf{E}_j = \mathrm{E}\{\mathbf{B}_{ij}X_{ij}\} = \mathrm{E}\{\mathbf{B}_j X_j\}$. Thus $\widehat{\boldsymbol{\theta}}_j = \mathbf{D}_{nj}^{-1}\mathbf{E}_{nj}$ and $\boldsymbol{\theta}_j^0 = \mathbf{D}_j^{-1}\mathbf{E}_j$.

Lemma A7. *Under conditions (C1) and (C2),*

(i) *there exists a constant* C_3 *such that for any* $\delta > 0$*,*

$$P\big(|\lambda_{\min}(\mathbf{D}_{nj}) - \lambda_{\min}(\mathbf{D}_j)| \geq r_n L_n \delta/n\big) \leq 2(r_n L_n)^2 \exp\Big(-\frac{\delta^2}{2(C_3 L_n^{-1} n + 2\delta/3)}\Big),$$
$$P\big(|\lambda_{\max}(\mathbf{D}_{nj} - \mathbf{D}_j)| \geq r_n L_n \delta/n\big) \leq 2(r_n L_n)^2 \exp\Big(-\frac{\delta^2}{2(C_3 L_n^{-1} n + 2\delta/3)}\Big),$$

(ii) *for some positive constant* c_1*, there exists some positive constant* c_2 *such that*

$$P\big(|\lambda_{\min}(\mathbf{D}_{nj})| \geq (1 + c_1)\lambda_{\min}(\mathbf{D}_j)\big) \leq 2(r_n L_n)^2 \exp\big(-c_2 a_0^{2r_n} r_n^{-2} L_n^{-3} n\big),$$

where $a_0 = (1 - \delta_0)/2$ *and* δ_0 *is defined in Lemma 1; and* *(iii)* *in addition, for any given constant c_2, there exists some positive constant c_3 such that*

$$P\big(\|\mathbf{D}_{nj}^{-1}\| \geq (1+c_3)\|\mathbf{D}_j^{-1}\|\big) \leq 2(r_nL_n)^2\exp\big(-c_2a_0^{2r_n}r_n^{-2}L_n^{-3}n\big).$$

Proof of Lemma A7. First, consider the proof of part (i). Denote $Q_{ij,s,t}^{(k,l)} = B_s(X_{ik})B_t(X_{il}) - \mathrm{E}\{B_s(X_{ik})B_t(X_{il})\}$ with $k,l \in S_j$ and $s,t = 1,\ldots,L_n$. Recalling that $\|B_t\|_\infty \leq 1$, we have $|Q_{ij,s,t}^{(k,l)}| \leq 2$ and $\mathrm{var}\{Q_{ij,s,t}^{(k,l)}\} \leq \mathrm{E}\{B_s^2(X_{ik})B_t^2(X_{il})\} \leq \mathrm{E}\{B_s^2(X_{ik})\} \leq C_3L_n^{-1}$ by the inequality (10). By Lemma A3, we have for any $\delta > 0$,

$$P\Big(\Big|n^{-1}\sum_{i=1}^n Q_{ij,s,t}^{(k,l)}\Big| > \frac{\delta}{n}\Big) \leq 2\exp\Big(-\frac{\delta^2}{2(C_3L_n^{-1}n + 2\delta/3)}\Big). \tag{A8}$$

Let $\mathbf{Q}_{nj} = \mathbf{D}_{nj} - \mathbf{D}_j$. It follows from Lemma 5 of [7] that $|\lambda_{\min}(\mathbf{D}_{nj}) - \lambda_{\min}(\mathbf{D}_j)| \leq \max\{|\lambda_{\min}(\mathbf{Q}_{nj})|, |\lambda_{\min}(-\mathbf{Q}_{nj})|\}$. Besides, it is easy to derive that for any $|S_j|L_n \times 1$ vector $\|\mathbf{a}\| = 1$, $|\mathbf{a}^T\mathbf{Q}_{nj}\mathbf{a}| \leq L_n|S_j| \cdot \|\mathbf{Q}_{nj}\|_\infty$, which implies that

$$|\lambda_{\min}(\mathbf{Q}_{nj})| \leq L_n|S_j| \cdot \|\mathbf{Q}_{nj}\|_\infty, \text{ and } |\lambda_{\max}(\mathbf{Q}_{nj})| \leq L_n|S_j| \cdot \|\mathbf{Q}_{nj}\|_\infty. \tag{A9}$$

This in conjunction with (A8) and the union bound of probability yields that

$$\begin{aligned} &P\big(\big|\lambda_{\min}(\mathbf{D}_{nj}) - \lambda_{\min}(\mathbf{D}_j)\big| \geq r_nL_n\delta/n\big) \\ &\leq P\big(\|\mathbf{Q}_{nj}\|_\infty \geq \delta/n\big) \leq 2(r_nL_n)^2\exp\Big(-\frac{\delta^2}{2(C_3L_n^{-1}n + 2\delta/3)}\Big) \end{aligned} \tag{A10}$$

and

$$P\big(\big|\lambda_{\max}(\mathbf{D}_{nj} - \mathbf{D}_j)\big| \geq r_nL_n\delta/n\big) \leq 2(r_nL_n)^2\exp\Big(-\frac{\delta^2}{2(C_3L_n^{-1}n + 2\delta/3)}\Big). \tag{A11}$$

Next, consider the proof of part (ii). Let $c_1^* = 2c_1C_1/(1-\delta_0)$, where $c_1 \in (0,1)$. Employing the result (A10) and taking $\delta = c_1^*a_0^{r_n}r_n^{-1}L_n^{-2}n$, we have

$$\begin{aligned} &P\big(\big|\lambda_{\min}(\mathbf{D}_{nj}) - \lambda_{\min}(\mathbf{D}_j)\big| \geq c_1\lambda_{\min}(\mathbf{D}_j)\big) \\ &\leq P\big(\big|\lambda_{\min}(\mathbf{D}_{nj}) - \lambda_{\min}(\mathbf{D}_j)\big| \geq c_1^*a_0^{r_n}L_n^{-1}\big) \\ &\leq 2(r_nL_n)^2\exp\Big(-\frac{c_1^{*2}a_0^{2r_n}r_n^{-2}L_n^{-4}n^2}{2(C_3L_n^{-1}n + 2c_1^*a_0^{r_n}r_n^{-1}L_n^{-2}n/3)}\Big) \\ &\leq 2(r_nL_n)^2\exp\big(-c_2a_0^{2r_n}r_n^{-2}L_n^{-3}n\big) \end{aligned} \tag{A12}$$

for some positive constant c_2. This implies the part (ii).

Last, consider the proof of part (iii). Let $A = \lambda_{\min}(\mathbf{D}_{nj})$ and $B = \lambda_{\min}(\mathbf{D}_j)$. Obviously, we know that $A, B > 0$. Using the same arguments as in [7], we can show that for $a \in (0,1)$, $|A^{-1} - B^{-1}| \geq cB^{-1}$ implies $|A - B| \geq aB$, where $c = \frac{1}{1-a} - 1$. Thus, $|\lambda_{\min}^{-1}(\mathbf{D}_{nj}) - \lambda_{\min}^{-1}(\mathbf{D}_j)| \geq (1/(1-c_1) - 1)\lambda_{\min}^{-1}(\mathbf{D}_j)$ implies $|\lambda_{\min}(\mathbf{D}_{nj}) - \lambda_{\min}(\mathbf{D}_j)| \geq c_1\lambda_{\min}(\mathbf{D}_j)$. Hence, using the fact that $\lambda_{\min}^{-1}(\mathbf{A}) = \lambda_{\max}(\mathbf{A}^{-1}) = \|\mathbf{A}^{-1}\|$ for any real symmetric invertible matrix $\mathbf{A}$, we have

$$\begin{aligned} &P\big(\big|\|\mathbf{D}_{nj}^{-1})\|\big| \geq (1+c_3)\|\mathbf{D}_j^{-1}\|\big) \\ &\leq P\big(\big|\|\mathbf{D}_{nj}^{-1})\| - \|\mathbf{D}_j^{-1}\|\big| \geq c_3\|\mathbf{D}_j^{-1}\|\big) \\ &\leq P\big(\big|\lambda_{\min}(\mathbf{D}_{nj}) - \lambda_{\min}(\mathbf{D}_j)\big| \geq c_1\lambda_{\min}(\mathbf{D}_j)\big) \\ &\leq 2(r_nL_n)^2\exp\big(-c_2a_0^{2r_n}r_n^{-2}L_n^{-3}n\big), \end{aligned} \tag{A13}$$

where $c_3 = 1/(1-c_1) - 1$. This completes the proof. □

Lemma A8. *Under conditions (C1)–(C3), for every $1 \le j \le p$ and for any given positive constant c_1^*, there exist some positive constants c_2^* such that*

$$P(\|\widehat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j^0\| \ge c_1^* a_0^{-r_n} r_n^{1/2} L_n) \le [8(r_n L_n)^2 + 4 r_n L_n] \exp\big(-c_2^* a_0^{2r_n} r_n^{-2} L_n^{-3} n\big).$$

Proof of Lemma A8. By the definitions of $\widehat{\boldsymbol{\theta}}_j$ and $\boldsymbol{\theta}_j^0$ and a simple algebra operation, we have

$$\widehat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j^0 = (\mathbf{D}_{nj}^{-1} - \mathbf{D}_j^{-1})\mathbf{E}_{nj} + \mathbf{D}_j^{-1}(\mathbf{E}_{nj} - \mathbf{E}_j) \triangleq I_{n1} + I_{n2} \text{ (say)}. \tag{A14}$$

In the following, we need to find the exponential tail probabilities for I_{n1} and I_{n2}, respectively.

We first deal with the first term I_{n1}. Since $\mathbf{D}_{nj}^{-1} - \mathbf{D}_j^{-1} = \mathbf{D}_{nj}^{-1}(\mathbf{D}_j - \mathbf{D}_{nj})\mathbf{D}_j^{-1}$, we have

$$\begin{aligned} \|I_{n1}\|^2 &= \mathbf{E}_{nj}^T \mathbf{D}_j^{-1}(\mathbf{D}_j - \mathbf{D}_{nj})\mathbf{D}_{nj}^{-1}\mathbf{D}_{nj}^{-1}(\mathbf{D}_j - \mathbf{D}_{nj})\mathbf{D}_j^{-1}\mathbf{E}_{nj} \\ &\le \|\mathbf{D}_j^{-1}\|^2 \|\mathbf{D}_{nj}^{-1}\|^2 \|\mathbf{D}_{nj} - \mathbf{D}_j\|^2 \|\mathbf{E}_{nj}\|^2. \end{aligned}$$

Thus, it follows from the triangle inequality and Lemma 1 that

$$\begin{aligned} \|I_{n1}\| &\le \lambda_{\min}^{-1}(\mathbf{D}_j)\|\mathbf{D}_{nj}^{-1}\| \cdot \big|\lambda_{\max}(\mathbf{D}_{nj} - \mathbf{D}_j)\big| \cdot \|\mathbf{E}_{nj}\| \\ &\le C_1^{-1} a_0^{-|\mathcal{S}_j|+1} L_n \|\mathbf{D}_{nj}^{-1}\| \cdot \big|\lambda_{\max}(\mathbf{D}_{nj} - \mathbf{D}_j)\big| \cdot \|\mathbf{E}_j\| \\ &\quad + C_1^{-1} a_0^{-|\mathcal{S}_j|+1} L_n \|\mathbf{D}_{nj}^{-1}\| \cdot \big|\lambda_{\max}(\mathbf{D}_{nj} - \mathbf{D}_j)\big| \cdot \|\mathbf{E}_{nj} - \mathbf{E}_j\| \\ &\triangleq I_{n1}^{(1)} + I_{n1}^{(2)} \text{ (say)}. \end{aligned}$$

For $I_{n1}^{(1)}$, it follows that

$$\begin{aligned} \|\mathbf{E}_j\|^2 &= \sum_{k \in \mathcal{S}_j} \sum_{l=1}^{L_n} \big[\mathrm{E}\{B_l(X_{ik})X_{ij}\}\big]^2 \le \sum_{k \in \mathcal{S}_j} \sum_{l=1}^{L_n} \mathrm{E}\big[B_l^2(X_{ik})X_{ij}^2\big] \\ &\le \sum_{k \in \mathcal{S}_j} \sum_{l=1}^{L_n} \mathrm{E}\big[B_l^2(X_{ik})\mathrm{E}\{X_{ij}^2 | \mathbf{X}_{-j}\}\big] \le 2K_1 K_2^2 C_3 r_n = C_4 r_n, \end{aligned}$$

where $C_4 = 2K_1K_2^2C_3$ and the last inequality holds by applying Lemma A1 and the result in (10). Using the above result, we have

$$I_{n1}^{(1)} \le C_1^{-1} C_4^{1/2} a_0^{-r_n+1} r_n^{1/2} L_n \|\mathbf{D}_{nj}^{-1}\| \cdot \big|\lambda_{\max}(\mathbf{D}_{nj} - \mathbf{D}_j)\big|.$$

Let $C_5 = (1 + c_3)a_0^2 C_1^{-2} C_4^{1/2}$, then for any $\delta > 0$, we have

$$\begin{aligned} P(|I_{n1}^{(1)}| \ge C_5 a_0^{-2r_n} r_n^{3/2} L_n^3 \delta / n) &\le P\big(\|\mathbf{D}_{nj}^{-1}\| \ge (1 + c_3)\|\mathbf{D}_j^{-1}\|\big) \\ &\quad + P\big(\big|\lambda_{\max}(\mathbf{D}_{nj} - \mathbf{D}_j)\big| \ge r_n L_n \delta / n\big). \end{aligned}$$

Therefore, by Lemma A7, it follows that

$$\begin{aligned} &P\big(|I_{n1}^{(1)}| \ge C_5 a_0^{-2r_n} r_n^{3/2} L_n^3 \delta / n\big) \\ &\le 2(r_n L_n)^2 \exp\big(-c_2 a_0^{2r_n} r_n^{-2} L_n^{-3} n\big) + 2(r_n L_n)^2 \exp\Big(-\frac{\delta^2}{2(C_3 L_n^{-1} n + 2\delta/3)}\Big). \end{aligned} \tag{A15}$$

For $I_{n1}^{(2)}$, note that $\mathbf{E}_{nj} - \mathbf{E}_j = \frac{1}{n}\sum_{i=1}^n \left[\mathbf{B}_{ij}X_{ij} - \mathrm{E}\{\mathbf{B}_{ij}X_{ij}\}\right]$ is an $|\mathcal{S}_j|L_n \times 1$ vector, whose $((k-1)L_n+1)$th component is $\frac{1}{n}\sum_{i=1}^n \left[B_l(X_{ik})X_{ij} - \mathrm{E}\{B_l(X_{ik})X_{ij}\}\right]$, where $k \in \mathcal{S}_j$ and $l = 1, \ldots, L_n$. Let $Z_{iklj} = B_l(X_{ik})X_{ij} - \mathrm{E}\{B_l(X_{ik})X_{ij}\}$. Then, for every $r \geq 2$, we have

$$\begin{aligned}
\mathrm{E}\{|Z_{iklj}|^r\} &\leq 2^r\mathrm{E}\{|B_l(X_{ik})X_{ij}|^r\} \leq 2^r\mathrm{E}\{B_l^2(X_{ik})|X_{ij}|^r\} \\
&\leq 2^r\mathrm{E}\{B_l^2(X_{ik})\mathrm{E}(|X_{ij}|^r|\mathbf{X}_{-j})\} \leq 2^r K_1 K_2^r r! C_3 L_n^{-1} \\
&= r!(2K_2)^{r-2} 8K_1K_2^2C_3L_n^{-1}/2,
\end{aligned}$$

where we have used the C_r inequality that $|x+y|^r \leq 2^{r-1}(|x|^r + |y|^r)$ for $r \geq 2$, the fact that $\|B_l\|_\infty \leq 1$ as well as Lemma A1. It follows from Lemma A2 that for any $\delta > 0$,

$$P\Big(\Big|\frac{1}{n}\sum_{i=1}^n Z_{iklj}\Big| \geq \frac{\delta}{n}\Big) \leq 2\exp\Big(-\frac{\delta^2}{c_4L_n^{-1}n + c_5\delta}\Big), \tag{A16}$$

where $c_4 = 16K_1K_2^2C_3$ and $c_5 = 4K_2$. Employing the union bound of probability and the inequality (A16), we further have

$$P\big(\|\mathbf{E}_{nj} - \mathbf{E}_j\| \geq r_n^{1/2}L_n^{1/2}\delta/n\big) \leq 2r_nL_n\exp\Big(-\frac{\delta^2}{c_4L_n^{-1}n + c_5\delta}\Big). \tag{A17}$$

Let $C_6 = (1+c_3)C_1^{-2}a_0^2$. Similar to the derivation of (A15) and by Lemma 1 and Lemma A7 and (A17), we obtain

$$\begin{aligned}
&P\big(|I_{n1}^{(2)}| \geq C_6a_0^{-2r_n}r_n^{3/2}L_n^{7/2}\delta^2/n^2\big) \\
&\leq P\big(\|\mathbf{D}_{nj}^{-1}\| \cdot |\lambda_{\max}(\mathbf{D}_{nj} - \mathbf{D}_j)| \geq C_6C_1a_0^{-r_n-1}r_nL_n^2\delta/n\big) \\
&\quad + P\big(\|\mathbf{E}_{nj} - \mathbf{E}_j\| \geq r_n^{1/2}L_n^{1/2}\delta/n\big) \\
&\leq P\big(\|\mathbf{D}_{nj}^{-1}\| \geq (1+c_3)\|\mathbf{D}_j^{-1}\|\big) + P\big(|\lambda_{\max}(\mathbf{D}_{nj} - \mathbf{D}_j)| > r_nL_n\delta/n\big) \\
&\quad + P\big(\|\mathbf{E}_{nj} - \mathbf{E}_j\| \geq r_n^{1/2}L_n^{1/2}\delta/n\big) \\
&\leq 2(r_nL_n)^2\exp\big(-c_2a_0^{2r_n}r_n^{-2}L_n^{-3}n\big) + 2(r_nL_n)^2\exp\Big(-\frac{\delta^2}{2(C_3L_n^{-1}n + 2\delta/3)}\Big) \\
&\quad + 2r_nL_n\exp\Big(-\frac{\delta^2}{c_4L_n^{-1}n + c_5\delta}\Big)
\end{aligned} \tag{A18}$$

Hence, combining (A15) and (A18) gives

$$\begin{aligned}
&P\big(\|I_{n1}\| \geq C_5a_0^{-2r_n}r_n^{3/2}L_n^3\delta/n + C_6a_0^{-2r_n}r_n^{3/2}L_n^{7/2}\delta^2/n^2\big) \\
&\leq P\big(|I_{n1}^{(1)}| \geq C_5a_0^{-2r_n}r_n^{3/2}L_n^3\delta/n\big) + P\big(|I_{n1}^{(2)}| \geq C_6a_0^{-2r_n}r_n^{3/2}L_n^{7/2}\delta^2/n^2\big) \\
&\leq 4(r_nL_n)^2\exp\big(-c_2a_0^{2r_n}r_n^{-2}L_n^{-3}n\big) + 4(r_nL_n)^2\exp\Big(-\frac{\delta^2}{2(C_3L_n^{-1}n + 2\delta/3)}\Big) \\
&\quad + 2r_nL_n\exp\Big(-\frac{\delta^2}{c_4L_n^{-1}n + c_5\delta}\Big) \\
&\leq 2[2(r_nL_n)^2 + r_nL_n]\exp\Big(-\frac{\delta^2}{c_6L_n^{-1}n + c_7\delta}\Big) \\
&\quad + 4(r_nL_n)^2\exp\big(-c_2a_0^{2r_n}r_n^{-2}L_n^{-3}n\big),
\end{aligned} \tag{A19}$$

where $c_6 = \max(2C_3, c_4)$ and $c_7 = \max(c_5, 4/3)$.

Next, we deal with the second term I_{n2}. Since $\|I_{n2}\|^2 = (\mathbf{E}_{nj} - \mathbf{E}_j)^T \mathbf{D}_j^{-1} \mathbf{D}_j^{-1} (\mathbf{E}_{nj} - \mathbf{E}_j) \leq \|\mathbf{D}_j^{-1}\|^2 \|\mathbf{E}_{nj} - \mathbf{E}_j\|^2$, we have $\|I_{n2}\| \leq \lambda_{\min}^{-1}(\mathbf{D}_j)\|\mathbf{E}_{nj} - \mathbf{E}_j\| \leq C_1^{-1} a_0^{-r_n+1} L_n \|\mathbf{E}_{nj} - \mathbf{E}_j\|$ by Lemma 1. Then, it follows from (A17) that

$$\begin{aligned} P\big(\|I_{n2}\| \geq C_1^{-1} a_0^{-r_n+1} r_n^{1/2} L_n^{3/2} \delta/n\big) &\leq P\big(\|\mathbf{E}_{nj} - \mathbf{E}_j\| \geq r_n^{1/2} L_n^{1/2} \delta/n\big) \\ &\leq 2 r_n L_n \exp\Big(- \frac{\delta^2}{c_4 L_n^{-1} n + c_5 \delta} \Big). \end{aligned} \tag{A20}$$

Putting (A14), (A19) and (A20) together, we find that

$$\begin{aligned} &P\big(\|\widehat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j^0\| \geq C_5 a_0^{-2r_n} r_n^{3/2} L_n^3 \delta/n + C_6 a_0^{-2r_n} r_n^{3/2} L_n^{7/2} \delta^2/n^2 + C_1^{-1} a_0^{-r_n+1} r_n^{1/2} L_n^{3/2} \delta/n\big) \\ &\leq P\big(\|I_{n1}\| \geq C_5 a_0^{-2r_n} r_n^{3/2} L_n^3 \delta/n + C_6 a_0^{-2r_n} r_n^{3/2} L_n^{7/2} \delta^2/n^2\big) \\ &\quad + P\big(\|I_{n2}\| \geq C_1^{-1} a_0^{-r_n+1} r_n^{1/2} L_n^{3/2} \delta/n\big) \\ &\leq 4[(r_n L_n)^2 + r_n L_n] \exp\Big(- \frac{\delta^2}{c_6 L_n^{-1} n + c_7 \delta} \Big) \\ &\quad + 4 (r_n L_n)^2 \exp\big(- c_2 a_0^{2r_n} r_n^{-2} L_n^{-3} n \big), \end{aligned} \tag{A21}$$

Using (A21) with $\delta = a_0^{r_n} r_n^{-1} L_n^{-2} n$, we have

$$P(\|\widehat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j^0\| \geq c_1^* a_0^{-r_n} r_n^{1/2} L_n) \leq [8(r_n L_n)^2 + 4 r_n L_n] \exp\big(- c_2^* a_0^{2r_n} r_n^{-2} L_n^{-3} n \big)$$

for some positive constant c_2^* and sufficiently large n, where $c_1^* = C_5 + C_6 + C_1^{-1} a_0$. Hence, the desired result follows. □

Lemma A9. *Under conditions (C1)–(C5), for any given constant $C > 0$ and for every $1 \leq j \leq p$, there exist some positive constants c_9 and c_{11} such that*

$$P(\|\widehat{\boldsymbol{\alpha}}_j - \boldsymbol{\alpha}_j^0\| \geq C (r_n L_n)^{1/2} n^{-\kappa}) \leq 2 \exp\big(- c_9 a_0^{2r_n} r_n^2 n^{1-4\kappa} \big) + \exp\big(- c_{11} a_0^{2r_n} L_n^{-2} n^{1-2\kappa} \big).$$

Proof of Lemma A9. Write $W_n(\boldsymbol{\alpha}_j) = \frac{1}{n}\sum_{i=1}^n \{\rho_\tau(Y_i - \mathbf{B}_{ij}^T \boldsymbol{\alpha}_j) - \rho_\tau(Y_i)\}$ and $W(\boldsymbol{\alpha}_j) = \mathrm{E}\{\rho_\tau(Y - \mathbf{B}_j^T \boldsymbol{\alpha}_j) - \rho_\tau(Y)\}$. By Lemma A.2 of [13], we have, for any $\epsilon > 0$,

$$\begin{aligned} &P\big(\|\widehat{\boldsymbol{\alpha}}_j - \boldsymbol{\alpha}_j^0\| \geq \epsilon\big) \\ &\leq P\Big(\sup_{\|\boldsymbol{\alpha}_j - \boldsymbol{\alpha}_j^0\| \leq \epsilon} \big|W_n(\boldsymbol{\alpha}_j) - W(\boldsymbol{\alpha}_j)\big| \geq \frac{1}{2} \inf_{\|\boldsymbol{\alpha}_j - \boldsymbol{\alpha}_j^0\| = \epsilon} W(\boldsymbol{\alpha}_j) - W(\boldsymbol{\alpha}_j^0) \Big) \end{aligned} \tag{A22}$$

Taking $\epsilon = C(r_n L_n)^{1/2} n^{-\kappa}$ in (A22), where C is any given positive constant, we first show that there exists some positive constant c_8 such that

$$\inf_{\|\boldsymbol{\alpha}_j - \boldsymbol{\alpha}_j^0\| = C(r_n L_n)^{1/2} n^{-\kappa}} W(\boldsymbol{\alpha}_j) - W(\boldsymbol{\alpha}_j^0) \geq c_8 a_0^{r_n} r_n n^{-2\kappa}. \tag{A23}$$

To this end, let $\boldsymbol{\alpha}_j = \boldsymbol{\alpha}_j^0 + C(r_n L_n)^{1/2} n^{-\kappa} \mathbf{u}$ with $\|\mathbf{u}\| = 1$. Invoking the Knight's identity ([42], p121), i.e., $\rho_\tau(u - v) - \rho_\tau(u) = -v[\tau - I(u < 0)] + \int_0^v [I(u \leq s) - I(u \leq 0)] \mathrm{d}s$, we have

$$W(\boldsymbol{\alpha}_j) - W(\boldsymbol{\alpha}_j^0) = \mathrm{E}\Big\{ \int_0^{C(r_n L_n)^{1/2} n^{-\kappa} \mathbf{B}_j^T \mathbf{u}} I(0 < Y - \mathbf{B}_j^T \boldsymbol{\alpha}_j^0 \leq s) \mathrm{d}s \Big\}, \tag{A24}$$

where we have used the result that $\mathrm{E}\{\mathbf{B}_j\psi_\tau(Y-\mathbf{B}_j^T\boldsymbol{\alpha}_j^0)\}=0$ by the definition of $\boldsymbol{\alpha}_j^0$. Note that the right hand side of (A24) equals

$$\begin{aligned}&\mathrm{E}\Big\{\int_0^{C(r_nL_n)^{1/2}n^{-\kappa}\mathbf{B}_j^T\mathbf{u}}\mathrm{E}\{I(0<Y-\mathbf{B}_j^T\boldsymbol{\alpha}_j^0\le s)\big|\mathbf{X}\}\mathrm{d}s\Big\}\\&=\mathrm{E}\Big\{\int_0^{C(r_nL_n)^{1/2}n^{-\kappa}\mathbf{B}_j^T\mathbf{u}}f_{Y|\mathbf{X}}(y^*)s\mathrm{d}s\Big\},\end{aligned}$$

for y^* between $\mathbf{B}_j^T\boldsymbol{\alpha}_j^0$ and $\mathbf{B}_j^T\boldsymbol{\alpha}_j^0+s$. By condition (C4), it follows that

$$\begin{aligned}W(\boldsymbol{\alpha}_j)-W(\boldsymbol{\alpha}_j^0)&\ge\frac{1}{2}c_{3f}C^2r_nL_nn^{-2\kappa}\mathrm{E}\{(\mathbf{B}_j^T\mathbf{u})^2\}\\&\ge\frac{1}{2}c_{3f}C^2r_nL_nn^{-2\kappa}\lambda_{\min}(\mathrm{E}\{\mathbf{B}_j\mathbf{B}_j^T\})\\&\ge\frac{1}{2}c_{3f}C^2C_1a_0^{r_n-1}r_nn^{-2\kappa}=c_8a_0^{r_n}r_nn^{-2\kappa},\end{aligned}$$

where $c_8=\frac{1}{2}c_{3f}C^2C_1a_0^{-1}$ and $a_0=(1-\delta_0)/2$. This proves (A23). Hence, by (A22), it reduces to derive that

$$\begin{aligned}&P\big(\|\widehat{\boldsymbol{\alpha}}_j-\boldsymbol{\alpha}_j^0\|\ge C(r_nL_n)^{1/2}n^{-\kappa}\big)\\&\le P\Big(\sup_{\|\boldsymbol{\alpha}_j-\boldsymbol{\alpha}_j^0\|\le C(r_nL_n)^{1/2}n^{-\kappa}}\big|W_n(\boldsymbol{\alpha}_j)-W(\boldsymbol{\alpha}_j)\big|\ge\frac{1}{2}c_8a_0^{r_n}r_nn^{-2\kappa}\Big)\\&\le P\Big(\sup_{\|\boldsymbol{\alpha}_j-\boldsymbol{\alpha}_j^0\|\le C(r_nL_n)^{1/2}n^{-\kappa}}\big|\{W_n(\boldsymbol{\alpha}_j)-W_n(\boldsymbol{\alpha}_j^0)\}-\{W(\boldsymbol{\alpha}_j)-W(\boldsymbol{\alpha}_j^0)\}\big|\ge\frac{1}{4}c_8a_0^{r_n}r_nn^{-2\kappa}\Big)\\&\quad+P\Big(\big|W_n(\boldsymbol{\alpha}_j^0)-W(\boldsymbol{\alpha}_j^0)\big|\ge\frac{1}{4}c_8a_0^{r_n}r_nn^{-2\kappa}\Big)\\&\triangleq J_{n1}+J_{n2}.\end{aligned}\tag{A25}$$

In what follows, we first consider J_{n2}. Let $U_{ij}=[\rho_\tau(Y_i-\mathbf{B}_{ij}^T\boldsymbol{\alpha}_j^0)-\rho_\tau(Y_i)]-\mathrm{E}[\rho_\tau(Y-\mathbf{B}_j^T\boldsymbol{\alpha}_j^0)-\rho_\tau(Y)]$ and then $W_n(\boldsymbol{\alpha}_j^0)-W(\boldsymbol{\alpha}_j^0)=\frac{1}{n}\sum_{i=1}^nU_{ij}$. Note that using the Knight's identity, we have $|\rho_\tau(u-v)-\rho_\tau(u)|\le|v|\max\{\tau-1,\tau\}\le|v|$. So, by using condition (C5), it follows that

$$|U_{ij}|\le2\big|\rho_\tau(Y_i-\mathbf{B}_{ij}^T\boldsymbol{\alpha}_j^0)-\rho_\tau(Y_i)\big|\le2\sup_{i,j}\big|\mathbf{B}_{ij}^T\boldsymbol{\alpha}_j^0\big|\le2M_1,$$

and

$$\mathrm{var}(U_{ij})\le\mathrm{E}\{[\rho_\tau(Y_i-\mathbf{B}_{ij}^T\boldsymbol{\alpha}_j^0)-\rho_\tau(Y_i)]^2\}\le\mathrm{E}\Big\{\sup_{i,j}|\mathbf{B}_{ij}^T\boldsymbol{\alpha}_j^0|^2\Big\}\le M_1^2.$$

According to Lemma A3, we have

$$\begin{aligned}J_{n2}&=P\Big(\Big|\frac{1}{n}\sum_{i=1}^nU_{ij}\Big|\ge\frac{1}{4}c_8a_0^{r_n}r_nn^{-2\kappa}\Big)\\&\le2\exp\Big(-\frac{16^{-1}c_8^2a_0^{2r_n}r_n^2n^{2-4\kappa}}{2(nM_1^2+M_1c_8a_0^{r_n}r_nn^{1-2\kappa}/6)}\Big)\\&\le2\exp\big(-c_9a_0^{2r_n}r_n^2n^{1-4\kappa}\big)\end{aligned}\tag{A26}$$

for some positive constant c_9, provided $a_0^{r_n}r_nn^{-2\kappa}=o(1)$.

Next, we consider J_{n1}. Define $V_{ij}(\boldsymbol{\alpha}_j) = \rho_\tau(Y_i - \mathbf{B}_{ij}^T\boldsymbol{\alpha}_j) - \rho_\tau(Y_i - \mathbf{B}_{ij}^T\boldsymbol{\alpha}_j^0)$ and so $W_n(\boldsymbol{\alpha}_j) - W_n(\boldsymbol{\alpha}_j^0) = \frac{1}{n}\sum_{i=1}^n V_{ij}(\boldsymbol{\alpha}_j)$. This leads to

$$J_{n1} = P\Big(\sup_{\|\boldsymbol{\alpha}_j-\boldsymbol{\alpha}_j^0\|\le C(r_nL_n)^{1/2}n^{-\kappa}}\Big|\frac{1}{n}\sum_{i=1}^n\big[V_{ij}(\boldsymbol{\alpha}_j) - \mathrm{E}\{V_{ij}(\boldsymbol{\alpha}_j)\}\big]\Big| \ge \frac{1}{4}c_8 a_0^{r_n} r_n n^{-2\kappa}\Big). \tag{A27}$$

Again, using the Knight's identity, we obtain

$$\begin{aligned}|V_{ij}(\boldsymbol{\alpha}_j)| &\le \big|\mathbf{B}_{ij}^T(\boldsymbol{\alpha}_j-\boldsymbol{\alpha}_j^0)[I(Y_i-\mathbf{B}_{ij}^T\boldsymbol{\alpha}_j^0<0)-\tau]\big| \\ &\quad+\Big|\int_0^{\mathbf{B}_{ij}^T(\boldsymbol{\alpha}_j-\boldsymbol{\alpha}_j^0)}\{I(Y_i-\mathbf{B}_{ij}^T\boldsymbol{\alpha}_j^0\le s)-I(Y_i-\mathbf{B}_{ij}^T\boldsymbol{\alpha}_j^0\le 0)\}\mathrm{d}s\Big| \\ &\le 2\big|\mathbf{B}_{ij}^T(\boldsymbol{\alpha}_j-\boldsymbol{\alpha}_j^0)\big| \le 2(|\mathcal{S}_j|L_n)^{1/2}\|\boldsymbol{\alpha}_j-\boldsymbol{\alpha}_j^0\|,\end{aligned}$$

where the last line is because $\|B_k\|_\infty \le 1$. Thus, it follows that

$$\begin{aligned}\sup_{\|\boldsymbol{\alpha}_j-\boldsymbol{\alpha}_j^0\|\le C(r_nL_n)^{1/2}n^{-\kappa}}\big|V_{ij}(\boldsymbol{\alpha}_j)\big| &\le 2(|\mathcal{S}_j|L_n)^{1/2}\Big\{\sup_{\|\boldsymbol{\alpha}_j-\boldsymbol{\alpha}_j^0\|\le C(r_nL_n)^{1/2}n^{-\kappa}}\|\boldsymbol{\alpha}_j-\boldsymbol{\alpha}_j^0\|\Big\} \\ &\le 2Cr_nL_nn^{-\kappa}.\end{aligned} \tag{A28}$$

Let $\varepsilon_1,\ldots,\varepsilon_n$ be a Rademacher sequence independent of $V_{ij}(\boldsymbol{\alpha}_j)$. By Lemmas A4 and A5, we have

$$\begin{aligned}&\mathrm{E}\Big\{\sup_{\|\boldsymbol{\alpha}_j-\boldsymbol{\alpha}_j^0\|\le C(r_nL_n)^{1/2}n^{-\kappa}}\Big|\frac{1}{n}\sum_{i=1}^n[V_{ij}(\boldsymbol{\alpha}_j)-\mathrm{E}\{V_{ij}(\boldsymbol{\alpha}_j)\}]\Big|\Big\} \\ &\le 2\mathrm{E}\Big\{\sup_{\|\boldsymbol{\alpha}_j-\boldsymbol{\alpha}_j^0\|\le C(r_nL_n)^{1/2}n^{-\kappa}}\Big|\frac{1}{n}\sum_{i=1}^n\varepsilon_iV_{ij}(\boldsymbol{\alpha}_j)\Big|\Big\} \\ &= 2\mathrm{E}\Big\{\sup_{\|\boldsymbol{\alpha}_j-\boldsymbol{\alpha}_j^0\|\le C(r_nL_n)^{1/2}n^{-\kappa}}\Big|\frac{1}{n}\sum_{i=1}^n\varepsilon_i[\rho_\tau(Y_i-\mathbf{B}_{ij}^T\boldsymbol{\alpha}_j)-\rho_\tau(Y_i-\mathbf{B}_{ij}^T\boldsymbol{\alpha}_j^0)]\Big|\Big\} \\ &\le 4\mathrm{E}\Big\{\sup_{\|\boldsymbol{\alpha}_j-\boldsymbol{\alpha}_j^0\|\le C(r_nL_n)^{1/2}n^{-\kappa}}\Big|\frac{1}{n}\sum_{i=1}^n\varepsilon_i\mathbf{B}_{ij}^T(\boldsymbol{\alpha}_j-\boldsymbol{\alpha}_j^0)\Big|\Big\} \\ &\le 4C(r_nL_n)^{1/2}n^{-\kappa}\mathrm{E}\Big\{\Big\|\frac{1}{n}\sum_{i=1}^n\varepsilon_i\mathbf{B}_{ij}\Big\|\Big\} \\ &\le 4C(r_nL_n)^{1/2}n^{-\kappa}\Big\{\mathrm{E}\Big\|\frac{1}{n}\sum_{i=1}^n\varepsilon_i\mathbf{B}_{ij}\Big\|^2\Big\}^{1/2} \\ &= 4C(r_nL_n)^{1/2}n^{-\kappa}\Big\{n^{-2}\sum_{k\in\mathcal{S}_j}\sum_{l=1}^{L_n}\sum_{i=1}^n\mathrm{E}[\varepsilon_i^2B_l^2(X_{ik})]\Big\}^{1/2} \\ &\le c_{10}r_nL_n^{1/2}n^{-\frac{1}{2}-\kappa},\end{aligned}$$

where $c_{10} = 4CC_3^{1/2}$ and we have used (10) in the last line. With the above arguments, we can apply Lemma A6 to derive J_{n2} in equation (A27). Set

$$\mathrm{U} = \sup_{\|\boldsymbol{\alpha}_j-\boldsymbol{\alpha}_j^0\|\le C(r_nL_n)^{1/2}n^{-\kappa}}\Big|\frac{1}{n}\sum_{i=1}^n\big[V_{ij}(\boldsymbol{\alpha}_j)-\mathrm{E}\{V_{ij}(\boldsymbol{\alpha}_j)\}\big]\Big|.$$

Taking $t = \frac{1}{4}c_8 a_0^{r_n} r_n n^{-2\kappa} - c_{10} r_n L_n^{1/2} n^{-\frac{1}{2}-\kappa}$ in Lemma A6, we have

$$\begin{aligned} J_{n2} &= P\Big(\mathrm{U} \geq \frac{1}{4}c_8 a_0^{r_n} r_n n^{-2\kappa}\Big) = P\Big(\mathrm{U} \geq \mathrm{E}\{\mathrm{U}\} + (\frac{1}{4}c_8 a_0^{r_n} r_n n^{-2\kappa} - \mathrm{E}\{\mathrm{U}\})\Big) \\ &\leq P\Big(\mathrm{U} \geq \mathrm{E}\{\mathrm{U}\} + (\frac{1}{4}c_8 a_0^{r_n} r_n n^{-2\kappa} - c_{10} r_n L_n^{1/2} n^{-\frac{1}{2}-\kappa})\Big) \\ &\leq \exp\Big(-\frac{n(\frac{1}{4}c_8 a_0^{r_n} r_n n^{-2\kappa} - c_{10} r_n L_n^{1/2} n^{-\frac{1}{2}-\kappa})^2}{2(2C r_n L_n n^{-\kappa})^2}\Big) \\ &\leq \exp\big(-c_{11} a_0^{2r_n} L_n^{-2} n^{1-2\kappa}\big) \end{aligned} \tag{A29}$$

foe some positive constant c_{11}, provided $a_0^{-2r_n} L_n / n^{1-2\kappa} = o(1)$. Plugging (A26) and (A29) into (A25) gives the desired result. □

Lemma A10. *Under conditions (C1)–(C5), for every $1 \leq j \leq p$ and for any given constant c_5^*, there exist some positive constants c_6^* and c_7^* such that*

$$\begin{aligned} &P\Big(\Big|\frac{1}{n}\sum_{i=1}^{n}\psi_\tau(Y_i - \mathbf{B}_{ij}^T\widehat{\boldsymbol{\alpha}}_j)(X_{ij} - \mathbf{B}_{ij}^T\widehat{\boldsymbol{\theta}}_j) - \mathrm{E}\{\psi_\tau(Y_i - \mathbf{B}_{ij}^T\boldsymbol{\alpha}_j^0)(X_{ij} - \mathbf{B}_{ij}^T\boldsymbol{\theta}_j^0)\}\Big| \geq c_5^* r_n n^{-\kappa}\Big) \\ &\leq 7\exp\big(-c_6^* a_0^{2r_n} r_n^2 n^{1-4\kappa}\big) + [8(r_n L_n)^2 + 4 r_n L_n]\exp\big(-c_7^* a_0^{2r_n} r_n^{-2} L_n^{-3} n\big). \end{aligned}$$

Proof of Lemma A10. Since $\mathrm{E}\{\mathbf{B}_{ij}\psi_\tau(Y_i - \mathbf{B}_{ij}^T\boldsymbol{\theta}_j^0)\} = 0$ by definition, so $\mathrm{E}\{\psi_\tau(Y_i - \mathbf{B}_{ij}^T\boldsymbol{\theta}_j^0)(X_{ij} - \mathbf{B}_{ij}^T\boldsymbol{\theta}_j^0)\} = \mathrm{E}\{\psi_\tau(Y_i - \mathbf{B}_{ij}^T\boldsymbol{\theta}_j^0)X_{ij}\}$. A simple decomposition gives

$$\begin{aligned} &n^{-1}\sum_{i=1}^{n}\psi_\tau(Y_i - \mathbf{B}_{ij}^T\widehat{\boldsymbol{\alpha}}_j)(X_{ij} - \mathbf{B}_{ij}^T\widehat{\boldsymbol{\theta}}_j) - \mathrm{E}\{\psi_\tau(Y_i - \mathbf{B}_j^T\boldsymbol{\alpha}_j^0)X_{ij}\} \\ &= n^{-1}\sum_{i=1}^{n}[\psi_\tau(Y_i - \mathbf{B}_{ij}^T\boldsymbol{\alpha}_j^0)X_{ij} - \mathrm{E}\{\psi_\tau(Y - \mathbf{B}_j^T\boldsymbol{\alpha}_j^0)X_j\}] \\ &\quad + n^{-1}\sum_{i=1}^{n}\{\psi_\tau(Y_i - \mathbf{B}_{ij}^T\widehat{\boldsymbol{\alpha}}_j) - \psi_\tau(Y_i - \mathbf{B}_{ij}^T\boldsymbol{\alpha}_j^0)\}X_{ij} - n^{-1}\sum_{i=1}^{n}\psi_\tau(Y_i - \mathbf{B}_{ij}^T\widehat{\boldsymbol{\alpha}}_j)\mathbf{B}_{ij}^T\widehat{\boldsymbol{\theta}}_j \\ &\triangleq \Delta_{n1j} + \Delta_{n2j} + \Delta_{n3j}. \end{aligned} \tag{A30}$$

The rest is to find exponential bounds for the tail probabilities of $\Delta_{n1j}, \Delta_{n2j}$ and Δ_{n3j}, respectively.

For Δ_{n1j}, since $|\psi_\tau(u)| \leq \max(\tau, 1-\tau) \leq 1$, so it follows from the C_r inequality and Lemma A1 that for each $r \geq 2$,

$$\begin{aligned} &\mathrm{E}\big\{\big|\psi_\tau(Y_i - \mathbf{B}_{ij}^T\boldsymbol{\alpha}_j^0)X_{ij} - \mathrm{E}\{\psi_\tau(Y - \mathbf{B}_j^T\boldsymbol{\alpha}_j^0)X_j\}\big|^r\big\} \\ &\leq 2^r\mathrm{E}\big\{\big|\psi_\tau(Y_i - \mathbf{B}_{ij}^T\boldsymbol{\alpha}_j^0)X_{ij}\big|^r\big\} \leq 2^r\mathrm{E}\big\{|X_{ij}|^r\big\} \\ &= 2^r\mathrm{E}\big\{\mathrm{E}(|X_{ij}|^r|\mathbf{X}_{-j})\big\} \leq 2^r K_1 K_2^r r! = r!(2K_2)^{r-2} 8K_1 K_2^2/2. \end{aligned}$$

Invoking Lemma A2, for any $\delta > 0$, we have

$$P(|\Delta_{n1j}| \geq \delta/n) \leq 2\exp\Big(-\frac{\delta^2}{c_{12}n + c_{13}\delta}\Big), \tag{A31}$$

where $c_{12} = 16K_1K_2^2$ and $c_{13} = 4K_2$.

For Δ_{n2j}, note that for each $r \geq 2$,

$$\begin{aligned} P(|\Delta_{n2j}| \geq \delta/n) &\leq P\big(|\Delta_{n2j}| \geq \delta/n, \|\widehat{\boldsymbol{\alpha}}_j - \boldsymbol{\alpha}_j^0\| < C(r_nL_n)^{1/2}n^{-\kappa}\big) \\ &\quad + P\big(\|\widehat{\boldsymbol{\alpha}}_j - \boldsymbol{\alpha}_j^0\| \geq C(r_nL_n)^{1/2}n^{-\kappa}\big) \\ &\triangleq H_{n1j} + H_{n2j}, \end{aligned} \tag{A32}$$

where a direct application of Lemma A9 yields $H_{n2j} \leq 2\exp\left(-c_9 a_0^{2r_n} r_n^2 n^{1-4\kappa}\right) + \exp\left(-c_{11} a_0^{2r_n} L_n^{-2} n^{1-2\kappa}\right)$. Let $\widehat{\boldsymbol{\alpha}}_j = \boldsymbol{\alpha}_j^0 + C(r_n L_n)^{1/2} n^{-\kappa}\mathbf{u}$ with $\|\mathbf{u}\| \leq 1$. Denote

$$\Pi_{ij} = \sup_{\|\mathbf{u}\|\leq 1} |\{\psi_\tau(Y_i - \mathbf{B}_{ij}^T\boldsymbol{\alpha}_j^0 - C(r_n L_n)^{1/2} n^{-\kappa}\mathbf{B}_{ij}^T\mathbf{u}) - \psi_\tau(Y_i - \mathbf{B}_{ij}^T\boldsymbol{\alpha}_j^0)\}X_{ij}|.$$

Then,

$$H_{n1j} \leq P\Big(\Big|n^{-1}\sum_{i=1}^n \Pi_{ij}\Big| \geq \frac{\delta}{n}\Big). \tag{A33}$$

Furthermore, there exists a $\mathbf{u}^* = (\{\mathbf{u}_k^{*T}, k \in \mathcal{S}_j\})^T$ with $\|\mathbf{u}^*\| \leq 1$ and $\mathbf{u}_k^* \in \mathrm{R}^{L_n}$ such that

$$\begin{aligned}
\mathrm{E}\{\Pi_{ij}\} &= \mathrm{E}\{|\{\psi_\tau(Y_i - \mathbf{B}_{ij}^T\boldsymbol{\alpha}_j^0 - C(r_n L_n)^{1/2} n^{-\kappa}\mathbf{B}_{ij}^T\mathbf{u}^*) - \psi_\tau(Y_i - \mathbf{B}_{ij}^T\boldsymbol{\alpha}_j^0)\}X_{ij}|\} \\
&\leq \mathrm{E}\Big\{\Big|\int_{\mathbf{B}_{ij}^T\boldsymbol{\alpha}_j^0}^{\mathbf{B}_{ij}^T\boldsymbol{\alpha}_j^0 + C(r_n L_n)^{1/2} n^{-\kappa}\mathbf{B}_{ij}^T\mathbf{u}^*} f_{Y|\mathbf{X}}(y)\mathrm{d}y\Big||X_{ij}|\Big\} \\
&\leq c_{4f} C(r_n L_n)^{1/2} n^{-\kappa}\mathrm{E}\{|\mathbf{B}_{ij}^T\mathbf{u}^*||X_{ij}|\} \\
&\leq c_{4f} C(r_n L_n)^{1/2} n^{-\kappa}\sqrt{\mathrm{E}\{|\mathbf{B}_{ij}^T\mathbf{u}^*|^2\}\mathrm{E}\{|X_{ij}|^2\}} \\
&\leq c_{14} r_n n^{-\kappa}
\end{aligned}$$

for some positive constant c_{14}, where we have used condition (C4) in the third line, Cauchy–Schwarz inequality in the fourth line, Lemmas 1 and A1 in the last line. Analogously to (A31), we have for each $r \geq 2$,

$$\mathrm{E}\{|\Pi_{ij} - \mathrm{E}(\Pi_{ij})|^r\} \leq 2^r\mathrm{E}\{|\Pi_{ij}|^r\} \leq 2^r\mathrm{E}\{2^r|X_{ij}|^r\} \leq r!(4K_2)^{r-2} 32K_2^2 K_1/2$$

and it follows from Lemma A2 that for any $\delta > 0$,

$$P\Big(\Big|\frac{1}{n}\sum_{i=1}^n\{\Pi_i - \mathrm{E}(\Pi_{ij})\}\Big| \geq \frac{\delta}{n}\Big) \leq 2\exp\Big(-\frac{\delta^2}{c_{15}n + c_{16}\delta}\Big), \tag{A34}$$

where $c_{15} = 64K_1K_2^2$ and $c_{16} = 8K_2$. Setting $\delta = c_{14} r_n n^{1-\kappa}$ in (A34), we obtain

$$\begin{aligned}
&P\Big(\Big|\frac{1}{n}\sum_{i=1}^n \Pi_i\Big| \geq 2c_{14} r_n n^{-\kappa}\Big) \\
&\leq P\Big(\Big|\frac{1}{n}\sum_{i=1}^n\{\Pi_i - \mathrm{E}(\Pi_{ij})\}\Big| \geq 2c_{14} r_n n^{-\kappa} - \mathrm{E}(\Pi_{ij})\Big) \\
&\leq P\Big(\Big|\frac{1}{n}\sum_{i=1}^n\{\Pi_i - \mathrm{E}(\Pi_{ij})\}\Big| \geq c_{14} r_n n^{-\kappa}\Big) \\
&\leq 2\exp\left(-c_{17} r_n^2 n^{1-2\kappa}\right).
\end{aligned} \tag{A35}$$

As $r_n^2 n^{1-2\kappa}/(a_0^{2r_n} L_n^{-2} n^{1-2\kappa}) \to \infty$ as $n \to \infty$, combining (A32), (A33) and (A35), we obtain

$$\begin{aligned}
P(|\Delta_{n2j}| \geq 2c_{14} r_n n^{-\kappa}) &\leq 2\exp\left(-c_9 a_0^{2r_n} r_n^2 n^{1-4\kappa}\right) + 3\exp\left(-c_{11} a_0^{2r_n} L_n^{-2} n^{1-2\kappa}\right) \\
&\leq 5\exp\left(-c_{18} a_0^{2r_n} r_n^2 n^{1-4\kappa}\right)
\end{aligned} \tag{A36}$$

for some positive constant c_{18}.

Finally, we consider Δ_{n3j}. Denote $\Phi(\boldsymbol{\alpha}_j) = n^{-1}\sum_{i=1}^n \rho_\tau(Y_i - \mathbf{B}_{ij}^T\boldsymbol{\alpha}_j)$ and define its subdifferential as $\partial\Phi(\boldsymbol{\alpha}_j) = (\{\partial\Phi_{(k-1)L_n+l}(\boldsymbol{\alpha}_j) : k \in \mathcal{S}_j, l = 1, \ldots, L_n\})^T$ with

$$\partial\Phi_{(k-1)L_n+l}(\boldsymbol{\alpha}_j) = -n^{-1}\sum_{i=1}^n \psi_\tau(Y_i - \mathbf{B}_{ij}^T\boldsymbol{\alpha}_j)B_l(X_{ik}) - n^{-1}\sum_{i=1}^n I(Y_i - \mathbf{B}_{ij}^T\boldsymbol{\alpha}_j = 0)v_i B_l(X_{ik})$$

and $v_i \in [\tau - 1, \tau]$. Recalling the definition of $\widehat{\boldsymbol{\alpha}}_j$, there exists $v_i^* \in [\tau - 1.\tau]$ such that $\partial\Phi_{(k-1)L_n+l}(\widehat{\boldsymbol{\alpha}}_j) = 0$. This yields

$$\Delta_{n3j} = n^{-1}\sum_{i=1}^n I(Y_i - \mathbf{B}_{ij}^T\widehat{\boldsymbol{\alpha}}_j = 0)v_i\mathbf{B}_{ij}^T\widehat{\boldsymbol{\theta}}_j.$$

Thus, by condition (C5), it follows that

$$\begin{aligned}
|\Delta_{n3j}| &\leq n^{-1}\sum_{i=1}^n I(Y_i - \mathbf{B}_{ij}^T\widehat{\boldsymbol{\alpha}}_j = 0)|\mathbf{B}_{ij}^T\widehat{\boldsymbol{\theta}}_j| \\
&\leq n^{-1}\sum_{i=1}^n I(Y_i - \mathbf{B}_{ij}^T\widehat{\boldsymbol{\alpha}}_j = 0)|\mathbf{B}_{ij}^T\boldsymbol{\theta}_j^0| \\
&\qquad + n^{-1}\sum_{i=1}^n I(Y_i - \mathbf{B}_{ij}^T\widehat{\boldsymbol{\alpha}}_j = 0)|\mathbf{B}_{ij}^T(\widehat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j^0)| \\
&\leq n^{-1}\sum_{i=1}^n I(Y_i - \mathbf{B}_{ij}^T\widehat{\boldsymbol{\alpha}}_j = 0)(M_2 + (r_nL_n)^{1/2}\|\widehat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j^0\|).
\end{aligned} \tag{A37}$$

Using Lemma A8, we obtain

$$\begin{aligned}
&P(M_2 + (r_nL_n)^{1/2}\|\widehat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j^0\| \geq M_2 + c_1^* a_0^{-r_n} r_n L_n^{3/2}) \\
&\leq [8(r_nL_n)^2 + 4r_nL_n]\exp\big(-c_2^* a_0^{2r_n} r_n^{-2} L_n^{-3} n\big).
\end{aligned} \tag{A38}$$

Note that $P(n^{-1}\sum_{i=1}^n I(Y_i - \mathbf{B}_{ij}^T\widehat{\boldsymbol{\alpha}}_j = 0) \geq \epsilon) = 0$ for any $\epsilon > 0$. Letting $\epsilon = n^{-1}L_n^{-3/2}$, we thus have

$$P\Big(n^{-1}\sum_{i=1}^n I(Y_i - \mathbf{B}_{ij}^T\widehat{\boldsymbol{\alpha}}_j = 0) \geq n^{-1}L_n^{-3/2}\Big) = 0. \tag{A39}$$

Gathering (A37)–(A39) gives

$$\begin{aligned}
&P\big(|\Delta_{n3j}| \geq n^{-1}L_n^{-3/2}(M_2 + c_1^* a_0^{-r_n} r_n L_n^{3/2})\big) \\
&\leq [8(r_nL_n)^2 + 4r_nL_n]\exp\big(-c_2^* a_0^{2r_n} r_n^{-2} L_n^{-3} n\big).
\end{aligned} \tag{A40}$$

Furthermore, using (A31) with $\delta = c_{14}r_n n^{1-\kappa}$, we have

$$P\big(|\Delta_{n1j} \geq c_{14}r_n n^{-\kappa}\big) \leq 2\exp\big(-c_3^* r_n^2 n^{1-2\kappa}\big) \tag{A41}$$

for some positive constant c_3^*. Accordingly, by (A36), (A40) and (A41), we obtain

$$\begin{aligned}
&P\big(|\Delta_{n1j} + \Delta_{n2j} + \Delta_{n3j}| \geq 3c_{14}r_n n^{-\kappa} + n^{-1}L_n^{-3/2}(M_2 + c_1^* a_0^{-r_n} r_n L_n^{3/2})\big) \\
&\leq 2\exp\big(-c_3^* r_n^2 n^{1-2\kappa}\big) + 5\exp\big(-c_{18} a_0^{2r_n} r_n^2 n^{1-4\kappa}\big) \\
&\quad + [8(r_nL_n)^2 + 4r_nL_n]\exp\big(-c_2^* a_0^{2r_n} r_n^{-2} L_n^{-3} n\big) \\
&\leq 7\exp\big(-c_4^* a_0^{2r_n} r_n^2 n^{1-4\kappa}\big) + [8(r_nL_n)^2 + 4r_nL_n]\exp\big(-c_2^* a_0^{2r_n} r_n^{-2} L_n^{-3} n\big)
\end{aligned}$$

for some positive constant c_4^*. As a result, the desired result follows for some given positive constant $c_5^* = 3c_{14} + M_2 + c_1^*$ and for sufficiently large n. $\square$

Lemma A11. *Under conditions (C1)–(C5), for every $1 \le j \le p$ and for any given constant c_8^*, there exist some positive constants c_{10}^* and c_{12}^* such that*

$$\begin{aligned} P\big(|\widehat{\sigma}_j^2 - \sigma_j^2| \ge c_5^* r_n n^{-\kappa}\big) &\le [8(r_nL_n)^2 + 6r_nL_n + 2]\exp\big(-c_{13}^* a_0^{2r_n} L_n^{-3} n^{1-2\kappa}\big) \\ &\quad + [10(r_nL_n)^2 + 4r_nL_n]\exp\big(-c_{10}^* a_0^{4r_n} r_n^{-3} L_n^{-4} n^{1-\kappa}\big) \end{aligned}$$

when n is sufficiently large. In addition, for some $\tilde{c}_1 \in (0,1)$,

$$\begin{aligned} P\big(|\widehat{\sigma}_j^2 - \sigma_j^2| \ge \tilde{c}_1 \sigma_j^2\big) &\le [8(r_nL_n)^2 + 6r_nL_n + 2]\exp\big(-c_{13}^* a_0^{2r_n} L_n^{-3} n^{1-2\kappa}\big) \\ &\quad + [10(r_nL_n)^2 + 4r_nL_n]\exp\big(-c_{10}^* a_0^{4r_n} r_n^{-3} L_n^{-4} n^{1-\kappa}\big) \end{aligned}$$

Proof of Lemma A11. Recalling the definition of $\widehat{\sigma}_j^2$ and σ_j^2, we have

$$\begin{aligned} |\widehat{\sigma}_j^2 - \sigma_j^2| &\le \Big| n^{-1}\sum_{i=1}^n (X_{ij} - \mathbf{B}_{ij}^T\boldsymbol{\theta}_j^0)^2 - \mathrm{E}\{(X_{ij} - \mathbf{B}_{ij}^T\boldsymbol{\theta}_j^0)^2\}\Big| \\ &\quad + \Big| n^{-1}\sum_{i=1}^n (X_{ij} - \mathbf{B}_{ij}^T\widehat{\boldsymbol{\theta}}_j)^2 - n^{-1}\sum_{i=1}^n (X_{ij} - \mathbf{B}_{ij}^T\boldsymbol{\theta}_j^0)^2\Big| \\ &\triangleq \Xi_{n1j} + \Xi_{n2j}. \end{aligned} \tag{A42}$$

Let $\xi_{ij} = (X_{ij} - \mathbf{B}_{ij}^T\boldsymbol{\theta}_j^0)^2 - \mathrm{E}\{(X_{ij} - \mathbf{B}_{ij}^T\boldsymbol{\theta}_j^0)^2\}$. For every $r \ge 2$, by the C_r inequality and condition (C5), we have $\mathrm{E}\{|\xi_{ij}|^r\} \le 2^r \mathrm{E}\{(X_{ij} - \mathbf{B}_{ij}^T\boldsymbol{\theta}_j^0)^{2r}\} \le 2^{3r-1}\{\mathrm{E}|X_{ij}|^{2r} + M_2^{2r}\} \le 2^{3r-1}\{K_1K_2^{2r}(2r)! + M_2^{2r}\} \le 2^{3r}\tilde{K}_1\tilde{K}_2^{2r}(2r)! \le 2^{3r}\tilde{K}_1\tilde{K}_2^{2r}(2r)^r r! = r!(16r\tilde{K}_2^2)^{r-2}512(r\tilde{K}_2^2)^2\tilde{K}_1/2$ with $\tilde{K}_1 = \max(K_1, 1)$ and $\tilde{K}_2 = \max(K_2, M_2)$. Thus, by Lemma A2, it follows

$$P\Big(\Xi_{n1j} \ge \frac{1}{2}c_5^* r_n n^{-\kappa}\Big) \le 2\exp(-c_6^* r_n^2 n^{1-2\kappa}) \tag{A43}$$

for some positive constant c_6^*. In addition, it is easily derived that

$$\begin{aligned} \Xi_{n2j} &\le (\widehat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j^0)^T\mathbf{D}_{nj}(\widehat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j^0) + \Big|2n^{-1}\sum_{i=1}^n (X_{ij} - \mathbf{B}_{ij}^T\boldsymbol{\theta}_j^0)\mathbf{B}_{ij}^T(\widehat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j^0)\Big| \\ &\triangleq \Xi_{n2j}^{(1)} + \Xi_{n2j}^{(2)}, \end{aligned} \tag{A44}$$

where $\Xi_{n2j}^{(1)} \le \lambda_{\max}(\mathbf{D}_{nj})\|\widehat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j^0\|^2$. Similarly, applying the arguments used in deriving Lemma A7(ii), we have that for any constant $\tilde{c}_1 \in (0,1)$, there exists some finite positive constant c_7^* such that

$$P\big(|\lambda_{\max}(\mathbf{D}_{nj})| \ge (1+\tilde{c}_1)\lambda_{\max}(\mathbf{D}_j)\big) \le 2(r_nL_n)^2\exp\big(-c_7^* a_0^{2r_n} r_n^{-2} L_n^{-3} n\big).$$

This together with Lemma 1 yields

$$P\big(|\lambda_{\max}(\mathbf{D}_{nj})| \ge (1+\tilde{c}_1)C_2 r_n L_n^{-1}\big) \le 2(r_nL_n)^2\exp\big(-c_7^* a_0^{2r_n} r_n^{-2} L_n^{-3} n\big). \tag{A45}$$

Moreover, employing (A21) with $\delta = (1+\tilde{c}_1)^{-1/2}C_2^{-1/2}(c_5^*/4)^{1/2}c_1^{*-1} a_0^{2r_n} r_n^{-3/2} L_n^{-5/2} n^{1-\kappa/2}$, we have

$$\begin{aligned} &P(\|\widehat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j^0\| \ge (1+\tilde{c}_1)^{-1/2}C_2^{-1/2}(c_5^*/4)^{1/2}L_n^{1/2}n^{-\kappa/2}) \\ &\le 4[(r_nL_n)^2 + r_nL_n]\exp\big(-c_8^* a_0^{4r_n} r_n^{-3} L_n^{-4} n^{1-\kappa}\big) \\ &\quad + 4(r_nL_n)^2\exp\big(-c_2^* a_0^{2r_n} r_n^{-2} L_n^{-3} n\big) \\ &\le 4[2(r_nL_n)^2 + r_nL_n]\exp\big(-c_9^* a_0^{4r_n} r_n^{-3} L_n^{-4} n^{1-\kappa}\big) \end{aligned}$$

for some positive constants c_8^* and c_9^*. This in conjunction with (A45) gives

$$\begin{aligned} P\Big(\Xi_{n2j}^{(1)} \geq \frac{1}{4}c_5^* r_n n^{-\kappa}\Big) &\leq P\big(|\lambda_{\max}(\mathbf{D}_{nj})| \geq (1+\tilde{c}_1)C_2 r_n L_n^{-1}\big) \\ &\quad + P\big(\|\widehat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j^0\| \geq (1+\tilde{c}_1)^{-1/2} C_2^{-1/2} (c_5^*/4)^{1/2} L_n^{1/2} n^{-\kappa/2}\big) \\ &\leq [10(r_n L_n)^2 + 4 r_n L_n] \exp\big(-c_{10}^* a_0^{4r_n} r_n^{-3} L_n^{-4} n^{1-\kappa}\big) \qquad \text{(A46)} \end{aligned}$$

for some positive constant c_{10}^*. For $\Xi_{n2j}^{(2)}$, let $N_{iklj} = (X_{ij} - \mathbf{B}_{ij}^T \boldsymbol{\theta}_j^0) B_l(X_{ik}), k \in \mathcal{S}_j, l = 1, \ldots, L_n$, and then for every $r \geq 2$, $\mathrm{E}\{|N_{iklj}|^r\} \leq \mathrm{E}\{|X_{ij} - \mathbf{B}_{ij}^T \boldsymbol{\theta}_j^0|^r\} \leq 2^{r-1}\{\mathrm{E}|X_{ij}|^r + \sup_{i,j} |\mathbf{B}_{ij}^T \boldsymbol{\theta}_j^0|^r\} \leq 2^{r-1}(K_1 K_2^r r! + M_2^r) \leq 2^{r-1}(2\tilde{K}_1 \tilde{K}_2^r r!) = r!(2\tilde{K}_2)^{r-2} 8\tilde{K}_1 \tilde{K}_2^2/2$, where $\tilde{K}_1 = \max(K_1, 1)$ and $\tilde{K}_2 = \max(K_2, M_2)$. Thus, it follows from Lemma A2 that

$$P\Big(\Big|n^{-1}\sum_{i=1}^n N_{iklj}\Big| \geq \frac{1}{8} c_5^* c_1^{*-1} a_0^{r_n} L_n^{-3/2} n^{-\kappa}\Big) \leq 2\exp(-c_{11}^* a_0^{2r_n} L_n^{-3} n^{1-2\kappa}) \qquad \text{(A47)}$$

for some positive constant c_{11}^*. Note that $\|n^{-1}\sum_{i=1}^n (X_{ij} - \mathbf{B}_{ij}^T \boldsymbol{\theta}_j^0)\mathbf{B}_{ij}\| \leq (r_n L_n)^{1/2} \max_{k,l} |N_{iklj}|$. This together with (A47) and the union bound of probability gives

$$\begin{aligned} &P\Big(\Big\|n^{-1}\sum_{i=1}^n (X_{ij} - \mathbf{B}_{ij}^T \boldsymbol{\theta}_j^0)\mathbf{B}_{ij}\Big\| \geq \frac{1}{8} c_5^* c_1^{*-1} a_0^{r_n} r_n^{1/2} L_n^{-1} n^{-\kappa}\Big) \\ &\leq 2(r_n L_n)\exp(-c_{11}^* a_0^{2r_n} L_n^{-3} n^{1-2\kappa}). \qquad \text{(A48)} \end{aligned}$$

Using Lemma A8 and (A48), we obtain

$$\begin{aligned} P\Big(\Xi_{n2j}^{(2)} \geq \frac{1}{4}c_5^* r_n n^{-\kappa}\Big) &\leq P\Big(\Big\|n^{-1}\sum_{i=1}^n (X_{ij} - \mathbf{B}_{ij}^T \boldsymbol{\theta}_j^0)\mathbf{B}_{ij}\Big\| \|\widehat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j^0\| \geq \frac{1}{8}c_5^* r_n n^{-\kappa}\Big) \\ &\leq P\Big(\Big\|n^{-1}\sum_{i=1}^n (X_{ij} - \mathbf{B}_{ij}^T \boldsymbol{\theta}_j^0)\mathbf{B}_{ij}\Big\| \geq \frac{1}{8} c_1^{*-1} c_5^* a_0^{r_n} r_n^{1/2} L_n^{-1} n^{-\kappa}\Big) \\ &\quad + P(\|\widehat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j^0\| \geq c_1^* a_0^{-r_n} r_n^{1/2} L_n) \\ &\leq 2(r_n L_n)\exp(-c_{11}^* a_0^{2r_n} L_n^{-3} n^{1-2\kappa}) \\ &\quad + [8(r_n L_n)^2 + 4 r_n L_n]\exp\big(-c_2^* a_0^{2r_n} r_n^{-2} L_n^{-3} n\big) \\ &\leq [8(r_n L_n)^2 + 6 r_n L_n]\exp\big(-c_{12}^* a_0^{2r_n} r_n^{-2} L_n^{-3} n\big) \qquad \text{(A49)} \end{aligned}$$

for some positive constant c_{12}^*. Therefore, combining (A43), (A44), (A46) and (A49), we can conclude the first result of Lemma A11. Moreover, the assumption that $r_n n^{-\kappa} = o(1)$ implies $c_5^* r_n n^{-\kappa} \leq \tilde{c}_1 \sigma_j^2$ for large n. Hence, the second result of Lemma A11 follows from the first result. □

Proof of Theorem 1. (i) We first show the first assertion. Let $H_{n1j} = \frac{1}{n}\sum_{i=1}^n \psi_\tau(Y_i - \mathbf{B}_{ij}^T \widehat{\boldsymbol{\alpha}}_j)(X_{ij} - \mathbf{B}_{ij}^T \widehat{\boldsymbol{\theta}}_j)$, $H_{n2j} = \sqrt{\widehat{\sigma}_j^2} = \widehat{\sigma}_j$, $h_{1j} = \mathrm{E}\{\psi_\tau(Y_i - \mathbf{B}_{ij}^T \boldsymbol{\alpha}_j^0)(X_{ij} - \mathbf{B}_{ij}^T \boldsymbol{\theta}_j^0)\}$ and $h_{2j} = \sigma_j$. Then,

$$\begin{aligned} &\big|\widehat{\varrho}_\tau(Y, X_j|\mathbf{X}_{\mathcal{S}_j}) - \varrho_\tau^*(Y, X_j|\mathbf{X}_{\mathcal{S}_j})\big| \\ &= H_{n2j}^{-1} h_{2j}^{-1}\big|(H_{n1j} - h_{1j})h_{2j} - h_{1j}(H_{n2j} - h_{2j})\big| \\ &\leq H_{n2j}^{-1}|H_{n1j} - h_{1j}| + H_{n2j}^{-1} h_{2j}^{-1}|h_{1j}||H_{n2j} - h_{2j}|. \qquad \text{(A50)} \end{aligned}$$

We first show that for some given constant $C_7 = (\sqrt{1-\tilde{c}_1}+1)^{-1}M_3^{-1/2}c_5^*$, there exists a positive constant c_{13}^* such that

$$\begin{aligned} P(|H_{n2j}-h_{2j}| \geq C_7 r_n n^{-\kappa}) \quad &\leq \quad 2[8(r_nL_n)^2+6r_nL_n+2]\exp(-c_{13}^* a_0^{2r_n} L_n^{-3} n^{1-2\kappa}) \\ &\quad +2[10(r_nL_n)^2+4r_nL_n]\exp(-c_{10}^* a_0^{4r_n} r_n^{-3} L_n^{-4} n^{1-\kappa}). \end{aligned} \tag{A51}$$

To this end, using the fact that $\sqrt{x}-\sqrt{y}=(x-y)/(\sqrt{x}+\sqrt{y})$ for positive x and y, we have

$$\begin{aligned} P(|H_{n2j}-h_{2j}| \geq C_7 r_n n^{-\kappa}) \quad &= \quad P\big(|\widehat{\sigma}_j^2-\sigma_j^2| \geq C_7 r_n n^{-\kappa}(\widehat{\sigma}_j+\sigma_j)\big) \\ &\leq \quad P\big(|\widehat{\sigma}_j^2-\sigma_j^2| \geq C_7 r_n n^{-\kappa}(\widehat{\sigma}_j+\sigma_j), \widehat{\sigma}_j^2 > (1-\tilde{c}_1)\sigma_j^2\big) \\ &\qquad +P\big(\widehat{\sigma}_j^2 \leq (1-\tilde{c}_1)\sigma_j^2\big) \\ &\leq \quad P\big(|\widehat{\sigma}_j^2-\sigma_j^2| \geq c_5^* r_n n^{-\kappa}\big)+P\big(|\widehat{\sigma}_j^2-\sigma_j^2| \geq \tilde{c}_1\sigma_j^2\big), \end{aligned}$$

where the last line uses condition (C5). This together with Lemma A11 implies (A51). Notice that since $C_7 r_n n^{-\kappa}=o(1)$, we have, for sufficiently large n, there exists a constant $\widetilde{c}_2 \in (0,1)$ such that $C_7 r_n n^{-\kappa} \leq \tilde{c}_2 M_3^{1/2} \leq \tilde{c}_2\sigma_j$. Thus,

$$\begin{aligned} &P(H_{n2j} \leq (1-\tilde{c}_2)h_{2j}) \leq P(|H_{n2j}-h_{2j}| \geq \tilde{c}_2\sigma_j) \leq P(|H_{n2j}-h_{2j}| \geq C_7 r_n n^{-\kappa}) \\ &\leq 2[8(r_nL_n)^2+6r_nL_n+2]\exp(-c_{13}^* a_0^{2r_n} L_n^{-3} n^{1-2\kappa}) \\ &\quad +2[10(r_nL_n)^2+4r_nL_n]\exp(-c_{10}^* a_0^{4r_n} r_n^{-3} L_n^{-4} n^{1-\kappa}). \end{aligned} \tag{A52}$$

Accordingly,

$$\begin{aligned} &P\big(H_{n2j}^{-1}|H_{n1j}-h_{1j}| \geq (1-\tilde{c}_2)^{-1}M_3^{-1/2}c_5^* r_n n^{-\kappa}\big) \\ &\leq P\big(|H_{n1j}-h_{1j}| \geq (1-\tilde{c}_2)^{-1}M_3^{-1/2}c_5^* r_n n^{-\kappa} H_{n2j}, H_{n2j} > (1-\tilde{c}_2)h_{2j}\big) \\ &\quad +P(H_{n2j} \leq (1-\tilde{c}_2)h_{2j}) \\ &\leq P\big(|H_{n1j}-h_{1j}| \geq c_5^* r_n n^{-\kappa}\big)+P(H_{n2j} \leq (1-\tilde{c}_2)h_{2j}) \\ &\leq 7\exp\big(-c_6^* a_0^{2r_n} r_n^2 n^{1-4\kappa}\big)+[8(r_nL_n)^2+4r_nL_n]\exp\big(-c_7^* a_0^{2r_n} r_n^{-2} L_n^{-3} n\big) \\ &\quad +2[8(r_nL_n)^2+6r_nL_n+2]\exp(-c_{13}^* a_0^{2r_n} L_n^{-3} n^{1-2\kappa}) \\ &\quad +2[10(r_nL_n)^2+4r_nL_n]\exp(-c_{10}^* a_0^{4r_n} r_n^{-3} L_n^{-4} n^{1-\kappa}) \\ &\leq 7\exp\big(-c_6^* a_0^{2r_n} r_n^2 n^{1-4\kappa}\big)+[44(r_nL_n)^2+20r_nL_n+2]\exp(-c_{14}^* a_0^{2r_n} L_n^{-3} n^{1-2\kappa}), \end{aligned} \tag{A53}$$

where $c_{14}^* = \min(c_7^*, c_{13}^*, c_{10}^*)$ and the last inequality is due to $a_0^{-2r_n} r_n^3 L_n n^{-\kappa}=o(1)$. Moreover, observe that, by the definition of $\boldsymbol{\theta}_j^0$ and Lemma A1,

$$\begin{aligned} |h_{1j}| \quad &= \quad \big|\mathrm{E}\{\psi_\tau(Y_i-\mathbf{B}_{ij}^T\boldsymbol{\alpha}_j^0)(X_{ij}-\mathbf{B}_{ij}^T\boldsymbol{\theta}_j^0)\}\big| = \big|\mathrm{E}\{\psi_\tau(Y_i-\mathbf{B}_{ij}^T\boldsymbol{\alpha}_j^0)X_{ij}\}\big| \\ &\leq \quad \max(\tau,1-\tau)\mathrm{E}\{|X_{ij}|\} \leq \max(\tau,1-\tau)\{\mathrm{E}(X_{ij}^2)\}^{1/2} \leq M_4, \end{aligned}$$

where $M_4 = \max(\tau, 1-\tau)\sqrt{2K_1K_2^2}$. So it follows from condition (C5) and (A51) and (A51) that

$$
\begin{aligned}
& P\big(H_{n2j}^{-1}h_{2j}^{-1}|h_{1j}||H_{n2j}-h_{2j}| \geq (1-\tilde{c}_2)^{-1}M_4M_3^{-3/2}(\sqrt{1-\tilde{c}_1}+1)^{-1}c_5^* r_n n^{-\kappa}\big) \\
& \leq P\big(|H_{n2j}-h_{2j}| \geq (1-\tilde{c}_2)^{-1}M_3^{-1}(\sqrt{1-\tilde{c}_1}+1)^{-1}c_5^* r_n n^{-\kappa}H_{n2j}\big) \\
& \leq P\big(|H_{n2j}-h_{2j}| \geq (1-\tilde{c}_2)^{-1}M_3^{-1}(\sqrt{1-\tilde{c}_1}+1)^{-1}c_5^* r_n n^{-\kappa}H_{n2j}, H_{n2j} > (1-\tilde{c}_2)h_{2j}\big) \\
& \quad + P(H_{n2j} \leq (1-\tilde{c}_2)h_{2j}) \\
& \leq P\big(|H_{n2j}-h_{2j}| \geq C_7 r_n n^{-\kappa}\big) + P(H_{n2j} \leq (1-\tilde{c}_2)h_{2j}) \\
& \leq 4[8(r_nL_n)^2 + 6r_nL_n + 2]\exp(-c_{13}^* a_0^{2r_n} L_n^{-3} n^{1-2\kappa}) \\
& \quad + 4[10(r_nL_n)^2 + 4r_nL_n]\exp(-c_{10}^* a_0^{4r_n} r_n^{-3} L_n^{-4} n^{1-\kappa}).
\end{aligned} \tag{A54}
$$

Put $C = (1-\tilde{c}_2)^{-1}c_5^* M_3^{-1/2}[1 + \frac{M_4}{M_3}(\sqrt{1-\tilde{c}_1}+1)^{-1}]/\sqrt{\tau(1-\tau)}$. Therefore, a direct application of (A53) and (A54) as well as the fact that $|x-y| \geq \big||x|-|y|\big|$, we can obtain

$$
\begin{aligned}
& \max_{1\leq j\leq p} P\Big(|\widehat{u}_j - u_j| \geq C r_n n^{-\kappa}\Big) \\
& \leq 7\exp\big(-c_6^* a_0^{2r_n} r_n^2 n^{1-4\kappa}\big) + [116(r_nL_n)^2 + 60r_nL_n + 10]\exp(-c_{14}^* a_0^{2r_n} L_n^{-3} n^{1-2\kappa}).
\end{aligned}
$$

This together with the union bound of probability proves the first assertion.

(ii) Next, we show the second assertion. By the choice of $\nu_n = \widetilde{C}_0 r_n n^{-\kappa}$ with $\widetilde{C}_0 \leq C_0/2$ and condition (C6), we have

$$
\begin{aligned}
& P(\mathcal{M}_* \subset \widehat{\mathcal{M}}) \geq P\Big(\min_{j\in\mathcal{M}_*} \widehat{u}_j > \nu_n\Big) = P\Big(\min_{j\in\mathcal{M}_*} u_j - \min_{j\in\mathcal{M}_*} \widehat{u}_j < \min_{j\in\mathcal{M}_*} u_j - \nu_n\Big) \\
& \geq P\Big(\min_{j\in\mathcal{M}_*}(\widehat{u}_j - u_j) > \nu_n - \min_{j\in\mathcal{M}_*} u_j\Big) \geq P\Big(\min_{j\in\mathcal{M}_*} u_j - \max_{j\in\mathcal{M}_*}|\widehat{u}_j - u_j| > \nu_n\Big) \\
& = 1 - P\Big(\max_{j\in\mathcal{M}_*}|\widehat{u}_j - u_j| \geq \min_{j\in\mathcal{M}_*} u_j - \nu_n\Big) \geq 1 - P\Big(\max_{j\in\mathcal{M}_*}|\widehat{u}_j - u_j| \geq \nu_n\Big) \\
& \geq 1 - s_n\{7\exp\big(-c_6^* a_0^{2r_n} r_n^2 n^{1-4\kappa}\big) + [116(r_nL_n)^2 + 60r_nL_n + 10]\exp(-c_{14}^* a_0^{2r_n} L_n^{-3} n^{1-2\kappa})\}.
\end{aligned}
$$

Thus, this completes the proof. □

Proof of Theorem 2. . By the assumption that $\sum_{i=1}^{p} u_j^* = O(n^{\varsigma})$ which implies that the size of $\{j : u_j^* > \widetilde{C}_0 r_n n^{-\kappa}\}$ cannot exceed $O(r_n^{-1}n^{\kappa+\varsigma})$. Thus, it follows that for any $\delta > 0$, on the set $\mathcal{A}_n = \Big\{\max_{1\leq j\leq p}|\widehat{u}_j - u_j^*| \leq \delta r_n n^{-\kappa}\Big\}$, the size of $\{j : \widehat{u}_j > 2\delta r_n n^{-\kappa}\}$ cannot exceed the size of $\{j : u_j^* > \delta r_n n^{-\kappa}\}$, which is bounded by $O(r_n^{-1}n^{\kappa+\varsigma})$. Then, taking $\delta = \widetilde{C}_0$ and $\nu_n = 2\widetilde{C}_0 r_n n^{-\kappa}$, we have

$$
P\Big(|\widehat{\mathcal{M}}| \leq O(r_n^{-1}n^{\kappa+\kappa})\Big) \geq P(\mathcal{A}_n) \geq 1 - P\Big(\max_{1\leq j\leq p}|\widehat{u}_j - u_j^*| > \widetilde{C}_0 r_n n^{-\kappa}\Big).
$$

Therefore, the desired conclusion follows from part (i) of Theorem 1. □

References

1. Tibshirani, R. Regression shrinkage and selection via the Lasso. *J. R. Stat. Soc. B* **1996**, *58*, 267–288. [CrossRef]
2. Fan, J.; Li, R. Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Am. Stat. Assoc.* **2001**, *96*, 1348–1360. [CrossRef]
3. Zou, H.; Li, R. One-step sparse estimate in nonconcave penalized likelihood models. *Ann. Stat.* **2008**, *36*, 1509–1533. [PubMed]
4. Zhang, C. Nearly unbiased variable selection under minimax concave penalty. *Ann. Stat.* **2010**, *38*, 894–942. [CrossRef]
5. Fan, J.; Lv, J. Sure independence screening for ultrahigh dimensional feature space. *J. R. Stat. Soc. B* **2008**, *70*, 849–911. [CrossRef]
6. Cheng, M.; Honda, T.; Li, J.; Peng, H. Nonparametric independence screening and structure identification for ultra-high dimensional longitudinal data. *Ann. Stat.* **2014**, *42*, 1819–1849. [CrossRef]

7. Fan, J.; Feng, Y.; Song, R. Nonparametric independence screening in sparse ultra-high-dimensional additive models. *J. Am. Stat. Assoc.* **2011**, *106*, 544–557. [CrossRef]
8. Fan, J.; Ma, Y.; Dai, W. Nonparametric independent screening in sparse ultra-high dimensional varying coefficient models. *J. Am. Stat. Assoc.* **2014**, *109*, 1270–1284. [CrossRef]
9. Fan, J.; Song, R. Sure independence screening in generalized linear models with NP-dimensionality. *Ann. Stat.* **2010**, *38*, 3567–3604. [CrossRef]
10. Liu, J.; Li, R.; Wu, R. Feature selection for varying coefficient models with ultrahigh-dimensional covariates. *J. Am. Stat. Assoc.* **2014**, *109*, 266–274. [CrossRef]
11. Xia, X.; Li, J.; Fu, B. Conditional quantile correlation learning for ultrahigh dimensional varying coefficient models and its application in survival analysis. *Statist. Sinica* **2019**, *29*, 645–669. [CrossRef]
12. Chang, J.; Tang, C.Y.; Wu, Y. Local independence feature screening for nonparametric and semiparametric models by marginal empirical likelihood. *Ann. Stat.* **2016**, *44*, 515–539. [CrossRef] [PubMed]
13. He, X.; Wang, L.; Hong, H. Quantile-adaptive model-free variable screening for high-dimensional heterogeneous data. *Ann. Stat.* **2013**, *41*, 342–369. [CrossRef]
14. Liu, W.; Ke, Y.; Liu, J.; Li, R. Model-free feature screening and FDR control with knockoff features. *J. Am. Stat. Assoc.* **2022**, *117*, 428–443. [CrossRef]
15. Li, J.; Zheng, Q.; Peng, L.; Huang, Z. Survival impact index and ultrahigh-dimensional model-free screening with survival outcomes. *Biometrics* **2016**, *72*, 1145–1154. [CrossRef]
16. Li, R.; Zhong, W.; Zhu, L. Feature screening via distance correlation learning. *J. Am. Stat. Assoc.* **2012**, *107*, 1129–1139. [CrossRef]
17. Ma, S.; Li, R.; Tsai, C. Variable screening via quantile partial correlation. *J. Am. Stat. Assoc.* **2017**, *112*, 650–663. [CrossRef]
18. Mai, Q.; Zou, H. The fused Kolmogorov filter: A nonparametric model-free screening method. *Ann. Stat.* **2015**, *43*, 1471–1497. [CrossRef]
19. Wu, Y.; Yin, G. Conditional qunatile screening in ultrahigh-dimensional heterogeneous data. *Biometrika* **2015**, *102*, 65–76. [CrossRef]
20. Zhou, T.; Zhu, L.; Xu, C.; Li, R. Model-free forward screening via cumulative divergence. *J. Am. Stat. Assoc.* **2020**, *115*, 1393–1405. [CrossRef]
21. Barut, E.; Fan, J.; Verhasselt, A. Conditional sure independence screening. *J. Am. Stat. Assoc.* **2016**, *111*, 1266–1277. [CrossRef] [PubMed]
22. Xia, X.; Jiang, B.; Li, J.; Zhang, W. Low-dimensional confounder adjustment and high-dimensional penalized estimation for survival analysis. *Lifetime Data Anal.* **2016**, *22*, 549–569. [CrossRef] [PubMed]
23. Chu, W.; Li, R.; Meimherr, M. Feature screening for time-varying coefficient models with ultrahigh-dimensional longitudinal data. *Ann. Appl. Stat.* **2016**, *10*, 596–617. [CrossRef] [PubMed]
24. Liu, Y.; Wang, Q. Model-free feature screening for ultrahigh-dimensional data conditional on some variables. *Ann. I. Stat. Math.* **2018**, *70*, 283–301. [CrossRef]
25. Wen, C.; Pan, W.; Huang, M.; Wang, X. Sure independence screening adjusted for confounding covariates with ultrahigh-dimensional data. *Statist. Sinica* **2018**, *28*, 293–317.
26. Li, R.; Liu, J.; Lou, L. Variable selection via partial correlation. *Statist. Sinica* **2017**, *27*, 983–996. [CrossRef]
27. Li, G.; Li, Y.; Tsai, C.L. Quantile correlations and quantile autoregressive modeling. *J. Am. Stat. Assoc.* **2015**, *110*, 246–261. [CrossRef]
28. Xia, X.; Li, J. Copula-based partial correlation screening: A joint and robust approach. *Statist. Sinica* **2021**, *31*, 421–447. [CrossRef]
29. De Boor, C. *A Practical Guide to Splines*; Springer: New York, NY, USA, 2001.
30. Huang, J.Z.; Wu, C.; Zhou, L. Varying-coefficient models and basis function approximation for the analysis of repeated measurements. *Biometrika* **2002**, *89*, 111–128. [CrossRef]
31. Xia, X. Model averaging prediction for nonparametric varying-coefficient models with B-spline smoothing. *Stat. Pap.* **2022**, *62*, 2885–2905. [CrossRef]
32. Huang, J.; Horowitz, J.; Wei, F. Variable selection in nonparametric additive models. *Ann. Stat.* **2010**, *38*, 2282–2313. [CrossRef] [PubMed]
33. Stone, C. Additive regression and other nonparametric models. *Ann. Stat.* **1985**, *13*, 689–705. [CrossRef]
34. Zhou, S.; Shen, X.; Wolfe, D.A. Local asymptotics for regression splines and confidence regions. *Ann. Stat.* **1998**, *26*, 1760–1782.
35. Kalisch, M.; Bühlmann, P. Estimating high-dimensional directed acyclic graphs with the PC-algorithm. *J. Mach. Learn. Res.* **2007**, *8*, 613–636.
36. Chin, K.; DeVries, S.; Fridlyand, J.; Spellman, P.T.; Roydasgupta, R.; Kuo, W.L.; Lapuk, A.; Neve, R.M.; Qian, Z.; Ryder, T.; et al. Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. *Cancer Cell* **2006**, *10*, 529–541. [CrossRef]
37. Zhou, Y.; Liu, J.; Hao, Z.; Zhu, L. Model-free conditional feature screening with exposure variables. *Stat. Its Interface* **2019**, *12*, 239–251. [CrossRef]
38. Chen, Z.; Fan, J.; Li, R. Error variance estimation in ultrahigh dimensional additive models. *J. Am. Stat. Assoc.* **2018**, *113*, 315–327. [CrossRef]
39. Van der Vaart, A.W.; Wellner, J.A. *Weak Convergence and Empirical Processes*; Springer: New York, NY, USA, 1996.
40. Leddoux, M.; Talagrand, M. *Probability in Banach Spaces: Isoperimetry and Processes*; Springer: Berlin, Germany, 1991.

41. Massart, P. About the constants in talagrands concentration inequalities for empirical processes. *Ann. Probab.* **2000**, *28*, 863–884. [CrossRef]
42. Koenker, R. *Quantile Regression*; Cambridge University Press: Cambridge, UK, 2005.