BIOMETRICS 000, 000–000 DOI: 10.1111/j.1541-0420.2005.00454.x
000 0000

# Supporting Information for "Estimating and Inferring the Maximum Degree of Stimulus-Locked Time-Varying Brain Connectivity Networks" by Kean Ming Tan, Junwei Lu, Tong Zhang, and Han Liu

SUMMARY: In Web Appendix A, we generalize the inferential procedure in Algorithm 1 in the main manuscript to testing general topological structure of a time-varying graph. The theoretical results for the general framework are presented in Theorems S1 and S2. In Web Appendix B, we present a $U$-statistic type estimator for the case when there are multiple subjects. In Web Appendix C, we define some notation that will be used throughout the supplementary. The proofs of the results in 5.1 of the main manuscript are in Web Appendix D. The proof of Theorem 2 in the main manuscript is in Web Appendix F. Theorems 3 and 4 in the main manuscript are special cases of Theorems S1 and S2, respectively. Therefore, their proofs follow directly and are ommited. We collect the proofs of Theorems S1 and S2 in Web Appendix G and Web Appendix H, respectively. In Web Appendix I, we collect a series of lemmas on covering number for various functions. Finally, some technical lemmas on empirical process are presented in Web Appendix J.

## Web Appendix A. Inference on Topological Structure of Time-Varying Graph

In this section, we generalize Algorithm 1 in the main manuscript to testing various graph structures that satisfy the *monotone graph property*. In Web Appendix A.1, we briefly introduce some concepts on graph theory. These include the notion of isomorphism, graph property, monotone graph property, and critical edge set. In Web Appendix A.2, we provide a test statistic and an estimate of the quantile of the proposed test statistic using the Gaussian multiplier bootstrap. We then develop an algorithm to test the dynamic topological structure of a time-varying graph which satisfies the monotone graph property.

### Web Appendix A.1 *Graph Theory*

Let $G = (V, E)$ be an undirected graph where $V = \{1, \ldots, d\}$ is a set of nodes and $E \subseteq V \times V$ is a set of edges connecting pairs of nodes. Let $\mathcal{G}$ be the set of all graphs with the same number of nodes. For any two graphs $G = (V, E)$ and $G' = (V, E')$, we write $G \subseteq G'$ if $G$ is a subgraph of $G'$, that is, if $E \subseteq E'$. We start with introducing some concepts on graph theory (see, for instance, Chapter 4 of Lovász, 2012).

Definition S1: Two graphs $G = (V, E)$ and $G' = (V, E')$ are said to be isomorphic if there exists permutations $\pi : V \to V$ such that $(j, k) \in E$ if and only if $\{\pi(j), \pi(k)\} \in E'$.

The notion of isomorphism is used in the graph theory literature to quantify whether two graphs have the same topological structure, up to any permutation of the vertices (see Chapter 1.2 of Bondy and Murty, 1976). We provide two concrete examples on the notion of isomorphism in Figure S1.

[Figure 1 about here.]

Next, we introduce the notion of graph property. A graph property is a property of graphs that depends only on the structure of the graphs, that is, a graph property is invariant under permutation of vertices. A formal definition is given as follows.

Definition S2: For two graphs $G$ and $G'$ that are isomorphic, a graph property is a function $\mathcal{P}: \mathcal{G} \rightarrow \{0,1\}$ such that $\mathcal{P}(G) = \mathcal{P}(G')$. A graph $G$ satisfies the graph property $\mathcal{P}$ if $\mathcal{P}(G) = 1$.

Some examples of graph property are that the graph is connected, the graph has no more than $k$ connected components, the maximum degree of the graph is larger than $k$, the graph has no more than $k$ isolated nodes, the graph contains a clique of size larger than $k$, and the graph contains a triangle. For instance, the two graphs in Figures S1(i) and S1(ii) are isomorphic and satisfy the graph property of being connected.

Definition S3: For two graphs $G \subseteq G'$, a graph property $\mathcal{P}$ is monotone if $\mathcal{P}(G) = 1$ implies that $\mathcal{P}(G') = 1$.

In other words, we say that a graph property is monotone if the graph property is preserved under the addition of new edges. Many graph property that are of interest such as those given in the paragraph immediately after Definition S2 are monotone. In Figure S2, we present several examples of graph property that are monotone by showing that adding additional edges to the graph does not change the graph property. For instance, we see from Figure S2(a) that the existing graph with gray edges are connected. Adding the red edges to the existing graph, the graph remains connected and therefore the graph property is monotone. Another example is the graph with maximum degree at least three as in Figure S2(c). We see that adding the red dash edges to the graph preserves the graph property of having maximum degree at least three.

[Figure 2 about here.]

For a given graph $G = (V, E)$, we define the class of edge sets satisfying the graph property $\mathcal{P}$ as

$$\mathscr{P} = \{E \subseteq V \times V \mid \mathcal{P}(G) = 1\}. \tag{S1}$$

Finally, we introduce the notion of critical edge set in the following definition.

Definition S4: Given any edge set $E \subseteq V \times V$, we define the critical edge set of $E$ for a given monotone graph property $\mathcal{P}$ as

$$\mathcal{C}(E, \mathcal{P}) = \{e \mid e \notin E, \text{ there exists } E' \supseteq E \text{ such that } E' \in \mathscr{P} \text{ and } E' \backslash \{e\} \notin \mathscr{P}\}. \tag{S2}$$

For a given monotone graph property $\mathcal{P}$, the critical edge set is the set of edges that will change the graph property of the graph once added to the existing graph. We provide two examples in Figure S3. Suppose that $\mathcal{P}$ is the graph property of being connected. In Figure S3(a), we see that the graph is not connected, and thus $\mathcal{P}(G) = 0$. Adding any of the red dash edges in Figure S3(b) changes $\mathcal{P}(G) = 0$ to $\mathcal{P}(G) = 1$.

[Figure 3 about here.]

Web Appendix A.2 *An Algorithm for Topological Inference*

Throughout the rest of the paper, we denote $G(z) = \{V, E(z)\}$ as the graph at $Z = z$. We consider hypothesis testing problem of the form

$$\begin{aligned} &H_0 : \mathcal{P}\{G(z)\} = 0 \text{ for all } z \in [0, 1] \\ &H_1 : \text{there exists a } z_0 \in [0, 1] \text{ such that } \mathcal{P}\{G(z_0)\} = 1, \end{aligned} \tag{S3}$$

where $G(\cdot)$ is the true underlying graph and $\mathcal{P}$ is a given monotone graph property as defined in Definition S3. We provide two concrete examples of the hypothesis testing problem in (S3).

Example S1: Number of connected components:

$H_0$ : for all $z \in [0, 1]$, the number of connected components is greater than $k$,

$H_1$ : there exists a $z_0 \in [0, 1]$ such that the number of connected components is not greater than $k$.

Example S2: Maximum degree of the graph:

$H_0$ : for all $z \in [0, 1]$, the maximum degree of the graph is not greater than $k$,

$H_1$ : there exists a $z_0 \in [0, 1]$ such that the maximum degree of the graph is greater than $k$.

We now propose an algorithm to test the topological structure of a time-varying graph. The proposed algorithm is very general and is able to test the hypothesis problem of the form in (S3). Our proposed algorithm is motivated by the step-down algorithm in Romano and Wolf (2005) for testing multiple hypothesis simultaneously. The main crux of our algorithm is as follows. By Definition S4, the critical edge set $\mathcal{C}\{E_{t-1}(z), \mathcal{P}\}$ contains edges that may change the graph property from $\mathcal{P}\{G(z)\} = 0$ to $\mathcal{P}\{G(z)\} = 1$. Thus, at the $t$-th iteration of the proposed algorithm, it suffices to test whether the edges on the critical edge set $\mathcal{C}\{E_{t-1}(z), \mathcal{P}\}$ are rejected. Let $E_t(z) = E_{t-1}(z) \cup \mathcal{R}(z)$, where $\mathcal{R}(z)$ is the rejected edge set from the critical edge set $\mathcal{C}\{E_{t-1}(z), \mathcal{P}\}$. Since $\mathcal{P}$ is a monotone graph property, if there exists a $z_0 \in [0, 1]$ such that $E_t(z_0) \in \mathscr{P}$, we directly reject the null hypothesis $H_0 : \mathcal{P}\{G(z)\} = 0$ for all $z$. This is due to the definition of monotone graph property that adding more edges does not change the graph property. If $E_t(z_0) \notin \mathscr{P}$, we repeat this process until the null hypothesis is rejected or no more edges in the critical edge set are rejected. We summarize the procedure in Algorithm S1.

Finally, we generalize the theoretical results in Theorems 3 and 4 to the general testing procedure in Algorithm S1. Given a monotone graph property $\mathcal{P}$, let

$$\mathcal{G}_0 = (\boldsymbol{\Theta}(\cdot) \in \mathcal{U}_{s,M} \mid \mathcal{P}[G\{\boldsymbol{\Theta}(z)\}] = 0 \text{ for all } z \in [0, 1]).$$

We now show that the type I error of the proposed inferential method in Algorithm S1 can be controlled at a pre-specified level $\alpha$.

Theorem S1: *Under the same conditions in Theorem 2, we have*

$$\lim_{n\to\infty} \sup_{\boldsymbol{\Theta}(\cdot)\in\mathcal{G}_0} P_{\boldsymbol{\Theta}(\cdot)}\left(\psi_\alpha = 1\right) \leqslant \alpha.$$

In order to study the power analysis for testing graph structure that satisfies the monotone graph property, we define signal strength of a precision matrix $\boldsymbol{\Theta}$ as

$$\mathrm{Sig}(\boldsymbol{\Theta}) := \max_{E' \subseteq E(\boldsymbol{\Theta}), \mathcal{P}(E')=1} \min_{e \in E'} |\boldsymbol{\Theta}_e|. \tag{S4}$$

**Algorithm S1** Dynamic skip-down method.

**Input:** A monotone graph property $\mathcal{P}$; $\widehat{\mathbf{\Theta}}^{\mathrm{de}}(z)$ for $z \in [0,1]$.

**Initialize:** $t = 1$; $E_0(z) = \emptyset$ for $z \in [0,1]$.

**Repeat:**

(1) Compute the critical edge set $\mathcal{C}\{E_{t-1}(z), \mathcal{P}\}$ for $z \in [0,1]$ and the conditional quantile $c\{1-\alpha, \mathcal{C}(E_{t-1}, \mathcal{P})\} = \inf\Big(t \in \mathbb{R} \mid P[T^B_{\mathcal{C}(E_{t-1},\mathcal{P})} \leqslant t \mid \{(\boldsymbol{X}_i, \boldsymbol{Y}_i, Z_i)\}_{i\in[n]}] \geqslant 1-\alpha\Big)$, where $T^B_{\mathcal{C}(E_{t-1},\mathcal{P})}$ is the bootstrap statistic defined in (10) with the maximum taken over the edge set $\mathcal{C}\{E_{t-1}(z), \mathcal{P}\}$.

(2) Construct the rejected edge set

$$\mathcal{R}(z) = \left[ e \in \mathcal{C}\{E_{t-1}(z), \mathcal{P}\} \mid \sqrt{nh} \cdot |\widehat{\mathbf{\Theta}}^{\mathrm{de}}_e(z)| \cdot \sum_{i\in[n]} K_h(Z_i - z)/n > c\{1-\alpha, \mathcal{C}(E_{t-1}, \mathcal{P})\} \right].$$

(3) Update the rejected edge set $E_t(z) \leftarrow E_{t-1}(z) \cup \mathcal{R}(z)$ for $z \in [0,1]$.

(4) $t \leftarrow t+1$.

**Until:** There exists a $z_0 \in [0,1]$ such that $E_t(z_0) \in \mathscr{P}$, or $E_t(z) = E_{t-1}(z)$ for $z \in [0,1]$.

**Output:** $\psi_\alpha = 1$ if there exists a $z_0 \in [0,1]$ such that $E_t(z_0) \in \mathscr{P}$ and $\psi_\alpha = 0$ otherwise.

Under $H_1$ : there exists a $z_0 \in [0,1]$ such that $\mathcal{P}\{G(z_0)\} = 1$, we define the parameter space

$$\mathcal{G}_1(\theta; \mathcal{P}) = \Big(\mathbf{\Theta}(\cdot) \in \mathcal{U}_{s,M} \,\Big|\, \mathcal{P}[G\{\mathbf{\Theta}(z_0)\}] = 1 \text{ and } \mathrm{Sig}\{\mathbf{\Theta}(z_0)\} \geqslant \theta \text{ for some } z_0 \in [0,1]\Big). \tag{S5}$$

Again, we emphasize that the signal strength defined in (S4) is weaker than the typical minimal signal strength for testing a single edge in a graph $\min_{e\in E(\mathbf{\Theta})} |\mathbf{\Theta}_e|$. $\mathrm{Sig}(\mathbf{\Theta})$ only requires that there exists a subgraph satisfying the property of interest such that the minimal signal strength on that subgraph is above certain level. For example, for $\mathcal{P}(G) = 1$ if and only if $G$ is connected, it suffices for $\mathbf{\Theta}$ belongs to $\mathcal{G}_1(\theta; \mathcal{P})$ if the minimal signal strength on a spanning tree is larger than $\theta$. The following theorem presents the power analysis of our test.

Theorem S2: *Assume that the same conditions in Theorem 2 hold and select the smoothing parameter $h = o(1/n^{-1/5})$. Assume that $\theta \geqslant C\sqrt{\log(dn)}/n^{2/5}$ for some sufficiently large constant $C$. Under the alternative hypothesis $H_1 : \mathcal{P}(G) = 1$ in (S3), we have*

$$\lim_{n\to\infty} \inf_{\boldsymbol{\Theta}\in\mathcal{G}_1(\theta;\mathcal{P})} \mathbb{P}_{\boldsymbol{\Theta}}(\psi_\alpha = 1) = 1 \tag{S6}$$

*for any fixed $\alpha \in (0,1)$.*

Thus, we have shown in Theorem S2 that the power of the proposed inferential method increases to one asymptotically.

## Web Appendix B. A $U$-Statistic Type Estimator

The main manuscript primarily concerns the case when there are two subjects. In this section, we present a $U$-statistic type inter-subject covariance to accommodate the case when there are more than two subjects. First, we note that the same natural stimuli is given to all subjects. This motivates the following statistical model for each $Z = z$:

$$\boldsymbol{X}^{(\ell)} = \boldsymbol{S} + \boldsymbol{E}^{(\ell)}, \quad \boldsymbol{S}|Z = z \sim N_d\{\mathbf{0}, \boldsymbol{\Sigma}(z)\}, \quad \boldsymbol{E}^{(\ell)}|Z = z \sim N_d\{\mathbf{0}, \boldsymbol{L}^{(\ell)}(z)\},$$

where $\boldsymbol{X}^{(\ell)}$, $\boldsymbol{E}^{(\ell)}$, and $\boldsymbol{L}^{(\ell)}(z)$ are the data, subject specific effect, and the covariance matrix for the subject specific effect for the $\ell$th subject, respectively. Suppose that there $N$ subjects. Then, the following $U$-statistic type inter-subject covariance matrix can be constructed to estimate $\boldsymbol{\Sigma}(z)$:

$$\widehat{\boldsymbol{\Sigma}}_U(z) = \frac{1}{\binom{N}{2}} \sum_{1\leqslant \ell<\ell'\leqslant N} \left[ \frac{\sum_{i\in[n]} K_h(Z_i - z)\boldsymbol{X}_i^{(\ell)}\{\boldsymbol{X}_i^{(\ell')}\}^T}{\sum_{i\in[n]} K_h(Z_i - z)} \right]. \tag{S7}$$

We leave the theoretical analysis of the above estimator for future work.

## Web Appendix C. Preliminaries

In this section, we define some notation that will be used throughout the Appendix. Let $[n]$ denote the set $\{1,\ldots,n\}$ and let $[d]$ denote the set $\{1,\ldots,d\}$. For two scalars $a, b$, we define $a \vee b = \max(a,b)$. We denote the $\ell_q$-norm for the vector $\mathbf{v}$ as $\|\mathbf{v}\|_q = (\sum_{j\in[d]} |v_j|^q)^{1/q}$

for $1 \leqslant q < \infty$. In addition, we let $\mathrm{supp}(\mathbf{v}) = \{j : v_j \neq 0\}$, $\|\mathbf{v}\|_0 = |\mathrm{supp}(\mathbf{v})|$, and $\|\mathbf{v}\|_\infty = \max_{j\in[d]} |v_j|$, where $|\mathrm{supp}(\mathbf{v})|$ is the number of non-zero elements in $\mathbf{v}$. For a matrix $\mathbf{A} \in \mathbb{R}^{n_1\times n_2}$, we denote the $j$th column as $\mathbf{A}_j$. We denote the Frobenius norm of $\mathbf{A}$ by $\|\mathbf{A}\|_F^2 = \sum_{i\in[n_1]}\sum_{j\in[n_2]} A_{ij}^2$, the max norm $\|\mathbf{A}\|_{\max} = \max_{i\in[n_1],j\in[n_2]} |A_{ij}|$, and the operator norm $\|\mathbf{A}\|_2 = \sup_{\|\mathbf{v}\|_2=1} \|\mathbf{A}\mathbf{v}\|_2$. Given a function $f$, let $\dot{f}$ and $\ddot{f}$ be the first and second-order derivatives, respectively. For $1 \leqslant p < \infty$, let $\|f\|_p = (\int f^p)^{1/p}$ denote the $L_p$ norm of $f$ and let $\|f\|_\infty = \sup_x |f(x)|$. The total variation of $f$ is defined as $\|f\|_{\mathrm{TV}} = \int |\dot{f}|$. We use the Landau symbol $a_n = \mathcal{O}(b_n)$ to indicate the existence of a constant $C > 0$ such that $a_n \leqslant C \cdot b_n$ for two sequences $a_n$ and $b_n$. We write $a_n = o(b_n)$ if $\lim_{n\to\infty} a_n/b_n \to 0$. Let $C, C_1, C_2, \ldots$ be generic constants whose values may vary from line to line.

Let

$$\mathbb{P}_n(f) = \frac{1}{n}\sum_{i\in[n]} f(X_i) \qquad \text{and} \qquad \mathbb{G}_n(f) = \sqrt{n}\cdot[\mathbb{P}_n(f) - \mathbb{E}\{f(X_i)\}]. \tag{S8}$$

For notational convenience, for fixed $j, k \in [d]$, let

$$g_{z,jk}(Z_i, X_{ij}, Y_{ik}) = K_h(Z_i - z)X_{ij}Y_{ik}, \qquad w_z(Z_i) = K_h(Z_i - z), \tag{S9}$$

$$q_{z,jk}(Z_i, X_{ij}, Y_{ik}) = g_{z,jk}(Z_i, X_{ij}, Y_{ik}) - \mathbb{E}\{g_{z,jk}(Z, X_j, Y_k)\}, \tag{S10}$$

and let

$$k_z(Z_i) = w_z(Z_i) - \mathbb{E}\{w_z(Z)\}. \tag{S11}$$

Recall from 5 that $K(\cdot)$ can be any symmetric kernel function that satisfies (12) and that $K_h(Z_i - z) = K\{(Z_i - z)/h\}/h$. By the definition of $\widehat{\mathbf{\Sigma}}(z)$ in (S7), we have

$$\widehat{\mathbf{\Sigma}}_{jk}(z) = \frac{\sum_{i\in[n]} g_{z,jk}(Z_i, X_{ij}, Y_{ik})}{\sum_{i\in[n]} w_z(Z_i)} = \frac{\mathbb{P}_n(g_{z,jk})}{\mathbb{P}_n(w_z)}. \tag{S12}$$

In addition, let

$$J^{(1)}_{z,jk}(Z_i, \boldsymbol{X}_i, \boldsymbol{Y}_i) = \sqrt{h}\cdot\{\mathbf{\Theta}_j(z)\}^T\cdot\left[K_h(Z_i - z)\boldsymbol{X}_i\boldsymbol{Y}_i^T - \mathbb{E}\left\{K_h(Z - z)\boldsymbol{X}\boldsymbol{Y}^T\right\}\right]\cdot\mathbf{\Theta}_k(z), \tag{S13}$$

$$J^{(2)}_{z,jk}(Z_i) = \sqrt{h} \cdot \{\boldsymbol{\Theta}_j(z)\}^T \cdot [K_h(Z_i - z) - \mathbb{E}\{K_h(Z - z)\}] \cdot \boldsymbol{\Sigma}(z) \cdot \boldsymbol{\Theta}_k(z), \tag{S14}$$

$$J_{z,jk}(Z_i, \boldsymbol{X}_i, \boldsymbol{Y}_i) = J^{(1)}_{z,jk}(Z_i, \boldsymbol{X}_i, \boldsymbol{Y}_i) - J^{(2)}_{z,jk}(Z_i), \tag{S15}$$

and let

$$W_{z,jk}(Z_i, X_{ij}, Y_{ik}) = \sqrt{h} \cdot \{K_h(Z_i - z)X_{ij}Y_{ik} - K_h(Z_i - z)\boldsymbol{\Sigma}_{jk}(z)\}. \tag{S16}$$

For two functions $f$ and $g$, we define its convolution as

$$(f * g)(x) = \int f(x - z)g(z)dz. \tag{S17}$$

In our proofs, we will use the following property of the derivative of a convolution

$$\frac{\partial}{\partial x}(f * g) = \frac{\partial f}{\partial x} * g. \tag{S18}$$

Finally, our proofs use the following inequality

$$\int_0^{b_1} \sqrt{\log(b_2/\epsilon)}d\epsilon \leqslant \sqrt{b_1} \cdot \sqrt{\int_0^{b_1} \log(b_2/\epsilon)d\epsilon} = b_1 \cdot \sqrt{1 + \log(b_2/b_1)}, \tag{S19}$$

where the first inequality holds by an application of Jensen's inequality.

## Web Appendix D. Proof of Results in 5.1

In this section, we establish the uniform rate of convergence for $\widehat{\boldsymbol{\Sigma}}(z)$ and $\widehat{\boldsymbol{\Theta}}(z)$ over $z \in [0, 1]$. To prove Theorem 1, we first observe that

$$\sup_{z\in[0,1]} \left\|\widehat{\boldsymbol{\Sigma}}(z) - \boldsymbol{\Sigma}(z)\right\|_{\max} \leqslant \sup_{z\in[0,1]} \max_{j,k\in[d]} \left|\widehat{\boldsymbol{\Sigma}}_{jk}(z) - \mathbb{E}\left\{\widehat{\boldsymbol{\Sigma}}_{jk}(z)\right\}\right| + \sup_{z\in[0,1]} \max_{j,k\in[d]} \left|\mathbb{E}\left\{\widehat{\boldsymbol{\Sigma}}_{jk}(z)\right\} - \boldsymbol{\Sigma}_{jk}(z)\right|. \tag{S20}$$

The first term is known as the variance term and the second term is known as the bias term in the kernel smoothing literature (see, for instance, Chapter 2 of Pagan and Ullah, 1999). Both the variance and bias terms involve evaluating the quantity $\mathbb{E}\{\widehat{\boldsymbol{\Sigma}}_{jk}(z)\}$. From (S12), we see that $\widehat{\boldsymbol{\Sigma}}_{jk}(z)$ involves the quotient of two averages and it is not straightforward to evaluate its expectation. The following lemma quantifies $\mathbb{E}\{\widehat{\boldsymbol{\Sigma}}_{jk}(z)\}$ in terms of the expectations of its numerator and its denominator.

Lemma S1: *Under the following conditions*

$$\left|\frac{\mathbb{G}_n(w_z)}{\sqrt{n}\cdot\mathbb{E}\left\{\mathbb{P}_n(w_z)\right\}}\right| < 1 \qquad \text{and} \qquad \mathbb{E}\left\{\mathbb{P}_n(w_z)\right\} \neq 0, \tag{S21}$$

*we have*

$$\mathbb{E}\left\{\widehat{\boldsymbol{\Sigma}}_{jk}(z)\right\} = \frac{\mathbb{E}\left\{\mathbb{P}_n(g_{z,jk})\right\}}{\mathbb{E}\left\{\mathbb{P}_n(w_z)\right\}} + \frac{1}{n}\mathcal{O}\left[\mathbb{E}\Big\{\mathbb{G}_n(w_z)\cdot\mathbb{G}_n(g_{z,jk})\Big\} + \mathbb{E}\left\{\mathbb{G}_n^2(g_{z,jk})\right\}\right]. \tag{S22}$$

We note that (S22) only holds under the two conditions in (S21). In the proof of Theorem 1, we will show that the two conditions in (S21) hold for $n$ sufficiently large. To obtain upper bounds for the bias and variance terms in (S20), we use the following intermediate lemmas.

Lemma S2: *Assume that $h = o(1)$. Under Assumptions 1-2, we have*

$$\sup_{z\in[0,1]}\max_{j,k\in[d]}\Big|\mathbb{E}\{\mathbb{P}_n(g_{z,jk})\} - f_Z(z)\boldsymbol{\Sigma}_{jk}(z)\Big| = \mathcal{O}(h^2), \tag{S23}$$

$$\sup_{z\in[0,1]}\Big|\mathbb{E}\{\mathbb{P}_n(w_z)\} - f_Z(z)\Big| = \mathcal{O}(h^2), \tag{S24}$$

$$\sup_{z\in[0,1]}\max_{j,k\in[d]}\frac{1}{n}\left|\mathbb{E}\Big\{\mathbb{G}_n(g_{z,jk})\cdot\mathbb{G}_n(w_z)\Big\}\right| = \mathcal{O}\left(\frac{1}{nh}\right), \tag{S25}$$

*and*

$$\sup_{z\in[0,1]}\frac{1}{n}\mathbb{E}\Big\{\mathbb{G}_n^2(w_z)\Big\} = \mathcal{O}\left(\frac{1}{nh}\right). \tag{S26}$$

Lemma S3: *Assume that $h = o(1)$ and $\log^2(d/h)/(nh) = o(1)$. Under Assumptions 1-2, there exists a universal constant $C > 0$ such that*

$$\sup_{z\in[0,1]}\max_{j,k\in[d]}\Big|\mathbb{G}_n(w_z)\vee\mathbb{G}_n(g_{z,jk})\Big| \leqslant C\cdot\sqrt{\frac{\log(d/h)}{h}}, \tag{S27}$$

*with probability at least $1 - 3/d$.*

The proofs of Lemmas S1-S3 are deferred to Sections S1-S3, respectively. We now provide a proof of Theorem 1.

Web Appendix D.1 *Proof of Theorem 1*

Recall from (S20) that

$$\sup_{z\in[0,1]}\left\|\widehat{\boldsymbol{\Sigma}}(z)-\boldsymbol{\Sigma}(z)\right\|_{\max}\leqslant \sup_{z\in[0,1]}\max_{j,k\in[d]}\left|\widehat{\boldsymbol{\Sigma}}_{jk}(z)-\mathbb{E}\left\{\widehat{\boldsymbol{\Sigma}}_{jk}(z)\right\}\right|+\sup_{z\in[0,1]}\max_{j,k\in[d]}\left|\mathbb{E}\left\{\widehat{\boldsymbol{\Sigma}}_{jk}(z)\right\}-\boldsymbol{\Sigma}_{jk}(z)\right|$$
$$=I_1+I_2.$$

It suffices to obtain upper bounds for $I_1$ and $I_2$.

We first verify that the two conditions in (S21) hold. By Lemma S2, we have

$$\left|\mathbb{E}\{\mathbb{P}_n(w_z)\}\right|=\mathcal{O}(h^2)+f_Z(z)\geqslant \underline{f}_Z(z)>0,$$

where the last inequality follows from Assumption 1. Moreover,

$$\begin{aligned}\left|\frac{\mathbb{G}_n(w_z)}{\sqrt{n}\cdot\mathbb{E}\left\{\mathbb{P}_n(w_z)\right\}}\right| &\leqslant C\cdot\frac{1}{\sqrt{n}}|\mathbb{G}_n(w_z)|\cdot\frac{1}{f_Z(z)+\mathcal{O}(h^2)}\\ &\leqslant C_1\cdot\sqrt{\frac{\log(d/h)}{nh}}\cdot\frac{1}{f_Z(z)+\mathcal{O}(h^2)}\\ &<1,\end{aligned}$$

for sufficiently large $n$, where the first inequality is obtained by an application of Lemma S2, the second inequality is obtained by an application of Lemma S3, and the last inequality is obtained by the scaling assumptions $h=o(1)$ and $\log(d/h)/(nh)=o(1)$.

**Upper bound for $I_1$:** By (S35) in the proof of Lemma S1, we have

$$\widehat{\boldsymbol{\Sigma}}_{jk}(z)=\frac{\mathbb{G}_n(g_{z,jk})}{\sqrt{n}\mathbb{E}\left\{\mathbb{P}_n(w_z)\right\}}+\frac{\mathbb{E}\{\mathbb{P}_n(g_{z,jk})\}}{\mathbb{E}\left\{\mathbb{P}_n(w_z)\right\}}-\frac{\mathbb{G}_n(w_z)\mathbb{E}\{\mathbb{P}_n(g_{z,jk})\}}{\sqrt{n}\mathbb{E}^2\{\mathbb{P}_n(w_z)\}}+\frac{1}{n}\mathcal{O}\left[\left\{\mathbb{G}_n(w_z)\mathbb{G}_n(g_{j,zk})\right\}+\mathbb{G}_n^2(g_{z,jk})\right].$$

Thus, by Lemma S1, we have

$$\begin{aligned}I_1&=\sup_{z\in[0,1]}\max_{j,k\in[d]}\left|\underbrace{\frac{\mathbb{G}_n(g_{z,jk})}{\sqrt{n}\cdot\mathbb{E}\left\{\mathbb{P}_n(w_z)\right\}}}_{I_{11}}-\underbrace{\frac{\mathbb{G}_n(w_z)\cdot\mathbb{E}\{\mathbb{P}_n(g_{z,jk})\}}{\sqrt{n}\cdot\mathbb{E}^2\{\mathbb{P}_n(w_z)\}}}_{I_{12}}+I_{13}\right| \\ &\leqslant \sup_{z\in[0,1]}\max_{j,k\in[d]}\left\{|I_{11}|+|I_{12}|+|I_{13}|\right\},\end{aligned}\tag{S28}$$

where $I_{13}=\mathcal{O}[\{\mathbb{G}_n(w_z)\mathbb{G}_n(g_{j,zk})\}+\mathbb{G}_n^2(g_{z,jk})+\mathbb{E}\{\mathbb{G}_n(w_z)\cdot\mathbb{G}_n(g_{j,zk})\}+\mathbb{E}\{\mathbb{G}_n^2(g_{z,jk})\}]/n.$

We now provide upper bounds for $I_{11}$, $I_{12}$, and $I_{13}$. By an application of Lemmas S2 and

S3, we obtain

$$\sup_{z\in[0,1]}\max_{j,k\in[d]}|I_{11}| \leqslant n^{-1/2}\cdot \sup_{z\in[0,1]}\max_{j,k\in[d]}\left|\frac{\mathbb{G}_n(g_{z,jk})}{f_Z(z)+\mathcal{O}(h^2)}\right| \leqslant C\cdot\sqrt{\frac{\log(d/h)}{nh}}. \tag{S29}$$

Similarly, we have

$$\sup_{z\in[0,1]}\max_{j,k\in[d]}|I_{12}| \leqslant n^{-1/2}\cdot \sup_{z\in[0,1]}\max_{j,k\in[d]}\left|\frac{\mathbb{G}_n(g_{z,jk})\{f_Z(z)\boldsymbol{\Sigma}_{jk}(z)+\mathcal{O}(h^2)\}}{\{f_Z(z)+\mathcal{O}(h^2)\}^2}\right| \leqslant C\cdot\sqrt{\frac{\log(d/h)}{nh}}. \tag{S30}$$

For $I_{13}$, we have

$$\begin{aligned}
\sup_{z\in[0,1]}\max_{j,k\in[d]}|I_{13}| &\leqslant \sup_{z\in[0,1]}\max_{j,k\in[d]}\left|\frac{1}{n}\mathcal{O}\left[\Big\{\mathbb{G}_n(w_z)\cdot\mathbb{G}_n(g_{z,jk})\Big\}+\mathbb{G}_n^2(g_{z,jk})\right]\right| + \mathcal{O}\left(\frac{1}{nh}\right)\\
&\leqslant C\cdot\frac{\log(d/h)}{nh}+\mathcal{O}\left(\frac{1}{nh}\right)\\
&\leqslant C\cdot\frac{\log(d/h)}{nh},
\end{aligned} \tag{S31}$$

where the first and second inequalities follow from Lemmas S2 and S3, respectively. Combining (S29), (S30), and (S31), we have

$$I_1 \leqslant C\cdot\sqrt{\frac{\log(d/h)}{nh}}, \tag{S32}$$

with probability at least $1-3/d$.

**Upper bound for $I_2$:** By Lemmas S1 and S2, we have

$$\begin{aligned}
I_2 &= \sup_{z\in[0,1]}\max_{j,k\in[d]}\left|\frac{\mathbb{E}\{\mathbb{P}_n(g_{z,jk})\}}{\mathbb{E}\{\mathbb{P}_n(w_z)\}}-\boldsymbol{\Sigma}_{jk}(z)+\frac{1}{n}\mathcal{O}\left[\mathbb{E}\Big\{\mathbb{G}_n(w_z)\cdot\mathbb{G}_n(g_{z,jk})\Big\}+\mathbb{E}\left\{\mathbb{G}_n^2(g_{z,jk})\right\}\right]\right|\\
&\leqslant \sup_{z\in[0,1]}\max_{j,k\in[d]}\left|\frac{f_Z(z)\boldsymbol{\Sigma}_{jk}(z)+\mathcal{O}(h^2)}{f_Z(z)+\mathcal{O}(h^2)}-\boldsymbol{\Sigma}_{jk}(z)+\frac{1}{n}\mathcal{O}\left[\mathbb{E}\Big\{\mathbb{G}_n(w_z)\cdot\mathbb{G}_n(g_{z,jk})\Big\}+\mathbb{E}\left\{\mathbb{G}_n^2(g_{z,jk})\right\}\right]\right|\\
&= \sup_{z\in[0,1]}\max_{j,k\in[d]}\left|\frac{f_Z(z)\boldsymbol{\Sigma}_{jk}(z)+\mathcal{O}(h^2)}{f_Z(z)+\mathcal{O}(h^2)}-\boldsymbol{\Sigma}_{jk}(z)+\mathcal{O}\left(\frac{1}{nh}\right)\right|\\
&= \sup_{z\in[0,1]}\max_{j,k\in[d]}\left|\frac{\mathcal{O}(h^2)\boldsymbol{\Sigma}_{jk}(z)}{f_Z(z)+\mathcal{O}(h^2)}+\mathcal{O}\left(\frac{1}{nh}\right)\right|\\
&\leqslant C\cdot\left(h^2+\frac{1}{nh}\right),
\end{aligned} \tag{S33}$$

where the first inequality follows from (S23) and (S24), the second equality follows from (S25) and (S26), and the last inequality follows from the assumption that $h=o(1)$.

Combining the upper bounds (S32) and (S33), we obtain

$$\sup_{z\in[0,1]} \left\|\widehat{\boldsymbol{\Sigma}}(z) - \boldsymbol{\Sigma}(z)\right\|_{\max} \leqslant C \cdot \left\{h^2 + \sqrt{\frac{\log(d/h)}{nh}}\right\}$$

with probability at least $1 - 3/d$.

## Web Appendix E. Proof of Technical Lemmas in Appendix Web Appendix D

In this section, we provide the proofs of Lemmas S1-S3.

### Web Appendix E.1 *Proof of Lemma S1*

The proof of the lemma uses the following fact

$$(1+x)^{-1} = 1 - x + \mathcal{O}(x^2) \qquad \text{for any } |x| < 1. \tag{S34}$$

From (S12), we have

$$\begin{aligned}
\widehat{\boldsymbol{\Sigma}}_{jk}(z) &= \frac{\mathbb{P}_n(g_{z,jk})}{\mathbb{P}_n(w_z)} \\
&= \frac{\mathbb{P}_n(g_{z,jk}) - \mathbb{E}\left\{\mathbb{P}_n(g_{z,jk})\right\} + \mathbb{E}\left\{\mathbb{P}_n(g_{z,jk})\right\}}{\mathbb{E}\left\{\mathbb{P}_n(w_z)\right\}} \cdot \left[\frac{\mathbb{E}\left\{\mathbb{P}_n(w_z)\right\}}{\mathbb{P}_n(w_z)}\right] \\
&= \frac{n^{-1/2}\cdot \mathbb{G}_n(g_{z,jk}) + \mathbb{E}\left\{\mathbb{P}_n(g_{z,jk})\right\}}{\mathbb{E}\left\{\mathbb{P}_n(w_z)\right\}} \cdot \left[1 + \frac{\mathbb{P}_n(w_z) - \mathbb{E}\left[\mathbb{P}_n(w_z)\right]}{\mathbb{E}\left[\mathbb{P}_n(w_z)\right]}\right]^{-1} \\
&= \frac{n^{-1/2}\cdot \mathbb{G}_n(g_{z,jk}) + \mathbb{E}\left[\mathbb{P}_n(g_{z,jk})\right]}{\mathbb{E}\left\{\mathbb{P}_n(w_z)\right\}} \cdot \left[1 + \frac{\mathbb{G}_n(w_z)}{\sqrt{n}\cdot \mathbb{E}\left\{\mathbb{P}_n(w_z)\right\}}\right]^{-1}.
\end{aligned}$$

Under the conditions (S21) and by applying (S34), we have

$$\begin{aligned}
\widehat{\boldsymbol{\Sigma}}_{jk}(z) &= \frac{n^{-1/2}\cdot \mathbb{G}_n(g_{z,jk}) + \mathbb{E}\left\{\mathbb{P}_n(g_{z,jk})\right\}}{\mathbb{E}\left\{\mathbb{P}_n(w_z)\right\}} \cdot \left(1 - \frac{\mathbb{G}_n(w_z)}{\sqrt{n}\cdot \mathbb{E}\left\{\mathbb{P}_n(w_z)\right\}} + \mathcal{O}\left[\frac{\mathbb{G}_n^2(w_z)}{n\cdot \mathbb{E}^2\left\{\mathbb{P}_n(w_z)\right\}}\right]\right) \\
&= \frac{\mathbb{G}_n(g_{z,jk})}{\sqrt{n}\mathbb{E}\left\{\mathbb{P}_n(w_z)\right\}} + \frac{\mathbb{E}\{\mathbb{P}_n(g_{z,jk})\}}{\mathbb{E}\left\{\mathbb{P}_n(w_z)\right\}} - \frac{\mathbb{G}_n(w_z)\mathbb{E}\{\mathbb{P}_n(g_{z,jk})\}}{\sqrt{n}\mathbb{E}^2\{\mathbb{P}_n(w_z)\}} + \frac{1}{n}\mathcal{O}\left[\left\{\mathbb{G}_n(w_z)\mathbb{G}_n(g_{z,jk})\right\} + \mathbb{G}_n^2(g_{z,jk})\right].
\end{aligned} \tag{S35}$$

Note that $\mathbb{E}\{\mathbb{G}_n(f)\} = 0$ by the definition of $\mathbb{G}_n(f)$ in (S8). Taking expectation on both sides of (S35), we obtain

$$\mathbb{E}\left\{\widehat{\boldsymbol{\Sigma}}_{jk}(z)\right\} = \frac{\mathbb{E}\left\{\mathbb{P}_n(g_{z,jk})\right\}}{\mathbb{E}\left\{\mathbb{P}_n(w_z)\right\}} + \frac{1}{n}\mathcal{O}\left[\mathbb{E}\left\{\mathbb{G}_n(w_z)\cdot \mathbb{G}_n(g_{z,jk})\right\} + \mathbb{E}\left\{\mathbb{G}_n^2(g_{z,jk})\right\}\right],$$

as desired.

Web Appendix E.2 *Proof of Lemma S2*

To prove Lemma S2, we write the expectation as an integral and apply Taylor expansion to the density function and the covariance function. We will show that the higher-order terms of the Taylor expansion can be bounded by $\mathcal{O}(h^2)$. We start by proving (S23).

**Proof of (S23):** Recall from (S9) the definition of $g_{z,jk}(Z_i, X_{ij}, Y_{ik}) = K_h(Z_i - z)X_{ij}Y_{ik}$. Thus, we have

$$\begin{aligned}
\mathbb{E}\{\mathbb{P}_n(g_{z,jk})\} &= \mathbb{E}\left\{\frac{1}{h}K\left(\frac{Z-z}{h}\right)X_jY_k\right\} \\
&= \mathbb{E}\left\{\frac{1}{h}K\left(\frac{Z-z}{h}\right)\mathbb{E}(X_jY_k \mid Z)\right\} \\
&= \mathbb{E}\left\{\frac{1}{h}K\left(\frac{Z-z}{h}\right)\mathbb{E}(S_jS_k \mid Z)\right\} \\
&= \mathbb{E}\left\{\frac{1}{h}K\left(\frac{Z-z}{h}\right)\mathbf{\Sigma}_{jk}(Z)\right\} \\
&= \int \frac{1}{h}K\left(\frac{Z-z}{h}\right)\mathbf{\Sigma}_{jk}(Z)f_Z(Z)dZ \\
&= \int K(u)\mathbf{\Sigma}_{jk}(uh+z)f_Z(uh+z)du,
\end{aligned} \tag{S36}$$

where the third equality hold using the fact that the subject-specific effects are independent between two subjects, and the last equality holds by a change of variable, $u = (Z-z)/h$. Applying Taylor expansions to $\mathbf{\Sigma}_{jk}(uh+z)$ and $f_Z(uh+z)$, we have

$$\mathbf{\Sigma}_{jk}(u+zh) = \mathbf{\Sigma}_{jk}(z) + uh \cdot \dot{\mathbf{\Sigma}}_{jk}(z) + u^2h^2 \cdot \ddot{\mathbf{\Sigma}}_{jk}(z') \tag{S37}$$

and

$$f_Z(u+zh) = f_Z(z) + uh \cdot \dot{f}_Z(z) + u^2h^2 \cdot \ddot{f}_Z(z''), \tag{S38}$$

where $z'$ and $z''$ are between $z$ and $uh+z$. Substituting (S37) and (S38) into the last expression of (S36), we have

$$\int K(u)\left\{\mathbf{\Sigma}_{jk}(z) + uh \cdot \dot{\mathbf{\Sigma}}_{jk}(z) + u^2h^2 \cdot \ddot{\mathbf{\Sigma}}_{jk}(z')\right\} \cdot \left\{f_Z(z) + uh \cdot \dot{f}_Z(z) + u^2h^2 \cdot \ddot{f}_Z(z'')\right\} du. \tag{S39}$$

By (12), we have $\int uK(u)du = 0$ and $\int u^l K(u)du < \infty$ for $l = 1, 2, 3, 4$. By Assumptions 1 and 2, we have

$$
\begin{aligned}
&h^2 \int u^2 K(u) \ddot{\boldsymbol{\Sigma}}_{jk}(z') f_Z(z) du \leqslant h^2 C M_\sigma \bar{f}_Z = \mathcal{O}(h^2), \\
&h^2 \int u^2 K(u) \dot{\boldsymbol{\Sigma}}_{jk}(z) \dot{f}_Z(z) du \leqslant h^2 C M_\sigma \bar{f}_Z = \mathcal{O}(h^2), \\
&h^2 \int u^2 K(u) \boldsymbol{\Sigma}_{jk}(z) \ddot{f}_Z(z'') du \leqslant h^2 C M_\sigma \bar{f}_Z = \mathcal{O}(h^2).
\end{aligned} \tag{S40}
$$

Substituting (S40) into (S39) and bounding the other higher-order terms by $\mathcal{O}(h^2)$, we obtain

$$
\mathbb{E}\{\mathbb{P}_n(g_{z,jk})\} = \boldsymbol{\Sigma}_{jk}(z) f_Z(z) + \mathcal{O}(h^2),
$$

for all $z \in [0,1]$ and $j,k \in [d]$. This implies that

$$
\sup_{z\in[0,1]} \max_{j,k\in[d]} |\mathbb{E}\{\mathbb{P}_n(g_{z,jk})\} - \boldsymbol{\Sigma}_{jk}(z) f_Z(z)| = \mathcal{O}(h^2).
$$

The proof of (S24) follows from the same set of argument.

**Proof of (S25):** Recall from (S9) the definition of $w_z(Z_i) = K_h(Z_i - z)$. Thus, we have

$$
\begin{aligned}
&\frac{1}{n}\mathbb{E}\Big\{\mathbb{G}_n(g_{z,jk}) \cdot \mathbb{G}_n(w_z)\Big\} \\
&= \mathbb{E}\Big\{\mathbb{P}_n(g_{z,jk}) \cdot \mathbb{P}_n(w_z)\Big\} - \mathbb{E}\{\mathbb{P}_n(g_{z,jk})\} \cdot \mathbb{E}\{\mathbb{P}_n(w_z)\} \\
&= \mathbb{E}\left[\left\{\frac{1}{n}\sum_{i\in[n]} K_h(Z_i - z) X_{ij} Y_{ik}\right\} \cdot \left\{\frac{1}{n}\sum_{i\in[n]} K_h(Z_i - z)\right\}\right] - \mathbb{E}\{\mathbb{P}_n(g_{z,jk})\} \cdot \mathbb{E}\{\mathbb{P}_n(w_z)\} \\
&= \frac{1}{n}\mathbb{E}\left\{K_h^2(Z - z) S_j S_k\right\} + \frac{1}{n^2}\mathbb{E}\left\{\sum_{i\in[n]}\sum_{i'\neq i} K_h(Z_i - z) K_h(Z_{i'} - z) X_{ij} Y_{ik}\right\} - \mathbb{E}\{\mathbb{P}_n(g_{z,jk})\} \cdot \mathbb{E}\{\mathbb{P}_n(w_z)\} \\
&= \frac{1}{n}\mathbb{E}\left\{K_h^2(Z - z)\boldsymbol{\Sigma}_{jk}(Z)\right\} + \frac{n-1}{n}\left[\mathbb{E}\left\{K_h(Z - z)\right\} \cdot \mathbb{E}\left\{K_h(Z - z)\boldsymbol{\Sigma}_{jk}(Z)\right\}\right] - \mathbb{E}\{\mathbb{P}_n(g_{z,jk})\}\mathbb{E}\{\mathbb{P}_n(w_z)\} \\
&= \underbrace{\frac{1}{n}\mathbb{E}\left\{K_h^2(Z - z)\boldsymbol{\Sigma}_{jk}(Z)\right\}}_{I_1} - \underbrace{\frac{1}{n}\mathbb{E}\{\mathbb{P}_n(g_{z,jk})\}\mathbb{E}\{\mathbb{P}_n(w_z)\}}_{I_2},
\end{aligned} \tag{S41}
$$

where the second to the last equality follows from the fact that $Z_i$ and $Z_{i'}$ are independent.

We now obtain an upper bound for $I_1$. By (12) and Assumptions 1-2, we have

$$
I_1 = \frac{1}{nh}\int \frac{1}{h} K^2\left(\frac{Z - z}{h}\right) \boldsymbol{\Sigma}_{jk}(Z) f_Z(Z) dZ \leqslant \frac{1}{nh} \cdot M_\sigma \cdot \bar{f}_Z \int \frac{1}{h} K^2\left(\frac{Z - z}{h}\right) dZ = \mathcal{O}\left(\frac{1}{nh}\right), \tag{S42}
$$

where the last equality holds by a change of variable. Moreover, by (S23) and (S24), we have

$$I_2 = \frac{1}{n}\left\{f_Z(z)\mathbf{\Sigma}_{jk}(z) + \mathcal{O}(h^2)\right\} \cdot \{f_Z(z) + \mathcal{O}(h^2)\} = \mathcal{O}\left(\frac{1}{n}\right). \tag{S43}$$

Substituting (S42) and (S43) into (S41), and taking the supreme over $z \in [0,1]$ and $j,k \in [d]$ on both sides of the equation, we obtain

$$\sup_{z\in[0,1]} \max_{j,k\in[d]} \left|\frac{1}{n}\mathbb{E}\Big\{\mathbb{G}_n(g_{z,jk}) \cdot \mathbb{G}_n(w_z)\Big\}\right| = \mathcal{O}\left(\frac{1}{nh}\right) + \mathcal{O}\left(\frac{1}{n}\right) = \mathcal{O}\left(\frac{1}{nh}\right),$$

where the last equality holds by the scaling assumption of $h = o(1)$. The proof of (S26) follows from the same set of argument.

Web Appendix E.3 *Proof of Lemma S3*

The proof of Lemma S3 involves obtaining upper bounds for the supreme of the empirical processes $\mathbb{G}_n(w_z)$ and $\mathbb{G}_n(g_{z,jk})$. To this end, we apply the Talagrand's inequality in Lemma S20. Let $\mathcal{F}$ be a function class. In order to apply Talagrand's inequality, we need to evaluate the quantities $\eta$ and $\tau^2$ such that

$$\sup_{f\in\mathcal{F}} \|f\|_\infty \leqslant \eta \qquad \text{and} \qquad \sup_{f\in\mathcal{F}} \mathrm{Var}(f(X)) \leqslant \tau^2.$$

Talagrand's inequality in Lemma S20 provides an upper bound for the supreme of an empirical process in terms of its expectation. By Lemma S21, the expectation can then be upper bounded as a function of the covering number of the function class $\mathcal{F}$, denoted as $N\{\mathcal{F}, L_2(Q), \epsilon\}$. The following lemmas provide upper bounds for the supreme of the empirical processes $\mathbb{G}_n(w_z)$ and $\mathbb{G}_n(g_{z,jk})$, respectively. The proofs are deferred to Sections Web Appendix E.3.1 and Web Appendix E.3.2, respectively.

Lemma S4: *Assume that $h = o(1)$ and $\log(d/h)/(nh) = o(1)$. Under Assumptions 1-2, for sufficiently large $n$, there exists a universal constant $C > 0$ such that*

$$\sup_{z\in[0,1]} |\mathbb{G}_n(w_z)| \leqslant C \cdot \sqrt{\frac{\log(d/h)}{h}}, \tag{S44}$$

*with probability at least $1 - 1/d$.*

Lemma S5: *Assume that $h = o(1)$ and $\log^2(d/h)/(nh) = o(1)$. Under Assumptions 1-2, for sufficiently large $n$, there exists a universal constant $C > 0$ such that*

$$\sup_{z\in[0,1]} \max_{j,k\in[d]} |\mathbb{G}_n(g_{z,jk})| \leqslant C \cdot \sqrt{\frac{\log(d/h)}{h}}, \tag{S45}$$

*with probability at least $1 - 2/d$.*

Applying Lemmas S4 and S5, we obtain

$$\begin{aligned}\sup_{z\in[0,1]} \max_{j,k\in[d]} \left|\mathbb{G}_n(w_z) \vee \mathbb{G}_n(g_{z,jk})\right| &\leqslant \sup_{z\in[0,1]} |\mathbb{G}_n(w_z)| + \sup_{z\in[0,1]} \max_{j,k\in[d]} |\mathbb{G}_n(g_{z,jk})| \\ &\leqslant C \cdot \sqrt{\frac{\log(d/h)}{h}},\end{aligned}$$

with probability at least $1 - 3/d$, as desired.

Web Appendix E.3.1 *Proof of Lemma S4* . The proof of Lemma S4 uses the set of arguments as detailed in the beginning of Web Appendix E.3. Recall from (S9) and (S11) the definition of $w_z(Z_i) = K_h(Z_i - z)$ and $k_z(Z_i) = w_z(Z_i) - \mathbb{E}\{w_z(Z)\}$, respectively. We consider the class of function

$$\mathcal{K} = \{k_z \mid z \in [0, 1]\}. \tag{S46}$$

First, note that

$$\begin{aligned}\sup_{z\in[0,1]} \|k_z\|_\infty &= \sup_{z\in[0,1]} \|w_z(Z_i) - \mathbb{E}\{w_z(Z)\}\|_\infty \\ &\leqslant \frac{1}{h}\|K\|_\infty + \bar{f}_Z + \mathcal{O}(h^2) \\ &\leqslant \frac{2}{h}\|K\|_\infty,\end{aligned} \tag{S47}$$

where the first inequality holds by (12) and Lemma S2, and the last inequality holds by the scaling assumption $h = o(1)$ for sufficiently large $n$.

Next, we obtain an upper bound for the variance of $k_z(Z_i)$. Note that

$$\begin{aligned}\sup_{z\in[0,1]} \mathrm{Var}\{k_z(Z)\} &= \sup_{z\in[0,1]} \mathbb{E}\left([w_z(Z) - \mathbb{E}\{w_z(Z)\}]^2\right) \\ &\leqslant \underbrace{\sup_{z\in[0,1]} 2\mathbb{E}\{w_z^2(Z)\}}_{I_1} + \underbrace{\sup_{z\in[0,1]} 2\mathbb{E}^2\{w_z(Z)\}}_{I_2},\end{aligned}$$

where we apply the inequality $(x-y)^2 \leqslant 2x^2 + 2y^2$ for two scalars $x, y$. By Lemma S2, we have $I_2 \leqslant 2\{\bar{f}_Z + \mathcal{O}(h^2)\}^2$. Also, by a change of variable and second-order Taylor expansion on the marginal density $f_Z(\cdot)$, we have

$$
\begin{aligned}
I_1 &= 2 \sup_{z\in[0,1]} \int \frac{1}{h^2} K^2\left(\frac{Z-z}{h}\right) f_Z(Z) dZ \\
&= 2 \sup_{z\in[0,1]} \frac{1}{h} \int K^2(u) f_Z(uh+z) du \\
&= 2 \sup_{z\in[0,1]} \frac{1}{h} \int K^2(u) \left\{ f_Z(z) + uh\dot{f}_Z(z) + u^2h^2\ddot{f}_Z(z') \right\} du \quad \text{for } z' \in (z, u+zh) \\
&\leqslant \frac{2}{h} \bar{f}_Z \|K\|_2^2 + \mathcal{O}(1) + \mathcal{O}(h).
\end{aligned} \tag{S48}
$$

Thus, for sufficiently large $n$ and the assumption that $h = o(1)$, we have

$$
\sup_{z\in[0,1]} \operatorname{Var}\{k_z(Z)\} \leqslant \frac{3}{h} \cdot \bar{f}_Z \cdot \|K\|_2^2. \tag{S49}
$$

By Lemma S16, the covering number for the function class $\mathcal{K}$ satisfies

$$
\sup_{Q} N\{\mathcal{K}, L_2(Q), \epsilon\} \leqslant \left( \frac{4 \cdot \|K\|_{\mathrm{TV}} \cdot C_K^{4/5} \cdot \bar{f}_Z^{1/5}}{h\epsilon} \right)^5. \tag{S50}
$$

We are now ready to obtain an upper bound for the supreme of the empirical process, $\sup_{z\in[0,1]} |\mathbb{G}_n(w_z)|$. By Lemma S21 with $A = 2 \cdot \|K\|_{\mathrm{TV}} \cdot C_K^{4/5} \cdot \bar{f}_Z^{1/5} / \|K\|_\infty$, $\|F\|_{L_2(\mathbb{P}_n)} = 2 \cdot \|K\|_\infty / h$, $V = 5$, $\sigma_P^2 = 3 \cdot \bar{f}_Z \cdot \|K\|_2^2/h$, for sufficiently large $n$, we obtain

$$
\begin{aligned}
\mathbb{E}\left\{ \sup_{z\in[0,1]} \frac{1}{\sqrt{n}} \cdot |\mathbb{G}_n(w_z)| \right\} &= \mathbb{E}\left( \sup_{z\in[0,1]} \frac{1}{n} \left| \sum_{i\in[n]} [w_z(Z_i) - \mathbb{E}\{w_z(Z)\}] \right| \right) \\
&\leqslant C \cdot \left\{ \sqrt{\frac{\log(1/h)}{nh}} + \frac{\log(1/h)}{n} \right\} \\
&\leqslant C \cdot \sqrt{\frac{\log(1/h)}{nh}},
\end{aligned} \tag{S51}
$$

where $C > 0$ is some sufficiently large constant. By Lemma S20 with $\tau^2 = 3\bar{f}_Z \cdot \|K\|_2^2/h$, $\eta = 2 \cdot \|K\|_\infty/h$, $\mathbb{E}[Y] \leqslant C \cdot \sqrt{\log(1/h)/(nh)}$, and picking $t = \sqrt{\log(d)/n}$, for sufficiently large $n$, we have

$$\begin{aligned}\sup_{z\in[0,1]} \frac{1}{\sqrt{n}} \cdot |\mathbb{G}_n(w_z)| &= \sup_{z\in[0,1]} \frac{1}{n} \left| \sum_{i\in[n]} (w_z(Z_i) - \mathbb{E}\{w_z(Z)\} \right| \\ &\leqslant C \cdot \left( \sqrt{\frac{\log(1/h)}{nh}} + \sqrt{\frac{\log(d)}{nh}} \cdot \sqrt{1 + \sqrt{\frac{\log(1/h)}{nh}} + \frac{\log(d)}{nh}} \right) \\ &\leqslant C \cdot \sqrt{\frac{\log(d/h)}{nh}},\end{aligned}$$

with probability $1-1/d$, where the last expression holds by the assumption that $\log(d/h)/(nh) = o(1)$ and $h = o(1)$. Multiplying both sides of the above equation by $\sqrt{n}$ completes the proof of Lemma S4.

Web Appendix E.3.2 *Proof of Lemma S5* . The proof of Lemma S5 uses the set of arguments as detailed in the beginning of Web Appendix E.3. For convenience, we prove Lemma S5 by conditioning on the event

$$\mathcal{A} = \left\{ \max_{i\in[n]} \max_{j\in[d]} \max(|X_{ij}|, |Y_{ij}|) \leqslant M_X \cdot \sqrt{\log d} \right\}. \quad \text{(S52)}$$

Since $X_{ij}$ and $Y_{ij}$ conditioned on $Z$ are Gaussian random variables, the event $\mathcal{A}$ occurs with probability at least $1 - 1/d$ for sufficiently large constant $M_X > 0$.

Recall from (S9) and (S10) the definition of $g_{z,jk}(Z_i, X_{ij}, Y_{ik}) = K_h(Z_i - z)X_{ij}Y_{ik}$ and $q_{z,jk}(Z_i, X_{ij}, Y_{ik}) = g_{z,jk}(Z_i, X_{ij}, Y_{ik}) - \mathbb{E}\{g_{z,jk}(Z, X_j, Y_k)\}$, respectively. We consider the function class

$$\mathcal{Q} = \{q_{z,jk} \mid z \in [0,1], j, k \in [d]\}. \quad \text{(S53)}$$

We first obtain an upper bound for the function class

$$
\begin{aligned}
\sup_{z\in[0,1]}\max_{j,k\in[d]}\|q_{z,jk}\|_\infty &= \sup_{z\in[0,1]}\max_{j,k\in[d]}\|g_{z,jk}(Z_i,X_{ij},Y_{ik})-\mathbb{E}\{g_{z,jk}(Z,X_j,Y_k)\}\|_\infty \\
&\leqslant \sup_{z\in[0,1]}\max_{j,k\in[d]}\|g_{z,jk}(Z_i,X_{ij},Y_{ik})\|_\infty + \sup_{z\in[0,1]}\max_{j,k\in[d]}\|\mathbb{E}\{g_{z,jk}(Z,X_j,Y_k)\}\|_\infty \\
&\leqslant \sup_{z\in[0,1]}\max_{j,k\in[d]}\|K_h(Z_i-z)X_{ij}Y_{ik}\|_\infty + \bar{f}_Z\cdot M_\sigma + \mathcal{O}(h^2) \\
&\leqslant \frac{1}{h}\cdot M_X^2\cdot\|K\|_\infty\cdot\log d + \bar{f}_Z\cdot M_\sigma + \mathcal{O}(h^2) \\
&\leqslant \frac{2}{h}\cdot M_X^2\cdot\|K\|_\infty\cdot\log d,
\end{aligned}
\tag{S54}
$$

where the second inequality holds by Assumptions 1-2 and Lemma S2, the third inequality holds by (12) and by conditioning on the event $\mathcal{A}$, and the last inequality holds by the scaling assumption $h=o(1)$ for sufficiently large $n$.

Next, we obtain an upper bound for the variance of $q_{z,jk}(Z_i,X_{ij},Y_{ik})$. Note that

$$
\begin{aligned}
\sup_{z\in[0,1]}\max_{j,k\in[d]}\mathrm{Var}\{q_{z,jk}(Z,X_j,Y_k)\} &= \sup_{z\in[0,1]}\max_{j,k\in[d]}\mathbb{E}\left[(g_{z,jk}(Z,X_j,Y_k)-\mathbb{E}\{g_{z,jk}(Z,X_j,Y_k)\})^2\right] \\
&\leqslant \underbrace{\sup_{z\in[0,1]}\max_{j,k\in[d]}2\mathbb{E}\left\{g_{z,jk}^2(Z,X_j,Y_k)\right\}}_{I_1} + \underbrace{\sup_{z\in[0,1]}\max_{j,k\in[d]}2\mathbb{E}^2\{g_{z,jk}(Z,X_j,Y_k)\}}_{I_2},
\end{aligned}
$$

where we apply the inequality $(x-y)^2\leqslant 2x^2+2y^2$ for two scalars $x,y$. By Lemma S2, we have $I_2\leqslant 2\left\{\bar{f}_Z\cdot M_\sigma+\mathcal{O}(h^2)\right\}^2$. Also, by a change of variable and second-order Taylor expansion on the marginal density $f_Z(\cdot)$ as in (S48), we have

$$
\begin{aligned}
I_1 &= 2\sup_{z\in[0,1]}\max_{j,k\in[d]}\mathbb{E}\left\{K_h^2(Z-z)\cdot\mathbb{E}\left(X_j^2Y_k^2\mid Z\right)\right\} \\
&\leqslant 2\kappa\sup_{z\in[0,1]}\mathbb{E}\left\{K_h^2(Z-z)\right\} \\
&\leqslant \frac{2\kappa}{h}\cdot\bar{f}_Z\cdot\|K\|_2^2+\mathcal{O}(1)+\mathcal{O}(h),
\end{aligned}
$$

where the first inequality follows from the fact that $|\mathbb{E}(X_j^2Y_k^2\mid Z)|\leqslant\kappa$ for some $\kappa<\infty$ since these are Gaussian random variables, and the second inequality follows from (S48). Thus, for sufficiently large $n$ and the assumption that $h = o(1)$, we have

$$\sup_{z\in[0,1]} \max_{j,k\in[d]} \mathrm{Var}\{q_{z,j,k}(Z, X_j, Y_k)\} \leqslant \frac{3\kappa}{h} \cdot \bar{f}_Z \cdot \|K\|_2^2. \tag{S55}$$

By Lemma S17, the covering number for the function class $\mathcal{Q}$ satisfies

$$\sup_Q N\{\mathcal{Q}, L_2(Q), \epsilon\} \leqslant \left( \frac{4\|K\|_{\mathrm{TV}} \cdot C_K^{4/5} \cdot \bar{f}_Z^{1/5} \cdot M_\sigma^{1/5} \cdot d^{1/10} \cdot M_X^{2/5} \cdot \log^{2/5} d}{h\epsilon} \right)^5. \tag{S56}$$

We now obtain an upper bound for the supreme of the empirical process, $\sup_{z\in[0,1]} \max_{j,k\in[d]} |\mathbb{G}_n(g_{z,jk})|$. By Lemma S21 with $A = 2 \cdot \|K\|_{\mathrm{TV}} \cdot C_K^{4/5} \cdot \bar{f}_Z^{1/5} \cdot M_\sigma^{1/5} \cdot d^{1/10} / \|K\|_\infty$, $\|F\|_{L_2(\mathbb{P}_n)} = 2 \cdot \|K\|_\infty \cdot M_X^2 \cdot \log d/h$, $V = 5$, $\sigma_P^2 = (3\kappa/h) \cdot \bar{f}_Z \cdot \|K\|_2^2$, for sufficiently large $n$, we obtain

$$\begin{aligned}
\mathbb{E}\left\{ \sup_{z\in[0,1]} \max_{j,k\in[d]} \frac{1}{\sqrt{n}} \cdot |\mathbb{G}_n(g_{z,jk})| \right\} &= \mathbb{E}\left( \sup_{z\in[0,1]} \max_{j,k\in[d]} \frac{1}{n} \cdot \left| \sum_{i\in[n]} [g_{z,jk}(Z_i, X_{ij}, Y_{ik}) - \mathbb{E}\{g_{z,jk}(Z, X_j, Y_k)\}] \right| \right) \\
&\leqslant C \cdot \left\{ \sqrt{\frac{\log(d/h)}{nh}} + \frac{\log(d/h)}{n} \right\} \\
&\leqslant C \cdot \sqrt{\frac{\log(d/h)}{nh}},
\end{aligned} \tag{S57}$$

where the last inequality holds by the assumption $\log(d/h)/nh = o(1)$. By Lemma S20 with $\tau^2 = 3 \cdot \kappa \cdot \bar{f}_Z \cdot \|K\|_2^2/h$, $\eta = 2 \cdot \|K\|_\infty \cdot M_X^2 \cdot \log d/h$, $\mathbb{E}[Y] \leqslant C \cdot \sqrt{\log(d/h)/(nh)}$, and picking $t = \sqrt{\log d/n}$, for sufficiently large $n$, we have

$$\begin{aligned}
\sup_{z\in[0,1]} \max_{j,k\in[d]} \frac{1}{\sqrt{n}} \cdot |\mathbb{G}_n(g_{z,jk})| &= \sup_{z\in[0,1]} \max_{j,k\in[d]} \frac{1}{n} \cdot \left| \sum_{i\in[n]} [g_{z,jk}(Z_i, X_{ij}, Y_{ik}) - \mathbb{E}\{g_{z,jk}(Z, X_j, Y_k)\}] \right| \\
&\leqslant C \cdot \left\{ \sqrt{\frac{\log(d/h)}{nh}} + \sqrt{\frac{\log d}{nh}} \cdot \sqrt{1 + \log d \cdot \sqrt{\frac{\log(d/h)}{nh}}} + \frac{\log^2 d}{nh} \right\} \\
&\leqslant C \cdot \sqrt{\frac{\log(d/h)}{nh}},
\end{aligned}$$

with probability at least $1 - 2/d$. The second inequality holds by the assumption that $\log^2(d/h)/(nh) = o(1)$. Multiplying both sides of the equation by $\sqrt{n}$, we completed the proof of Lemma S5.

**Web Appendix F. Proof of Theorem 2**

In this section, we provide the proof of Theorem 2. To prove Theorem 2, we use a similar set of arguments in the series of work on Gaussian multiplier bootstrap of the supreme of empirical process (see, for instance, Chernozhukov et al., 2013, 2014a,b). Recall from (9) and (10) that

$$T_E = \sup_{z\in[0,1]} \max_{(j,k)\in E(z)} \sqrt{nh} \cdot \left| \widehat{\mathbf{\Theta}}^{\text{de}}_{jk}(z) - \mathbf{\Theta}_{jk}(z) \right| \cdot \mathbb{P}_n(w_z) \tag{S58}$$

and

$$T_E^B = \sup_{z\in[0,1]} \max_{(j,k)\in E(z)} \sqrt{nh} \cdot \left| \frac{\sum_{i\in[n]} \left\{ \widehat{\mathbf{\Theta}}_j(z) \right\}^T K_h(Z_i - z) \left\{ \boldsymbol{X}_i \boldsymbol{Y}_i^T \widehat{\mathbf{\Theta}}_k(z) - \mathbf{e}_k \right\} \xi_i / n}{\left\{ \widehat{\mathbf{\Theta}}_j(z) \right\}^T \widehat{\mathbf{\Sigma}}_j(z)} \right|, \tag{S59}$$

respectively, where $\xi_i \sim N(0,1)$. Note that for notational convenience, we drop the subscript $E$ from $T_E$ and $T_E^B$ throughout the proof.

We aim to show that $T^B$ is a good approximation of $T$. However, $T$ and $T^B$ are not exact averages. To apply the results in Chernozhukov et al., 2014a, we define four intermediate processes:

$$T_0 = \sup_{z\in[0,1]} \max_{(j,k)\in E(z)} \sqrt{nh} \cdot \left| \sum_{i\in[n]} \left\{ \mathbf{\Theta}_j(z) \right\}^T K_h(Z_i - z) \left\{ \boldsymbol{X}_i \boldsymbol{Y}_i^T - \mathbf{\Sigma}(z) \right\} \mathbf{\Theta}_k(z)/n \right|; \tag{S60}$$

$$\begin{aligned} T_{00} = \sup_{z\in[0,1]} \max_{(j,k)\in E(z)} \sqrt{nh} \cdot \Bigg| & \sum_{i\in[n]} \left\{ \mathbf{\Theta}_j(z) \right\}^T K_h(Z_i - z) \left\{ \boldsymbol{X}_i \boldsymbol{Y}_i^T - \mathbf{\Sigma}(z) \right\} \mathbf{\Theta}_k(z)/n \\ & - \left\{ \mathbf{\Theta}_j(z) \right\}^T \left( \Big[ \mathbb{E}\{K_h(Z - z)\boldsymbol{X}\boldsymbol{Y}^T\} - \mathbb{E}\{K_h(Z - z)\}\mathbf{\Sigma}(z) \Big] \right) \mathbf{\Theta}_k(z)/n \Bigg|; \end{aligned} \tag{S61}$$

$$T_0^B = \sup_{z\in[0,1]} \max_{(j,k)\in E(z)} \sqrt{nh} \cdot \left| \sum_{i\in[n]} \Big[ \left\{ \mathbf{\Theta}_j(z) \right\}^T K_h(Z_i - z) \left\{ \boldsymbol{X}_i \boldsymbol{Y}_i^T - \mathbf{\Sigma}(z) \right\} \mathbf{\Theta}_k(z) \Big] \xi_i / n \right|, \tag{S62}$$

$$T_{00}^B = \sup_{z\in[0,1]} \max_{(j,k)\in E(z)} \sqrt{nh} \cdot \Bigg| \sum_{i\in[n]} \Big\{ \{\boldsymbol{\Theta}_j(z)\}^T K_h(Z_i - z) \left\{\boldsymbol{X}_i \boldsymbol{Y}_i^T - \boldsymbol{\Sigma}(z)\right\} \boldsymbol{\Theta}_k(z)/n$$
$$- \{\boldsymbol{\Theta}_j(z)\}^T \Big(\Big[\mathbb{E}\{K_h(Z-z)\boldsymbol{X}\boldsymbol{Y}^T\} - \mathbb{E}\{K_h(Z-z)\}\boldsymbol{\Sigma}(z)\Big]\Big)\boldsymbol{\Theta}_k(z)\Big\} \cdot \xi_i/n\Bigg|; \tag{S63}$$

where $\xi_i \overset{\text{i.i.d.}}{\sim} N(0,1)$.

To prove Theorem 2, we show that $T_{00}$ is a good approximation of $T$ and that $T_{00}^B$ is a good approximation of $T^B$. We then show that there exists a Gaussian process $W$ such that both $T_{00}^B$ and $T_{00}$ can be accurately approximated by $W$. This is done by applications of Theorems A.1 and A.2 in Chernozhukov et al. (2014a). The following summarizes the chain of empirical and Gaussian processes that we are going to study

$$T \longleftrightarrow T_0 \longleftrightarrow T_{00} \longleftrightarrow W \longleftrightarrow T_{00}^B \longleftrightarrow T_0^B \longleftrightarrow T^B.$$

The following lemma provides an approximation error between the statistic $T$ and the intermediate empirical process $T_{00}$.

Lemma S6: *Assume that $h^2 + \sqrt{\log(d/h)/nh} = o(1)$. Under Assumptions 1-2, for sufficiently large $n$, there exists a universal constant $C > 0$ such that*

$$|T - T_{00}| \leqslant C \cdot \left\{\sqrt{nh^5} + s \cdot \sqrt{nh^9} + \frac{s \cdot \log(d/h)}{\sqrt{nh}} + \cdot s \cdot h^2 \cdot \sqrt{\log(d/h)}\right\},$$

*with probability at least $1 - 1/d$.*

*Proof.* The proof is deferred to Web Appendix F.2.

We now apply Theorems A.1 and A.2 in Chernozhukov et al. (2014a) to show that there exists a Gaussian process $W$ such that the quantities $|T_{00} - W|$ and $|T_{00}^B - W|$ can be controlled, respectively. The results are stated in the following lemmas.

Lemma S7: *Assume that $\log^6 s \cdot \log^4(d/h)/(nh) = o(1)$. Under Assumptions 1-2, for* *sufficiently large $n$, there exists universal constants $C, C' > 0$ such that*

$$P\left[|T_{00} - W| \geqslant C \cdot \left\{\frac{\log^6(s) \cdot \log^4(d/h)}{nh}\right\}^{1/8}\right] \leqslant C' \cdot \left\{\frac{\log^6(s) \cdot \log^4(d/h)}{nh}\right\}^{1/8}.$$

*Proof.* The proof is deferred to Web Appendix F.3.

Lemma S8: *Assume that $\log^4(s) \cdot \log^3(d/h)/(nh) = o(1)$. Under Assumptions 1-2, for sufficiently large $n$, there exists universal constants $C, C'' > 0$ such that*

$$P\left[|T_{00}^B - W| > C \cdot \left\{\frac{\log^4(s) \cdot \log^3(d/h)}{nh}\right\}^{1/8} \;\Big|\; \{(Z_i, \boldsymbol{X}_i, \boldsymbol{Y}_i)\}_{i \in [n]}\right] \leqslant C'' \cdot \left\{\frac{\log^4(s) \cdot \log^3(d/h)}{nh}\right\}^{1/8},$$

*with probability at least $1 - 3/n$.*

*Proof.* The proof is deferred to Web Appendix F.4.

Finally, the following lemma provides an upper bound on the difference between $T^B$ and $T_{00}^B$, conditioned on the data $\{(Z_i, \boldsymbol{X}_i, \boldsymbol{Y}_i)\}_{i \in [n]}$.

Lemma S9: *Assume that $s \cdot \sqrt{h^3 \log^3(d/h)} + s \cdot \sqrt{\log^4(d/h)/nh^2} + \sqrt{h^5 \log n} = o(1)$. Under Assumptions 1-2, for sufficiently large $n$, there exists universal constants $C, C'' > 0$ such that, with probability at least $1 - 1/d$,*

$$P\left[|T^B - T_{00}^B| > C \cdot \sqrt{h^3 \log^3(d/h)} + s \cdot \sqrt{\frac{\log^4(d/h)}{nh^2}} + \sqrt{h^5 \log n} \;\Big|\; \{(Z_i, \boldsymbol{X}_i, \boldsymbol{Y}_i)\}_{i \in [n]}\right] \leqslant 2/d + 1/n.$$

*Proof.* The proof is deferred to Web Appendix F.5.

With Lemmas S6-S9, we are now ready to prove Theorem 2.

### Web Appendix F.1 *Proof of Theorem 2*

Recall that for notational convenience, we drop the subscript $E$ from $T_E$ and $T_E^B$ throughout the proof. In this section, we show that $T$ can be well-approximated by the $(1-\alpha)$-conditional quantile of $T^B$, i.e., $P\{T \geqslant c(1-\alpha)\} \leqslant \alpha$. For notational convenience, we let $r = r_1 + r_2 +$

$r_3 + r_4$, where

$$r_1 = \sqrt{nh^5} + s \cdot \sqrt{nh^9} + \frac{s \cdot \log(d/h)}{\sqrt{nh}} + \cdot s \cdot h^2 \cdot \sqrt{\log(d/h)}$$

$$r_2 = \left\{ \frac{\log^6 s \cdot \log^4(d/h)}{nh} \right\}^{1/8}$$

$$r_3 = \left\{ \frac{\log^4 s \cdot \log^3(d/h)}{nh} \right\}^{1/8}$$

$$r_4 = \sqrt{h^3 \log^3(d/h)} + s \cdot \sqrt{\frac{\log^4(d/h)}{nh^2}} + \sqrt{h^5 \log n}.$$

These are the scaling that appears in Lemmas S6-S9. By Lemmas S6 and S7, it can be shown that

$$P(|T - W| \geqslant 2r_2) \leqslant P(|T - T_{00}| + |T_{00} - W| \geqslant 2r_2) \leqslant 2r_2, \tag{S64}$$

since $r_2 \geqslant r_1$ and $r_2 \geqslant 1/d$. With some abuse of notation, throughout the proof, we write $P_\xi(T^B \geqslant t)$ to indicate $P[T^B \geqslant t \mid \{(Z_i, \boldsymbol{X}_i, \boldsymbol{Y}_i)\}_{i\in[n]}]$. By Lemmas S8 and S9, we have

$$P_\xi(|T^B - W| \geqslant 2r_2) \leqslant P_\xi(|T^B - T^B_{00}| + |T^B_{00} - W| \geqslant 2r_2) \leqslant 2r_2, \tag{S65}$$

since $r_2 \geqslant r_3$ and $r_2 \geqslant 2/d + 1/n$. Define the event

$$\mathcal{E} = \left(P[|T^B_{00} - W| > r_2 \mid \{(Z_i, \boldsymbol{X}_i, \boldsymbol{Y}_i)\}_{i\in[n]} \leqslant r_2]\right),$$

and note that $P(\mathcal{E}) \geqslant 1 - 2/d - 4/n$ by Lemmas S8 and S9. Throughout the proof, we condition on the event $\mathcal{E}$.

By the triangle inequality, we obtain

$$\begin{aligned} P\{T \leqslant c(1-\alpha)\} &\geqslant 1 - P\{T - W + W + 2r_2 \geqslant c(1-\alpha) + 2r_2\} \\ &\geqslant 1 - P(|T - W| \geqslant 2r_2) - P\{W \geqslant c(1-\alpha) - 2r_2\} \\ &\geqslant P\{|W| \leqslant c(1-\alpha) - 2r_2\} - 2r_2, \end{aligned} \tag{S66}$$

where the last inequality follows from (S64). By a similar argument and by (S65), we have

$$\begin{aligned} P\{W \leqslant c(1-\alpha) - 2r_2\} &\geqslant P_\xi\{T^B \leqslant c(1-\alpha) - 4r_2\} - 2r_2 \\ &\geqslant P_\xi\{T^B \leqslant c(1-\alpha)\} - 2r_2 - P_\xi\{|T^B - c(1-\alpha)| \leqslant r_2\}, \end{aligned} \tag{S67}$$

where the last inequality follows from the fact that $P(X \leqslant t - \epsilon) - P(X \leqslant t) \geqslant -P(|X - t| \leqslant$

$\epsilon$) for any $\epsilon > 0$. Thus, combining (S66) and (S67), we obtain

$$P\{T \leqslant c(1-\alpha)\} \geqslant 1 - \alpha - 4r_2 - P_\xi\{|T^B - c(1-\alpha)| \leqslant r_2\}. \tag{S68}$$

It remains to show that the quantity $P_\xi\{|T^B - c(1-\alpha)| \leqslant r_2\}$ converges to zero as we increase $n$.

By the definition of $T_{00}$ and from (S15), we have

$$T_{00} = \sup_{z\in[0,1]} \max_{j,k\in[d]} \frac{1}{\sqrt{n}} \left| \sum_{i\in[n]} J_{z,jk}(Z_i, \boldsymbol{X}_i, \boldsymbol{Y}_i) \right| \quad \text{and} \quad T_{00}^B = \sup_{z\in[0,1]} \max_{j,k\in[d]} \frac{1}{\sqrt{n}} \left| \sum_{i\in[n]} J_{z,jk}(Z_i, \boldsymbol{X}_i, \boldsymbol{Y}_i)\xi_i \right|.$$

Let $\widehat{\sigma}^2_{z,jk} = \sum_{i=1}^n J^2_{z,jk}(Z_i, \boldsymbol{X}_i, \boldsymbol{Y}_i)/n$ be the conditional variance, and let $\underline{\sigma} = \inf_{z,jk} \widehat{\sigma}_{z,jk}$ and $\bar{\sigma} = \sup_{z,jk} \widehat{\sigma}_{z,jk}$. By Lemma A.1 of Chernozhukov et al. (2014b) and Theorem 3 of Chernozhukov et al. (2013), we obtain

$$\begin{aligned} &P_\xi\{|T^B - c(1-\alpha)| \leqslant r_2\} \\ &\leqslant C \cdot \bar{\sigma}/\underline{\sigma} \cdot r_2 \cdot \{\mathbb{E}[T^B \mid \{(Z_i, \boldsymbol{X}_i, \boldsymbol{Y}_i)\}_{i\in[n]}] + \sqrt{1 \vee \log(\underline{\sigma}/r_2)}\} \\ &\leqslant C \cdot \bar{\sigma}/\underline{\sigma} \cdot r_2 \cdot \{\mathbb{E}[T^B_{00} \mid \{(Z_i, \boldsymbol{X}_i, \boldsymbol{Y}_i)\}_{i\in[n]}] + \mathbb{E}[|T^B - T^B_{00}| \mid \{(Z_i, \boldsymbol{X}_i, \boldsymbol{Y}_i)\}_{i\in[n]}] + \sqrt{1 \vee \log(\underline{\sigma}/r_2)}\}. \end{aligned} \tag{S69}$$

We first calculate the quantity $\bar{\sigma}$. By (S90), we have

$$\sup_{z\in[0,1]} \max_{j,k\in[d]} \|J^2_{z,jk}(Z_i, \boldsymbol{X}_i, \boldsymbol{Y}_i)\|_\infty \leqslant C \cdot \log^2 s/h. \tag{S70}$$

Moreover, by (S90), we have

$$\sup_{z\in[0,1]} \max_{j,k\in[d]} \mathbb{E}[J^4_{z,jk}(Z_i, \boldsymbol{X}_i, \boldsymbol{Y}_i)] \leqslant C \cdot \log^4 s/h^2. \tag{S71}$$

Define the function class $\mathcal{J}' = \{J^2_{z,jk}(\cdot) \mid z \in [0,1], j,k \in [d]\}$. By Lemmas S15, S18 and S19, we have

$$\sup_Q N\{\mathcal{J}', L_2(Q), \epsilon\} \leqslant C \cdot d^2 \cdot \left( \frac{d^{17/24} \cdot \log^{3/4} d}{h^{11/12} \cdot \epsilon} \right)^{24}. \tag{S72}$$

Thus, applying Lemma S21 with $\sigma_P^2 = C \cdot \log^4 s/h^2$ and $\|F\|_{L_2(\mathbb{P}_n)} \leqslant C \cdot d^2 \cdot (d^{17/24} \cdot \log^{3/4} d/h^{11/12})^{24}$, we have

$$\mathbb{E}\left[ \sup_{z\in[0,1]} \max_{j,k\in[d]} \frac{1}{n} \left| \sum_{i\in[n]} J^2_{z,jk}(Z_i, \boldsymbol{X}_i, \boldsymbol{Y}_i) - \mathbb{E}\{J^2_{z,jk}(Z, \boldsymbol{X}, \boldsymbol{Y})\} \right| \right] \leqslant C \cdot \sqrt{\frac{\log^5(d/h)}{nh^2}}.$$

By an application of the Markov's inequality, we obtain

$$P\left(\sup_{z\in[0,1]}\max_{j,k\in[d]}\left[\frac{1}{n}\sum_{i\in[n]}J^2_{z,jk}(Z_i,\boldsymbol{X}_i,\boldsymbol{Y}_i)-\mathbb{E}\{J^2_{z,jk}(Z_i,\boldsymbol{X}_i,\boldsymbol{Y}_i)\}\right]\geqslant C\cdot\left\{\frac{\log^5(d/h)}{nh^2}\right\}^{1/4}\right)\leqslant C\cdot\left\{\frac{\log^5(d/h)}{nh^2}\right\}^{1/4}. \tag{S73}$$

Thus, we have with probability at least $1-C\cdot\left\{\log^5(d/h)/(nh^2)\right\}^{1/4}$,

$$\bar{\sigma}^2=\sup_{z\in[0,1]}\max_{j,k\in[d]}\frac{1}{n}\sum_{i\in[n]}J^2_{z,jk}(Z_i,\boldsymbol{X}_i,\boldsymbol{Y}_i)\leqslant\sup_{z\in[0,1]}\max_{j,k\in[d]}\mathbb{E}\{J^2_{z,jk}(Z_i,\boldsymbol{X}_i,\boldsymbol{Y}_i)\}+C\cdot\left\{\frac{\log^5(d/h)}{nh^2}\right\}^{1/4}\leqslant C\cdot\log^2 s, \tag{S74}$$

where the last inequality follows from (S95) for sufficiently large $n$. By Lemma S10, we have $\inf_{z,j,k}\mathbb{E}\{J^2_{z,jk}(Z,\boldsymbol{X},\boldsymbol{Y})\}\geqslant c>0$. Therefore, we have

$$\underline{\sigma}^2=\inf_{z,j,k}\frac{1}{n}\sum_{i=1}^{n}J^2_{z,jk}(Z_i,\boldsymbol{X}_i,\boldsymbol{Y}_i)\geqslant c-\sup_{z,j,k}\frac{1}{n}\sum_{i=1}^{n}[J^2_{z,jk}(Z_i,\boldsymbol{X}_i,\boldsymbol{Y}_i)-\mathbb{E}\{J^2_{z|(j,k)}(Z,\boldsymbol{X},\boldsymbol{Y})\}]\geqslant c/2>0,$$

with probability at least $1-C\cdot\left\{\log^5(d/h)/(nh^2)\right\}^{1/4}$.

Next, we calculate the quantity $\mathbb{E}[T^B_{00}\mid\{(Z_i,\boldsymbol{X}_i,\boldsymbol{Y}_i)\}_{i\in[n]}]$. By Dudley's inequality (see, e.g., Corollary 2.2.8 in Van Der Vaart and Wellner, 1996) and (S96), we obtain

$$\mathbb{E}[T^B_{00}\mid\{(Z_i,\boldsymbol{X}_i,\boldsymbol{Y}_i)\}_{i\in[n]}]\leqslant C\cdot\log s\cdot\sqrt{\log(d/h)}. \tag{S75}$$

Moreover, by Lemma S9, we have

$$\mathbb{E}[|T^B-T^B_{00}|\mid\{(Z_i,\boldsymbol{X}_i,\boldsymbol{Y}_i)\}_{i\in[n]}]\leqslant C\cdot\sqrt{h^3\log^3(d/h)}+s\cdot\sqrt{\frac{\log^4(d/h)}{nh^2}}+\sqrt{h^5\log n}\leqslant r_2, \tag{S76}$$

with probability at least $1-2/d-1/n$. Substituting (S74), (S75), and (S76) into (S69), we obtain

$$P_\xi\{|T^B-c(1-\alpha)|\leqslant r_2\}\leqslant C\cdot\left\{\frac{\log^{22}s\cdot\log^8(d/h)}{nh}\right\}^{1/8}. \tag{S77}$$

Thus, substituting (S77) into (S68), we have

$$P\{T\leqslant c(1-\alpha)\}\geqslant 1-\alpha-4r_2-\frac{\log^{22}s\cdot\log^8(d/h)}{nh}.$$

By the scaling assumptions, $r_2=o(1)$ and $\log^{22}s\cdot\log^8(d/h)/(nh)=o(1)$. Thus, this implies that

$$\lim_{n\to\infty}P\{T\leqslant c(1-\alpha)\}\geqslant 1-\alpha,$$

which implies that

$$\lim_{n\to\infty} P\{T \geqslant c(1-\alpha)\} \leqslant \alpha,$$

as desired.

Web Appendix F.2 *Proof of Lemma S6*

In this section, we show that $|T - T_{00}|$ is upper bounded by the quantity

$$C \cdot \left\{ \sqrt{nh^5} + s \cdot \sqrt{nh^9} + \frac{s \cdot \log(d/h)}{\sqrt{nh}} + \cdot s \cdot h^2 \cdot \sqrt{\log(d/h)} \right\}$$

with high probability for sufficiently large constant $C > 0$. By the triangle inequality, we have $|T - T_{00}| \leqslant |T - T_0| + |T_0 - T_{00}|$. Thus, is suffices to obtain upper bounds for the terms $|T - T_0|$ and $|T_0 - T_{00}|$.

**Upper Bound for $|T - T_0|$:** Let $\widetilde{\boldsymbol{\Theta}}_k = \left(\widehat{\boldsymbol{\Theta}}_{1k}, \ldots, \widehat{\boldsymbol{\Theta}}_{(j-1)k}, \boldsymbol{\Theta}_{jk}, \widehat{\boldsymbol{\Theta}}_{(j+1)k}, \ldots, \widehat{\boldsymbol{\Theta}}_{dk}\right)^T \in \mathbb{R}^d$. Then, the statistics $T$ can be rewritten as

$$\begin{aligned} T &= \sup_{z\in[0,1]} \max_{(j,k)\in E(z)} \sqrt{nh} \cdot \left| \widehat{\boldsymbol{\Theta}}^{\text{de}}_{jk}(z) - \boldsymbol{\Theta}_{jk}(z) \right| \cdot \mathbb{P}_n(w_z) \\ &= \sup_{z\in[0,1]} \max_{(j,k)\in E(z)} \sqrt{nh} \cdot \left| \widehat{\boldsymbol{\Theta}}_{jk}(z) - \boldsymbol{\Theta}_{jk}(z) - \frac{\left\{\widehat{\boldsymbol{\Theta}}_j(z)\right\}^T \left\{\widehat{\boldsymbol{\Sigma}}(z)\widehat{\boldsymbol{\Theta}}_k - \mathbf{e}_k\right\}}{\left\{\widehat{\boldsymbol{\Theta}}_j(z)\right\}^T \widehat{\boldsymbol{\Sigma}}_j(z)} \right| \cdot \mathbb{P}_n(w_z) \qquad \text{(S78)} \\ &= \sup_{z\in[0,1]} \max_{(j,k)\in E(z)} \sqrt{nh} \cdot \left| \frac{\left\{\widehat{\boldsymbol{\Theta}}_j(z)\right\}^T \left\{\widehat{\boldsymbol{\Sigma}}(z)\widetilde{\boldsymbol{\Theta}}_k - \mathbf{e}_k\right\}}{\left\{\widehat{\boldsymbol{\Theta}}_j(z)\right\}^T \widehat{\boldsymbol{\Sigma}}_j(z)} \right| \cdot \mathbb{P}_n(w_z). \end{aligned}$$

To obtain an upper bound on the difference between $T$ and $T_0$, we make use of the following inequality:

$$\left| \frac{x}{1+\delta} - y \right| \leqslant 2 \cdot y \cdot |\delta| + 2 \cdot |x - y| \qquad \text{for any } |\delta| \leqslant \frac{1}{2}. \qquad \text{(S79)}$$

Recall from (S60) that

$$T_0 = \sup_{z\in[0,1]} \max_{(j,k)\in E(z)} \sqrt{nh} \cdot \left| \sum_{i\in[n]} \{\boldsymbol{\Theta}_j(z)\}^T K_h(Z_i - z) \left\{ \boldsymbol{X}_i \boldsymbol{Y}_i^T - \boldsymbol{\Sigma}(z) \right\} \boldsymbol{\Theta}_k(z)/n \right|.$$

Applying (S79) with $x = \{\widehat{\boldsymbol{\Theta}}_j(z)\}^T\{\widehat{\boldsymbol{\Sigma}}(z)\widetilde{\boldsymbol{\Theta}}_k - \mathbf{e}_k\}$, $\delta = \{\widehat{\boldsymbol{\Theta}}_j(z)\}^T\widehat{\boldsymbol{\Sigma}}_j(z) - 1$, and $y =$

$\{\boldsymbol{\Theta}_j(z)\}^T\{\widehat{\boldsymbol{\Sigma}}(z)-\boldsymbol{\Sigma}(z)\}\boldsymbol{\Theta}_k(z)$, and by the triangle inequality, we have

$$
\begin{aligned}
&|T-T_0|\\
&\leqslant \sup_{z\in[0,1]}\max_{(j,k)\in E(z)}\sqrt{nh}\cdot\left|\frac{\left\{\widehat{\boldsymbol{\Theta}}_j(z)\right\}^T\left\{\widehat{\boldsymbol{\Sigma}}(z)\widetilde{\boldsymbol{\Theta}}_k-\mathbf{e}_k\right\}\cdot\mathbb{P}_n(w_z)}{\left\{\widehat{\boldsymbol{\Theta}}_j(z)\right\}^T\widehat{\boldsymbol{\Sigma}}_j(z)}-\frac{1}{n}\sum_{i\in[n]}\{\boldsymbol{\Theta}_j(z)\}^T K_h(Z_i-z)\left\{\boldsymbol{X}_i\boldsymbol{Y}_i^T-\boldsymbol{\Sigma}(z)\right\}\boldsymbol{\Theta}_k(z)\right|\\
&\leqslant \sup_{z\in[0,1]}\max_{(j,k)\in E(z)}\sqrt{nh}\cdot\left|\frac{\left\{\widehat{\boldsymbol{\Theta}}_j(z)\right\}^T\left\{\widehat{\boldsymbol{\Sigma}}(z)\widetilde{\boldsymbol{\Theta}}_k-\mathbf{e}_k\right\}}{\left\{\widehat{\boldsymbol{\Theta}}_j(z)\right\}^T\widehat{\boldsymbol{\Sigma}}_j(z)}-\{\boldsymbol{\Theta}_j(z)\}^T\left\{\widehat{\boldsymbol{\Sigma}}(z)-\boldsymbol{\Sigma}(z)\right\}\boldsymbol{\Theta}_k(z)\right|\cdot|\mathbb{P}_n(w_z)|\\
&\leqslant \underbrace{2\sup_{z\in[0,1]}\max_{(j,k)\in E(z)}\sqrt{nh}\cdot\left[\{\boldsymbol{\Theta}_j(z)\}^T\left\{\widehat{\boldsymbol{\Sigma}}(z)-\boldsymbol{\Sigma}(z)\right\}\boldsymbol{\Theta}_k(z)\cdot\left|\left\{\widehat{\boldsymbol{\Theta}}_j(z)\right\}^T\widehat{\boldsymbol{\Sigma}}_j(z)-1\right|\right]\cdot|\mathbb{P}_n(w_z)|}_{I_1}\\
&+\underbrace{2\sup_{z\in[0,1]}\max_{(j,k)\in E(z)}\sqrt{nh}\cdot\left[\left\{\widehat{\boldsymbol{\Theta}}_j(z)\right\}^T\left\{\widehat{\boldsymbol{\Sigma}}(z)\widetilde{\boldsymbol{\Theta}}_k-\mathbf{e}_k\right\}-\{\boldsymbol{\Theta}_j(z)\}^T\left\{\widehat{\boldsymbol{\Sigma}}(z)-\boldsymbol{\Sigma}(z)\right\}\boldsymbol{\Theta}_k(z)\right]\cdot|\mathbb{P}_n(w_z)|}_{I_2}.
\end{aligned}
\tag{S80}
$$

It remains to obtain upper bounds for $I_1$ and $I_2$ in (S80).

**Upper bound for $I_1$:** By Corollary 1, we have

$$\sup_{z\in[0,1]}\max_{j\in[d]}\left|\left\{\widehat{\boldsymbol{\Theta}}_j(z)\right\}^T\widehat{\boldsymbol{\Sigma}}_j(z)-1\right|\leqslant C\cdot\left[h^2+\sqrt{\frac{\log(d/h)}{nh}}\right]. \tag{S81}$$

Moreover, by Lemmas S4 and S2, we have

$$\sup_{z\in[0,1]}|\mathbb{P}_n(w_z)|\leqslant|\mathbb{E}\{\mathbb{P}_n(w_z)\}|+C\cdot\sqrt{\frac{\log(d/h)}{nh}}=\bar{f}_Z+\mathcal{O}\left\{h^2+\sqrt{\frac{\log(d/h)}{nh}}\right\}, \tag{S82}$$

with probability at least $1-1/d$. Thus, by Holder's inequality, we have

$$
\begin{aligned}
I_1&\leqslant 2\sup_{z\in[0,1]}\max_{(j,k)\in E(z)}\sqrt{nh}\cdot\left|\left\{\widehat{\boldsymbol{\Theta}}_j(z)\right\}^T\widehat{\boldsymbol{\Sigma}}_j(z)-1\right|\cdot|\mathbb{P}_n(w_z)|\cdot\left|\{\boldsymbol{\Theta}_j(z)\}^T\left\{\widehat{\boldsymbol{\Sigma}}(z)-\boldsymbol{\Sigma}(z)\right\}\boldsymbol{\Theta}_k(z)\right|\\
&\leqslant 2\sup_{z\in[0,1]}\max_{(j,k)\in E(z)}\sqrt{nh}\cdot\left|\left\{\widehat{\boldsymbol{\Theta}}_j(z)\right\}^T\widehat{\boldsymbol{\Sigma}}_j(z)-1\right|\cdot|\mathbb{P}_n(w_z)|\cdot\|\boldsymbol{\Theta}_j(z)\|_1^2\cdot\|\widehat{\boldsymbol{\Sigma}}(z)-\boldsymbol{\Sigma}(z)\|_{\max}\\
&\leqslant 2\cdot M^2\cdot\sqrt{nh}\cdot C\cdot\left\{h^2+\sqrt{\frac{\log(d/h)}{nh}}\right\}\cdot\left[\bar{f}_Z+\mathcal{O}\left\{h^2+\sqrt{\frac{\log(d/h)}{nh}}\right\}\right]\cdot\left\{h^2+\sqrt{\frac{\log(d/h)}{nh}}\right\}\\
&\leqslant C\cdot\sqrt{nh}\cdot\left\{h^2+\sqrt{\frac{\log(d/h)}{nh}}\right\}^2,
\end{aligned}
\tag{S83}
$$

with probability greater than $1-4/d$, where the third inequality holds by Theorem 1, (S81), and (S82).

**Upper bound for $I_2$:** To obtain an upper bound for $I_2$, we first decompose the quantity

$\sqrt{nh}\cdot\{\widehat{\boldsymbol{\Theta}}_j(z)\}^T\{\widehat{\boldsymbol{\Sigma}}(z)\widetilde{\boldsymbol{\Theta}}_k-\mathbf{e}_k\}$ into the following

$$\begin{aligned}&\sqrt{nh}\cdot\left\{\widehat{\boldsymbol{\Theta}}_j(z)\right\}^T\left\{\widehat{\boldsymbol{\Sigma}}(z)\widetilde{\boldsymbol{\Theta}}_k-\mathbf{e}_k\right\}\\&=\underbrace{\sqrt{nh}\cdot\left\{\widehat{\boldsymbol{\Theta}}_j(z)\right\}^T\widehat{\boldsymbol{\Sigma}}(z)\left\{\widetilde{\boldsymbol{\Theta}}_k(z)-\boldsymbol{\Theta}_k(z)\right\}}_{I_{21}}+\underbrace{\sqrt{nh}\cdot\left\{\widehat{\boldsymbol{\Theta}}_j(z)\right\}^T\left\{\widehat{\boldsymbol{\Sigma}}(z)\;-\boldsymbol{\Sigma}(z)\right\}\boldsymbol{\Theta}_k(z)}_{I_{22}}.\end{aligned}$$

Next, we show that $I_{21}$ converges to zero and that the difference between $I_{22}$ and the term $\sqrt{nh}\cdot\{\boldsymbol{\Theta}_j(z)\}^T\{\widehat{\boldsymbol{\Sigma}}(z)-\boldsymbol{\Sigma}(z)\}\boldsymbol{\Theta}_k(z)$ is small.

**Upper bound for $I_{21}$:** By Holder's inequality and Corollary 1, we have

$$\begin{aligned}|I_{21}|&\leqslant\sup_{z\in[0,1]}\max_{(j,k)\in E(z)}\sqrt{nh}\cdot\left\|\left\{\widehat{\boldsymbol{\Theta}}_j(z)\right\}^T\widehat{\boldsymbol{\Sigma}}_{-j}(z)\right\|_\infty\cdot\left\|\widehat{\boldsymbol{\Theta}}_k(z)-\boldsymbol{\Theta}_k(z)\right\|_1\\&\leqslant C\cdot\sqrt{nh}\cdot s\cdot\left\{h^2+\sqrt{\frac{\log(d/h)}{nh}}\right\}^2\\&\leqslant C\cdot\left\{s\cdot\sqrt{nh^9}+\frac{s\cdot\log(d/h)}{\sqrt{nh}}+s\cdot h^2\cdot\sqrt{\log(d/h)}\right\},\end{aligned}\tag{S84}$$

with probability at least $1-1/d$.

**Decomposition of $I_{22}$:** By adding and subtracting terms, we have

$$I_{22}=\underbrace{\sqrt{nh}\cdot\left\{\widehat{\boldsymbol{\Theta}}_j(z)-\boldsymbol{\Theta}_j(z)\right\}^T\left\{\widehat{\boldsymbol{\Sigma}}(z)-\boldsymbol{\Sigma}(z)\right\}\boldsymbol{\Theta}_k(z)}_{I_{221}}+\underbrace{\sqrt{nh}\cdot\{\boldsymbol{\Theta}_j(z)\}^T\left\{\widehat{\boldsymbol{\Sigma}}(z)-\boldsymbol{\Sigma}(z)\right\}\boldsymbol{\Theta}_k(z)}_{I_{222}}.\tag{S85}$$

Similar to (S84), we have

$$\begin{aligned}|I_{221}|&\leqslant\sup_{z\in[0,1]}\max_{(j,k)\in E(z)}\sqrt{nh}\cdot\left\|\widehat{\boldsymbol{\Theta}}_j(z)-\boldsymbol{\Theta}_j(z)\right\|_1\cdot\left\|\widehat{\boldsymbol{\Sigma}}(z)-\boldsymbol{\Sigma}(z)\right\|_{\max}\cdot\|\boldsymbol{\Theta}_k(z)\|_1\\&\leqslant C\cdot\sqrt{nh}\cdot M\cdot s\cdot\left\{h^2+\sqrt{\frac{\log(d/h)}{nh}}\right\}^2\\&\leqslant C\cdot\left\{s\cdot\sqrt{nh^9}+\frac{s\cdot\log(d/h)}{\sqrt{nh}}+s\cdot h^2\cdot\sqrt{\log(d/h)}\right\},\end{aligned}\tag{S86}$$

where the second inequality holds by Holder's inequality, Corollary 1, and the fact that $\boldsymbol{\Theta}(z)\in\mathcal{U}_{s,M}$.

Combining the results (S84)-(S86), we have

$$
\begin{aligned}
I_2 &= 2 \sup_{z\in[0,1]} \max_{(j,k)\in E(z)} \sqrt{nh}\cdot\left[\left\{\widehat{\boldsymbol{\Theta}}_j(z)\right\}^T\left\{\widehat{\boldsymbol{\Sigma}}(z)\widetilde{\boldsymbol{\Theta}}_k-\mathbf{e}_k\right\}-\{\boldsymbol{\Theta}_j(z)\}^T\left\{\widehat{\boldsymbol{\Sigma}}(z)-\boldsymbol{\Sigma}(z)\right\}\boldsymbol{\Theta}_k(z)\right]\cdot|\mathbb{P}_n(w_z)|\\
&\leqslant 2\cdot\sup_{z\in[0,1]}|\mathbb{P}_n(w_z)|\cdot[I_{21}+I_{221}]\\
&\leqslant 2\cdot\left[\bar{f}_Z+\mathcal{O}\left\{h^2+\sqrt{\frac{\log(d/h)}{nh}}\right\}\right]\cdot(I_{21}+I_{221})\\
&\leqslant C\cdot\left\{s\cdot\sqrt{nh^9}+\frac{s\cdot\log(d/h)}{\sqrt{nh}}+s\cdot h^2\cdot\sqrt{\log(d/h)}\right\},
\end{aligned}
\tag{S87}
$$

where the third inequality follows from (S82).

Combining the upper bounds for $I_1$ in (S83) and $I_2$ in (S87), we have

$$
|T-T_0|\leqslant C\cdot\left\{s\cdot\sqrt{nh^9}+\frac{s\cdot\log(d/h)}{\sqrt{nh}}+s\cdot h^2\cdot\sqrt{\log(d/h)}\right\},
\tag{S88}
$$

with probability at least $1-1/d$.

**Upper bound for $|T_0-T_{00}|$:** Recall from (S61) the definition of $T_{00}$

$$
\begin{aligned}
T_{00}=\sup_{z\in[0,1]}\max_{(j,k)\in E(z)}\sqrt{nh}\cdot\Bigg|\sum_{i\in[n]}\{\boldsymbol{\Theta}_j(z)\}^T K_h(Z_i-z)\left\{\boldsymbol{X}_i\boldsymbol{Y}_i^T-\boldsymbol{\Sigma}(z)\right\}\boldsymbol{\Theta}_k(z)/n\\
-\{\boldsymbol{\Theta}_j(z)\}^T\left[\mathbb{E}\{K_h(Z-z)\boldsymbol{X}\boldsymbol{Y}^T\}-\mathbb{E}\{K_h(Z-z)\}\boldsymbol{\Sigma}(z)\right]\boldsymbol{\Theta}_k(z)/n\Bigg|;
\end{aligned}
$$

Using the triangle inequality $||x|-|y||\leqslant|x-y|$, we obtain

$$
\begin{aligned}
|T_0-T_{00}| &\leqslant \sqrt{nh}\cdot\sup_{z\in[0,1]}\max_{(j,k)\in E(z)}\left|\{\boldsymbol{\Theta}_j(z)\}^T\left[\mathbb{E}\{K_h(Z-z)\boldsymbol{X}\boldsymbol{Y}^T\}-\mathbb{E}\{K_h(Z-z)\}\boldsymbol{\Sigma}(z)\right]\boldsymbol{\Theta}_k(z)\right|\\
&\leqslant \sqrt{nh}\cdot\sup_{z\in[0,1]}\max_{(j,k)\in E(z)}\|\boldsymbol{\Theta}_j(z)\|_1\cdot\|\boldsymbol{\Theta}_k(z)\|_1\cdot|\mathbb{E}\{K_h(Z-z)X_jY_k\}-\mathbb{E}\{K_h(Z-z)\}\cdot\boldsymbol{\Sigma}_{jk}(z)|\\
&\leqslant \sqrt{nh}\cdot M^2\cdot\sup_{z\in[0,1]}\max_{(j,k)\in E(z)}|\mathbb{E}\{K_h(Z-z)X_jY_k\}-\mathbb{E}\{K_h(Z-z)\}\cdot\boldsymbol{\Sigma}_{jk}(z)|\\
&= \sqrt{nh}\cdot M^2\cdot\left|f_Z(z)\cdot\boldsymbol{\Sigma}_{jk}(z)+\mathcal{O}(h^2)-f_Z(z)\cdot\boldsymbol{\Sigma}_{jk}(z)+\boldsymbol{\Sigma}_{jk}(z)\cdot\mathcal{O}(h^2)\right|\\
&\leqslant M^2\cdot M_\sigma\cdot\sqrt{nh^5},
\end{aligned}
\tag{S89}
$$

where the second inequality follows from an application of Holder's inequality, the third inequality follows from the fact that $\boldsymbol{\Theta}(z) \in \mathcal{U}_{s,M}$, the first equality follows by an application of Lemma S2, and the last inequality follows from Assumption 2 and that $h^2 = o(1)$.

Thus, combining (S88) and (S89), there exists a constant $C > 0$ such that

$$|T - T_{00}| \leqslant C \cdot \left\{ \sqrt{nh^5} + s \cdot \sqrt{nh^9} + \frac{s \cdot \log(d/h)}{\sqrt{nh}} + \cdot s \cdot h^2 \cdot \sqrt{\log(d/h)} \right\},$$

with probability at least $1 - 1/d$.

Web Appendix F.3 *Proof of Lemma S7*

Recall from (S61) the definition

$$T_{00} = \sup_{z\in[0,1]} \max_{(j,k)\in E(z)} \sqrt{nh} \cdot \Bigg| \sum_{i\in[n]} \{\boldsymbol{\Theta}_j(z)\}^T K_h(Z_i - z) \left\{ \boldsymbol{X}_i \boldsymbol{Y}_i^T - \boldsymbol{\Sigma}(z) \right\} \boldsymbol{\Theta}_k(z)/n$$
$$- \{\boldsymbol{\Theta}_j(z)\}^T \left[ \mathbb{E}\{K_h(Z - z)\boldsymbol{X}\boldsymbol{Y}^T\} - \mathbb{E}\{K_h(Z - z)\}\boldsymbol{\Sigma}(z) \right] \boldsymbol{\Theta}_k(z)/n \Bigg|.$$

Recall from (S15) that $J_{z,jk}(Z_i, \boldsymbol{X}_i, \boldsymbol{Y}_i) = J^{(1)}_{z,jk}(Z_i, \boldsymbol{X}_i, \boldsymbol{Y}_i) - J^{(2)}_{z,jk}(Z_i)$, where $J^{(1)}_{z,jk}(Z_i, \boldsymbol{X}_i, \boldsymbol{Y}_i)$ and $J^{(2)}_{z,jk}(Z_i)$ are as defined in (S13) and (S14), respectively. Let $\mathcal{J} = \{J_{z,jk} \mid z \in [0,1],\ j,k \in [d]\}$. Then the intermediate empirical average $T_{00}$ can be written as

$$T_{00} = \sup_{z\in[0,1]} \max_{(j,k)\in E(z)} \left| \frac{1}{\sqrt{n}} \sum_{i\in[n]} J_{z,jk}(Z_i, \boldsymbol{X}_i, \boldsymbol{Y}_i) \right|.$$

In this section, we show that there exists a Gaussian process $W$ such that

$$|T_{00} - W| \leqslant C \cdot \left\{ \frac{\log^6 s \cdot \log^4(d/h)}{nh} \right\}^{1/8}$$

with high probability. To this end, we apply Theorem A.1 in Chernozhukov et al. (2014a), which involves the following quantities

- upper bound for $\sup_{z\in[0,1]} \max_{j,k\in[d]} \|J_{z,jk}(Z_i, \boldsymbol{X}_i, \boldsymbol{Y}_i)\|_\infty$;
- upper bound for $\sup_{z\in[0,1]} \max_{j,k\in[d]} \mathbb{E}\left\{J^2_{z,jk}(Z, \boldsymbol{X}, \boldsymbol{Y})\right\}$;
- covering number for the function class $\mathcal{J}$.

Let $\mathcal{S}_j(z)$ and $\mathcal{S}_k(z)$ to be the support of $\boldsymbol{\Theta}_j(z)$ and $\boldsymbol{\Theta}_k(z)$, respectively. Note that the cardinality for both sets are less than $s$. We now obtain the above quantities.

**Upper bound for $\sup_{z\in[0,1]}\max_{j,k\in[d]}\|J_{z,jk}(Z_i,\boldsymbol{X}_i,\boldsymbol{Y}_i)\|_\infty$:** We have with probability at least $1-1/(2s)$,

$$
\begin{aligned}
&\sup_{z\in[0,1]}\max_{j,k\in[d]}\|J_{z,jk}(Z_i,\boldsymbol{X}_i,\boldsymbol{Y}_i)\|_\infty\\
&\leqslant \sqrt{h}\cdot\sup_{z\in[0,1]}\max_{j,k\in[d]}\|\boldsymbol{\Theta}_j(z)\|_1\cdot\|\boldsymbol{\Theta}_k(z)\|_1\cdot\left(\max_{j\in\mathcal{S}_j(z),k\in\mathcal{S}_k(z)}\|q_{z,jk}\|_\infty+M_\sigma\cdot\|k_z\|_\infty\right)\\
&\leqslant \sqrt{h}\cdot M^2\cdot\left\{\frac{2}{h}\cdot M_X^2\cdot\|K\|_\infty\cdot\log(2s)+M_\sigma\cdot\frac{2}{h}\cdot\|K\|_\infty\right\} \qquad\qquad (S90)\\
&\leqslant \frac{4}{\sqrt{h}}\cdot M^2\cdot M_X^2\cdot M_\sigma\cdot\|K\|_\infty\cdot\log(2s)\\
&= C_1\cdot\frac{\log s}{\sqrt{h}},
\end{aligned}
$$

where the first inequality follows by Holder's inequality and the definition of $q_{z,jk}$ and $k_z$ and the second inequality follows from (S47) and (S54). Note that since we are only taking max over the set $\mathcal{S}_j(z)$ and $\mathcal{S}_k(z)$, instead of a $\log d$ factor from (S54), we obtain a $\log(2s)$ factor.

**Upper bound for $\sup_{z\in[0,1]}\max_{j,k\in[d]}\mathbb{E}\{J^2_{z,jk}(Z,\boldsymbol{X},\boldsymbol{Y})\}$:** By an application of the inequality $(x-y)^2\leqslant 2x^2+2y^2$, we have

$$
\begin{aligned}
\sup_{z\in[0,1]}\max_{j,k\in[d]}\mathbb{E}\left\{J^2_{z,jk}(Z,\boldsymbol{X},\boldsymbol{Y})\right\} &= \sup_{z\in[0,1]}\max_{j,k\in[d]}\mathbb{E}\left[\left\{J^{(1)}_{z,jk}(Z,\boldsymbol{X},\boldsymbol{Y})-J^{(2)}_{z,jk}(Z)\right\}^2\right]\\
&\leqslant \underbrace{2\sup_{z\in[0,1]}\max_{j,k\in[d]}\mathbb{E}\left[\left\{J^{(1)}_{z,jk}(Z,\boldsymbol{X},\boldsymbol{Y})\right\}^2\right]}_{I_1}+\underbrace{2\sup_{z\in[0,1]}\max_{j,k\in[d]}\mathbb{E}\left[\left\{J^{(2)}_{z,jk}(Z)\right\}^2\right]}_{I_2}.
\end{aligned}
$$

To obtain an upper bound for $I_1$, we need an upper bound for $\sup_{z\in[0,1]}\max_{j,k\in[d]}\mathbb{E}\{\max_{j\in\mathcal{S}_j(z),k\in\mathcal{S}_k(z)}q^2_{z,jk}\}$. Recall from (S9) the definition of $g_{z,jk}(Z_i,X_{ij},Y_{ik})=K_h(Z_i-z)X_{ij}Y_{ik}$ and that $q_{z,jk}(Z_i,X_{ij},Y_{ik})=g_{z,jk}(Z_i,X_{ij},Y_{ik})-\mathbb{E}\{g_{z,jk}(Z,X_j,Y_k)\}$. Thus, we have

$$
\begin{aligned}
\sup_{z\in[0,1]}\max_{j,k\in[d]}\mathbb{E}\left\{\max_{j\in\mathcal{S}_j(z),k\in\mathcal{S}_k(z)}q^2_{z,jk}\right\} &= \sup_{z\in[0,1]}\max_{j,k\in[d]}\mathbb{E}\left[\max_{j\in\mathcal{S}_j(z),k\in\mathcal{S}_k(z)\in[d]}\{g_{z,jk}-\mathbb{E}(g_{z,jk})\}^2\right]\\
&\leqslant 2\sup_{z\in[0,1]}\max_{j,k\in[d]}\mathbb{E}\left\{\max_{j\in\mathcal{S}_j(z),k\in\mathcal{S}_k(z)\in[d]}g^2_{z,jk}\right\}+2\sup_{z\in[0,1]}\max_{j,k\in[d]}\mathbb{E}^2(g_{z,jk}),
\end{aligned}
\qquad (S91)
$$

where we apply the fact that $(x-y)^2\leqslant 2x^2+2y^2$ to obtain the last inequality. By Lemma S2, we have $2\sup_{z\in[0,1]}\max_{j,k\in[d]}\mathbb{E}^2(g_{z,jk})\leqslant 2\left\{\bar{f}_Z\cdot M_\sigma+\mathcal{O}(h^2)\right\}^2$. Moreover, we have

$$
\begin{aligned}
2\sup_{z\in[0,1]}\max_{j,k\in[d]}\mathbb{E}\left\{\max_{j\in\mathcal{S}_j(z),k\in\mathcal{S}_k(z)\in[d]} g_{z,jk}^2\right\} &= 2\sup_{z\in[0,1]}\max_{j,k\in[d]}\mathbb{E}\left\{\max_{j\in\mathcal{S}_j(z),k\in\mathcal{S}_k(z)\in[d]} K_h^2(Z-z)X_j^2Y_k^2\right\}\\
&\leqslant 2\cdot M_X^4\cdot\log^2(2s)\sup_{z\in[0,1]}\max_{j,k\in[d]}\mathbb{E}\left\{K_h^2(Z-z)\right\}\\
&\leqslant 2\cdot M_X^4\cdot\log^2(2s)\cdot\left\{\frac{1}{h}\cdot\bar{f}_Z\cdot\|K\|_2^2+\mathcal{O}(1)+\mathcal{O}(h^2)\right\}\\
&\leqslant 3\cdot\bar{f}_Z\cdot\|K\|_2^2\cdot M_X^4\cdot\frac{\log^2(2s)}{h},
\end{aligned}
$$

with probability at least $1-1/(2s)$, where the second inequality follows from an application of Lemma S2.

Thus, by Holder's inequality, we have

$$
\begin{aligned}
I_1 &\leqslant 2\cdot h\cdot\sup_{z\in[0,1]}\max_{j,k\in[d]}\mathbb{E}\left[\left\{\|\boldsymbol{\Theta}_j(z)\|_1\cdot\|\boldsymbol{\Theta}_k(z)\|_1\cdot\max_{j\in\mathcal{S}_j(z),k\in\mathcal{S}_k(z)}|q_{z,jk}|\right\}^2\right]\\
&\leqslant 2\cdot h\cdot M^4\cdot\sup_{z\in[0,1]}\max_{j,k\in[d]}\mathbb{E}\left\{\max_{j\in\mathcal{S}_j(z),k\in\mathcal{S}_k(z)} q_{z,jk}^2\right\}\\
&\leqslant 2\cdot h\cdot M^4\cdot\left[3\cdot\bar{f}_Z\cdot\|K\|_2^2\cdot M_X^4\cdot\frac{\log^2(2s)}{h}+2\left\{\bar{f}_Z\cdot M_\sigma+\mathcal{O}(h^2)\right\}^2\right]\\
&\leqslant 8\cdot M^4\cdot\bar{f}_Z\cdot M_X^4\cdot\|K\|_2^2\cdot\log^2(2s),
\end{aligned}
\tag{S92}
$$

where the second inequality holds by the fact that $\boldsymbol{\Theta}(z)\in\mathcal{U}_{s,M}$.

Similarly, to obtain an upper bound for $I_2$, we use the fact from (S49) that

$$
\sup_{z\in[0,1]}\mathbb{E}\left\{k_z^2\right\}\leqslant\frac{3}{h}\cdot\bar{f}_Z\cdot\|K\|_2^2. \tag{S93}
$$

By Holder's inequality, we have

$$
\begin{aligned}
I_2 &\leqslant 2\cdot h\cdot\sup_{z\in[0,1]}\max_{j,k\in[d]}\mathbb{E}\left[\left\{\|\boldsymbol{\Theta}_j(z)\|_1\cdot\|\boldsymbol{\Theta}_k(z)\|_1\cdot\max_{(j,k)\in E(z)}|\boldsymbol{\Sigma}_{jk}(z)|\cdot|k_z|\right\}^2\right]\\
&\leqslant 2\cdot h\cdot M^4\cdot M_\sigma^2\cdot\sup_{z\in[0,1]}\mathbb{E}\left(k_z^2\right)\\
&\leqslant 6\cdot M_\sigma^2\cdot M^4\cdot\bar{f}_Z\cdot\|K\|_2^2,
\end{aligned}
\tag{S94}
$$

where the second inequality holds by Assumption 2 and by the fact that $\boldsymbol{\Theta}(z)\in\mathcal{U}_{s,M}$, and the last inequality holds by (S93).

Combining the upper bounds for $I_1$ (S92) and $I_2$ (S94), we have

$$\sup_{z\in[0,1]} \max_{j,k\in[d]} \mathbb{E}\left\{J^2_{z,jk}(Z,\boldsymbol{X},\boldsymbol{Y})\right\} \leqslant 8\cdot M^4\cdot \bar{f}_Z\cdot \|K\|_2^2\cdot \left\{M_\sigma^2 + M_X^4\cdot \log^2(2s)\right\} \leqslant C\cdot \log^2 s = \sigma_J^2, \tag{S95}$$

for sufficiently large $C > 0$.

**Covering number of the function class $\mathcal{J}$:** First, we note that the function class $\mathcal{J}$ is generated from the addition of two function classes

$$\mathcal{J}^{(1)}_{jk} = \left\{J^{(1)}_{z,jk} \mid z\in[0,1]\right\} \qquad \text{and} \qquad \mathcal{J}^{(2)}_{jk} = \left\{J^{(2)}_{z,jk} \mid z\in[0,1]\right\}.$$

Thus, to obtain the covering number of $\mathcal{J}$, we first obtain the covering number for the function classes $\mathcal{J}^{(1)}_{jk}$ and $\mathcal{J}^{(2)}_{jk}$. Then, we apply Lemma S15 to obtain the covering number of the function class $\mathcal{J}$. From Lemma S18, we have with probability at least $1-1/d$,

$$N\{\mathcal{J}^{(1)}_{jk}, L_2(Q),\epsilon\} \leqslant C\cdot \left(\frac{d^{5/4}\cdot \log^{3/2} d}{\sqrt{h}\cdot\epsilon}\right)^6.$$

Moreover, from Lemma S19, we have

$$N\{\mathcal{J}^{(2)}_{jk}, L_2(Q),\epsilon\} \leqslant C\cdot \left(\frac{d^{1/6}}{h^{4/3}\cdot\epsilon}\right)^6.$$

Applying Lemma S15 with $a_1 = d^{5/4}\cdot\log^{3/2} d/h^{1/2}$, $v_1 = 6$, $a_2 = d^{1/6}/h^{4/3}$, and $v_2 = 6$, we have

$$N\{\mathcal{J}, L_2(Q),\epsilon\} \leqslant C\cdot d^2\cdot \left(\frac{d^{17/24}\cdot \log^{3/4} d}{h^{11/12}\cdot\epsilon}\right)^{12}, \tag{S96}$$

where we multiply $d^2$ on the right hand side since the function class $\mathcal{J}$ is taken over all $j,k\in[d]$.

**Application of Theorem A.1 in Chernozhukov et al. (2014a):** Applying Theorem A.1 in Chernozhukov et al. (2014a) with $a = d^{65/24}\cdot\log^{7/4} d/h^{17/12}$, $b = C\cdot \log s/\sqrt{h}$, $\sigma_J = C\cdot\log s$, and

$$K_n = A\cdot\{\log n \vee \log(ab/\sigma_J)\} = C\cdot \log(d/h),$$

for sufficiently large constant $A, C > 0$, there exists a random process $W$ such that for any $\gamma \in (0, 1)$,

$$P\left[|T_{00} - W| \geqslant C \cdot \left\{\frac{bK_n}{(\gamma n)^{1/2}} + \frac{(b\sigma_J)^{1/2}K_n^{3/4}}{\gamma^{1/2}n^{1/4}} + \frac{b^{1/3}\sigma_J^{2/3}K_n^{2/3}}{\gamma^{1/3}n^{1/6}}\right\}\right] \leqslant C' \cdot \left(\gamma + \frac{\log n}{n}\right)$$

for some absolute constant $C'$. Picking $\gamma = \left\{\log^6 s \cdot \log^4(d/h)/(nh)\right\}^{1/8}$, we have

$$P\left[|T_{00} - W| \geqslant C \cdot \left\{\frac{\log^6 s \cdot \log^4(d/h)}{nh}\right\}^{1/8}\right] \leqslant C' \cdot \left\{\frac{\log^6 s \cdot \log^4(d/h)}{nh}\right\}^{1/8},$$

as desired.

### Web Appendix F.4 *Proof of Lemma S8*

Recall from the proof of Lemma S7 that

$$T_{00} = \sup_{z \in [0,1]} \max_{(j,k) \in E(z)} \left| \frac{1}{\sqrt{n}} \sum_{i \in [n]} J_{z,jk}(Z_i, \boldsymbol{X}_i, \boldsymbol{Y}_i) \right|.$$

We note that

$$T_{00}^B = \sup_{z \in [0,1]} \max_{(j,k) \in E(z)} \left| \frac{1}{\sqrt{n}} \sum_{i \in [n]} J_{z,jk}(Z_i, \boldsymbol{X}_i, \boldsymbol{Y}_i) \cdot \xi_i \right|,$$

where $\xi_i \overset{\text{i.i.d.}}{\sim} N(0, 1)$. To show that the term $|W - T_{00}^B|$ can be controlled, we apply Theorem A.2 in Chernozhukov et al. (2014a).

Let

$$\psi_n = \sqrt{\frac{\sigma_J^2 K_n}{n}} + \left(\frac{b^2\sigma_J^2 K_n^3}{n}\right)^{1/4} \qquad \text{and} \qquad \gamma_n(\delta) = \frac{1}{\delta}\left(\frac{b^2\sigma_J^2 K_n^3}{n}\right)^{1/4} + \frac{1}{n},$$

as defined in Theorem A.2 in Chernozhukov et al. (2014a). From the proof of Lemma S7, we have $b = C \cdot \log s/\sqrt{h}$, $K_n = C \cdot \log(d/h)$, and $\sigma_J = C \cdot \log s$. Since $b^2 K_n = C \cdot \log^2 s \cdot \log(d/h)/h \leqslant n \cdot \log^2 s$ for sufficiently large $n$, there exists a constant $C'' > 0$ such that

$$P\left[|T_{00}^B - W| > \psi_n + \delta \;\middle|\; \{(Z_i, \boldsymbol{X}_i, \boldsymbol{Y}_i)\}_{i \in [n]}\right] \leqslant C'' \cdot \gamma_n(\delta),$$

with probability at least $1 - 3/n$. Choosing $\delta = \left\{\log^4(s) \cdot \log^3(d/h)/(nh)\right\}^{1/8}$, we have

$$P\left[|T_{00}^B - W| > C \cdot \left\{\frac{\log^4(s) \cdot \log^3(d/h)}{nh}\right\}^{1/8} \;\middle|\; \{(Z_i, \boldsymbol{X}_i, \boldsymbol{Y}_i)\}_{i \in [n]}\right] \leqslant C'' \cdot \left\{\frac{\log^4(s) \cdot \log^3(d/h)}{nh}\right\}^{1/8},$$

with probability at least $1 - 3/n$.

Web Appendix F.5 *Proof of Lemma S9*

In this section, we show that $|T^B - T^B_{00}|$ is upper bounded by the quantity

$$C \cdot \left\{ s \cdot \sqrt{h^3 \log^3(d/h)} + s \cdot \sqrt{\frac{\log^4(d/h)}{nh^2}} + \sqrt{h^5 \log n} \right\}$$

with high probability for sufficiently large constant $C > 0$. Throughout the proof of this lemma, we conditioned on the data $\{(Z_i, \boldsymbol{X}_i, \boldsymbol{Y}_i)\}_{i\in[n]}$. By the triangle inequality, we have $|T^B - T^B_{00}| \leqslant |T^B - T^B_0| + |T^B_0 - T^B_{00}|$. Thus, it suffices to obtain upper bounds for the terms $|T^B - T^B_0|$ and $|T^B_0 - T^B_{00}|$.

**Upper bound for $|T^B - T^B_0|$:** Recall from (S59) and (S62) that

$$T^B = \sup_{z\in[0,1]} \max_{(j,k)\in E(z)} \sqrt{nh} \cdot \left| \frac{\sum_{i\in[n]} \left\{ \widehat{\boldsymbol{\Theta}}_j(z) \right\}^T K_h(Z_i - z) \left\{ \boldsymbol{X}_i \boldsymbol{Y}_i^T \widehat{\boldsymbol{\Theta}}_k(z) - \mathbf{e}_k \right\} \xi_i / n}{\left\{ \widehat{\boldsymbol{\Theta}}_j(z) \right\}^T \widehat{\boldsymbol{\Sigma}}_j(z)} \right|,$$

and that

$$T^B_0 = \sup_{z\in[0,1]} \max_{(j,k)\in E(z)} \sqrt{nh} \cdot \left| \sum_{i\in[n]} \left[ \left\{ \boldsymbol{\Theta}_j(z) \right\}^T K_h(Z_i - z) \left\{ \boldsymbol{X}_i \boldsymbol{Y}_i^T - \boldsymbol{\Sigma}(z) \right\} \boldsymbol{\Theta}_k(z) \right] \xi_i / n \right|,$$

respectively. Using the triangle inequality, we have

$$\begin{aligned}
|T^B - T^B_0| \leqslant \sqrt{nh} \cdot & \left| \sup_{z\in[0,1]} \max_{(j,k)\in E(z)} \left[ \frac{1}{n} \sum_{i\in[n]} \left\{ \widehat{\boldsymbol{\Theta}}_j(z) \right\}^T K_h(Z_i - z) \left\{ \boldsymbol{X}_i \boldsymbol{Y}_i^T \widehat{\boldsymbol{\Theta}}_k(z) - \mathbf{e}_k \right\} / \left\{ \widehat{\boldsymbol{\Theta}}_j(z) \right\}^T \widehat{\boldsymbol{\Sigma}}_j(z) \right.\right. \\
& \left.\left. - \frac{1}{n} \sum_{i\in[n]} \left\{ \boldsymbol{\Theta}_j(z) \right\}^T K_h(Z_i - z) \left\{ \boldsymbol{X}_i \boldsymbol{Y}_i^T - \boldsymbol{\Sigma}(z) \right\} \boldsymbol{\Theta}_k(z) \right] \xi_i \right| \\
\leqslant & \underbrace{2\sqrt{nh} \cdot \left| \sup_{z\in[0,1]} \max_{(j,k)\in E(z)} \frac{1}{n} \sum_{i\in[n]} \left\{ \widehat{\boldsymbol{\Theta}}_j(z) - \boldsymbol{\Theta}_j(z) \right\}^T K_h(Z_i - z) \left\{ \boldsymbol{X}_i \boldsymbol{Y}_i^T - \boldsymbol{\Sigma}(z) \right\} \boldsymbol{\Theta}_k(z) \xi_i \right|}_{I_1} \\
& + \underbrace{2\sqrt{nh} \cdot \left| \sup_{z\in[0,1]} \max_{(j,k)\in E(z)} \frac{1}{n} \sum_{i\in[n]} \left\{ \boldsymbol{\Theta}_j(z) \right\}^T K_h(Z_i - z) \boldsymbol{X}_i \boldsymbol{Y}_i^T \left( \widehat{\boldsymbol{\Theta}}_k(z) - \boldsymbol{\Theta}_k(z) \right) \xi_i \right|}_{I_2} \\
& + \underbrace{2\sqrt{nh} \cdot \left| \sup_{z\in[0,1]} \max_{(j,k)\in E(z)} \frac{1}{n} \sum_{i\in[n]} \left\{ \boldsymbol{\Theta}_j(z) \right\}^T K_h(Z_i - z) \left\{ \boldsymbol{X}_i \boldsymbol{Y}_i^T - \boldsymbol{\Sigma}(z) \right\} \boldsymbol{\Theta}_k(z) \xi_i \right| \cdot \left| \left\{ \widehat{\boldsymbol{\Theta}}_j(z) \right\}^T \widehat{\boldsymbol{\Sigma}}_j(z) - 1 \right|}_{I_3},
\end{aligned} \tag{S97}$$

where the second inequality holds by another application of the triangle inequality and inequality in (S79). We now obtain upper bounds for $I_1$, $I_2$, and $I_3$.

**Upper bound for $I_1$:** By an application of Holder's inequality, we have

$$I_1 \leqslant \sup_{z\in[0,1]} \max_{j,k\in[d]} \left\|\widehat{\boldsymbol{\Theta}}_j(z) - \boldsymbol{\Theta}_j(z)\right\|_1 \cdot \|\boldsymbol{\Theta}_k(z)\|_1 \cdot \sqrt{nh} \cdot \left|\sup_{z\in[0,1]} \max_{j,k\in[d]} \frac{1}{n}\sum_{i\in[n]} \{K_h(Z_i-z)X_{ij}Y_{ik} - K_h(Z_i-z)\boldsymbol{\Sigma}_{jk}(z)\}\,\xi_i\right|$$
$$\leqslant M\cdot C\cdot s\cdot\left\{h^2+\sqrt{\frac{\log(d/h)}{nh}}\right\}\cdot\sqrt{nh}\cdot\left|\sup_{z\in[0,1]} \max_{j,k\in[d]} \frac{1}{n}\sum_{i\in[n]} \{K_h(Z_i-z)X_{ij}Y_{ik} - K_h(Z_i-z)\boldsymbol{\Sigma}_{jk}(z)\}\,\xi_i\right|, \tag{S98}$$

where the last inequality follows from the fact that $\boldsymbol{\Theta}(z) \in \mathcal{U}_{s,M}$ and by an application of Corollary 1. For notational convenience, we use the notation as defined in (S16)

$$W_{z,jk}(Z_i, X_{ij}, Y_{ik}) = \sqrt{h}\cdot\{K_h(Z_i-z)X_{ij}Y_{ik} - K_h(Z_i-z)\boldsymbol{\Sigma}_{jk}(z)\}\,. \tag{S99}$$

Then, we have

$$\sqrt{\frac{h}{n}}\sum_{i\in[n]} \{K_h(Z_i-z)X_{ij}Y_{ik} - K_h(Z_i-z)\boldsymbol{\Sigma}_{jk}(z)\}\,\xi_i = \frac{1}{\sqrt{n}}\sum_{i\in[n]} W_{z,jk}(Z_i, X_{ij}, Y_{ik})\cdot\xi_i.$$

We note that conditioned on the data $\{(Z_i, \boldsymbol{X}_i, \boldsymbol{Y}_i)\}_{i\in[n]}$, the above expression is a Gaussian process. It remains to bound the supreme of the Gaussian process

$$\frac{1}{\sqrt{n}}\sum_{i\in[n]} W_{z,jk}(Z_i, X_{ij}, Y_{ik})\cdot\xi_i \sim N\left\{0, \frac{1}{n}\sum_{i\in[n]} W^2_{z,jk}(Z_i, X_{ij}, Y_{ik})\right\}$$

in probability.

To this end, we apply the Dudley's inequality (see, e.g., Corollary 2.2.8 in Van Der Vaart and Wellner, 1996) and the Borell's inequality (see, e.g., Proposition A.2.1 in Van Der Vaart and Wellner, 1996), which involves the following quantities:

- upper bound on the conditional variance $\sum_{i\in[n]} W^2_{z,jk}(Z_i, X_{ij}, Y_{ik})/n$;
- the covering number of the function class

$$\mathcal{W} = \{W_{z,jk}(\cdot) \mid z\in[0,1],\ j,k\in[d]\}$$

under the $L_2$ norm on the empirical measure.

**Upper bound for the conditional variance $\sum_{i=1}^n W^2_{z,jk}(Z_i, X_{ij}, Y_{ik})/n$ :** By the definition of $W_{z,jk}(Z_i, X_{ij}, Y_{ik})$ in (S99), we have

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n} W_{z,jk}^2(Z_i, X_{ij}, Y_{ik}) &= \frac{h}{n}\cdot\sum_{i\in[n]}\{K_h(Z_i - z)X_{ij}Y_{ik} - K_h(Z_i - z)\mathbf{\Sigma}_{jk}(z)\}^2 \\ &\leqslant h\cdot\max_{i\in[n]}\ \{K_h(Z_i - z)X_{ij}Y_{ik} - K_h(Z_i - z)\mathbf{\Sigma}_{jk}(z)\}^2 \\ &\leqslant 2h\cdot\max_{i\in[n]}\ \left\{K_h^2(Z_i - z)X_{ij}^2Y_{ik}^2 + K_h^2(Z_i - z)\mathbf{\Sigma}_{jk}^2(z)\right\} \\ &\leqslant 2h\cdot\left(\frac{1}{h^2}\cdot\|K\|_\infty^2\cdot M_X^4\cdot\log^2 d + \frac{1}{h^2}\cdot\|K\|_\infty^2\cdot M_\sigma^2\right) \\ &\leqslant C\cdot\frac{\log^2 d}{h},\end{aligned} \tag{S100}$$

with probability at least $1 - 1/d$. Note that the second inequality holds by the fact that $(x - y)^2 \leqslant 2x^2 + 2y^2$, and the third inequality holds by (12) and Assumption 2, and the fact that $\max(X_{ij}, Y_{ij}) \leqslant M_X\cdot\sqrt{\log d}$ with probability at least $1 - 1/d$.

**Covering number of the function class $\mathcal{W}$:** To obtain the covering number of the function class $\mathcal{W}$ under the $L_2$ norm on the empirical measure, it suffices to obtain the covering number $\sup_Q N\{\mathcal{W}, L_2(Q), \epsilon\}$. First, we note that $W_{z,jk} = \sqrt{h}\cdot\{g_{z,jk} - w_z\cdot\mathbf{\Sigma}_{jk}(z)\}$. From Lemma S16, we have $\mathcal{K}_1 = \{w_z(\cdot)\mid z\in[0,1]\}$ and that

$$\sup_Q \mathcal{N}\{\mathcal{K}_1, L_2(Q), \epsilon\} \leqslant \left(\frac{2\cdot C_K\cdot\|K\|_{\mathrm{TV}}}{h\epsilon}\right)^4.$$

Also, From Lemma S17, we have $\mathcal{G}_{1,jk} = \{g_{z,jk}(\cdot)\mid z\in[0,1]\}$ and that

$$\sup_Q N\{\mathcal{G}_{1,jk}, L_2(Q), \epsilon\} \leqslant \left(\frac{2\cdot M_X^2\cdot\log d\cdot C_K\cdot\|K\|_{\mathrm{TV}}}{h\epsilon}\right)^4.$$

Moreover, by Assumption 2, $\mathbf{\Sigma}_{jk}(z)$ is $M_\sigma$-Lipschitz. Thus, applying Lemmas S14 and S15, we obtain

$$\sup_Q N\{\mathcal{W}, L_2(Q), \epsilon\} \leqslant 2^{22}\cdot M_\sigma\cdot M_X^8\cdot C_K^8\cdot\|K\|_{\mathrm{TV}}^8\cdot\|K\|_\infty^5\cdot d^2\cdot\left(\frac{\log^{4/9} d}{h^{17/18}\epsilon}\right)^9 = C\cdot d^2\cdot\left(\frac{\log^{4/9} d}{h^{17/18}\epsilon}\right)^9, \tag{S101}$$

where the term $d^2$ appear on the right hand side because the function class $\mathcal{W}$ is over $j, k\in[d]$.

**Applying Dudley's inequality and Borell's inequality:** Applying Dudley's inequality (see Corollary 2.2.8 in Van Der Vaart and Wellner, 1996) with (S100) and (S101), we have

$$\mathbb{E}\left\{\sup_{z\in[0,1]}\max_{j,k\in[d]}\frac{1}{\sqrt{n}}\sum_{i\in[n]}W_{z,jk}(Z_i,X_{ij},Y_{ik})\cdot\xi_i\right\}\leqslant C\cdot\int_0^{C\cdot\sqrt{\frac{\log^2 d}{h}}}\sqrt{\log\left(\frac{d^{2/9}\cdot\log^{4/9}d}{h^{17/18}\epsilon}\right)}d\epsilon.$$

Applying (S19) with $b_1 = C\cdot\sqrt{\log^2 d/h}$ and $b_2 = d^{2/9}\cdot\log^{4/9} d/h^{17/18}$, we have

$$\mathbb{E}\left\{\sup_{z\in[0,1]}\max_{j,k\in[d]}\frac{1}{\sqrt{n}}\sum_{i\in[n]}W_{z,jk}(Z_i,X_{ij},Y_{ik})\cdot\xi_i\right\}\leqslant C\cdot\sqrt{\frac{\log^3(d/h)}{h}}, \tag{S102}$$

for some sufficiently large $C>0$.

By Borell's inequality (see Proposition A.2.1 in Van Der Vaart and Wellner, 1996), for $\lambda>0$, we have

$$P\left[\left|\sup_{z\in[0,1]}\max_{j,k\in E(z)}\frac{1}{\sqrt{n}}\sum_{i\in[n]}W_{z,jk}(Z_i,X_{ij},Y_{ik})\cdot\xi_i\right|\geqslant C\cdot\sqrt{\frac{\log^3(d/h)}{h}}+\lambda\;\middle|\;\{(Z_i,\boldsymbol{X}_i,\boldsymbol{Y}_i)\}_{i\in[n]}\right]$$
$$\leqslant 2\cdot\exp\left(-\frac{\lambda^2}{2\sigma_X^2}\right),$$

where $\sigma_X^2$ is the upper bound on the conditional variance. Picking $\lambda = C\cdot\sqrt{\frac{\log^3(d/h)}{h}}$, we have

$$P\left[\left|\sup_{z\in[0,1]}\max_{j,k\in[d]}\frac{1}{\sqrt{n}}\sum_{i\in[n]}W_{z,jk}(Z_i,X_{ij},Y_{ik})\cdot\xi_i\right|\geqslant C\cdot\sqrt{\frac{\log^3(d/h)}{h}}\;\middle|\;\{(Z_i,\boldsymbol{X}_i,\boldsymbol{Y}_i)\}_{i\in[n]}\right]\leqslant\frac{1}{d}. \tag{S103}$$

Thus, substituting (S103) into (S98), we have

$$\begin{aligned}I_1&\leqslant C\cdot M\cdot s\cdot\left\{h^2+\sqrt{\frac{\log(d/h)}{nh}}\right\}\cdot\sqrt{\frac{\log^3(d/h)}{h}}\\&\leqslant C\cdot s\cdot\sqrt{h^3\log^3(d/h)}+C\cdot s\cdot\sqrt{\frac{\log^4(d/h)}{nh^2}},\end{aligned}\tag{S104}$$

with probability $1-1/d$.

**Upper bound for $I_2$:** By an application of Holder's inequality, we have

$$
\begin{aligned}
I_2 &\leqslant \sqrt{nh} \cdot \sup_{z\in[0,1]} \max_{j,k\in[d]} \left\|\widehat{\boldsymbol{\Theta}}_j(z) - \boldsymbol{\Theta}_j(z)\right\|_1 \cdot \left\|\widehat{\boldsymbol{\Theta}}_k(z)\right\|_1 \cdot \left| \sup_{z\in[0,1]} \max_{j,k\in[d]} \frac{1}{n}\sum_{i\in[n]} \{K_h(Z_i - z)X_{ij}Y_{ik}\}\xi_i \right| \\
&\leqslant \sup_{z\in[0,1]} \max_{j,k\in[d]} \left[\left\|\widehat{\boldsymbol{\Theta}}_k(z) - \boldsymbol{\Theta}_k(z)\right\|_1 + \|\boldsymbol{\Theta}_k(z)\|_1\right] \cdot C \cdot s \cdot \left\{h^2 + \sqrt{\frac{\log(d/h)}{nh}}\right\} \\
&\quad \times \sqrt{nh} \cdot \left| \sup_{z\in[0,1]} \max_{j,k\in[d]} \frac{1}{n}\sum_{i\in[n]} \{K_h(Z_i - z)X_{ij}Y_{ik}\}\xi_i \right| \\
&\leqslant C \cdot M \cdot s \cdot \left\{h^2 + \sqrt{\frac{\log(d/h)}{nh}}\right\} \cdot \sqrt{nh} \cdot \left| \sup_{z\in[0,1]} \max_{j,k\in[d]} \frac{1}{n}\sum_{i\in[n]} \{K_h(Z_i - z)X_{ij}Y_{ik}\}\xi_i \right|,
\end{aligned}
\tag{S105}
$$

where the second inequality holds by triangle inequality and Corollary 1, and the last inequality holds by another application of Corollary 1 and the assumption that $h^2+\sqrt{\log(d/h)/(nh)} = o(1)$.

Recall the definition of $g_{z,jk}(Z_i, X_{ij}, Y_{ik}) = K_h(Z_i - z)X_{ij}Y_{ik}$. Conditioned on the data $\{(Z_i, \boldsymbol{X}_i, \boldsymbol{Y}_i)\}_{i\in[n]}$, we note that

$$
\sqrt{\frac{h}{n}}\sum_{i\in[n]} \{K_h(Z_i - z)X_{ij}Y_{ik}\}\xi_i = \frac{1}{\sqrt{n}}\sum_{i\in[n]} \sqrt{h}\cdot g_{z,jk}(Z_i, X_{ij}, Y_{ik})\cdot\xi_i \sim N\left\{0, \frac{h}{n}\sum_{i\in[n]} g_{z,jk}^2(Z_i, X_{ij}, Y_{ik})\right\}.
$$

Similar to the upper bound for $I_1$, we apply Dudley's inequality and Borell's inequality to bound the supreme of the Gaussian process in the last expression.

To this end, we need to obtain an upper bound for the conditional covariance. By (S54), we have

$$
\frac{h}{n}\sum_{i\in[n]} g_{z,jk}^2(Z_i, X_{ij}, Y_{ik}) \leqslant \frac{1}{h}\cdot M_X^4 \cdot \|K\|_\infty^4 \cdot \log^2 d,
\tag{S106}
$$

with probability at least $1 - 1/d$. In addition, by an application of Lemma S17, the covering number for the class of function $\{\sqrt{h}\cdot g_{z,jk}(\cdot) \mid z\in[0,1],\ j,k\in[d]\}$ is

$$
\sup_Q N\left[\left\{\sqrt{h}\cdot g_{z,jk}(\cdot) \mid z\in[0,1],\ j,k\in[d]\right\}, L_2(Q), \epsilon\right] \leqslant d^2\cdot\left(\frac{2\cdot M_X^2\cdot\log d\cdot C_K\cdot\|K\|_{\mathrm{TV}}}{\sqrt{h}\epsilon}\right)^4.
\tag{S107}
$$

By an application of Dudley's inequality, we have

$$
\mathbb{E}\left\{\sup_{z\in[0,1]} \max_{j,k\in[d]} \frac{1}{\sqrt{n}}\sum_{i\in[n]} \sqrt{h}\cdot g_{z,jk}(Z_i, X_{ij}, Y_{ik})\cdot\xi_i\right\} \leqslant C\cdot\int_0^{\sqrt{M_X^4\cdot\|K\|_\infty^4\cdot\frac{\log^2 d}{h}}} \sqrt{\log\left(\frac{d^{1/2}\cdot\log d}{h^{1/2}\epsilon}\right)}d\epsilon.
$$

Applying (S19) with $b_1 = \sqrt{M_X^4 \cdot \|K\|_\infty^4 \cdot \log^2 d/h}$ and $b_2 = d^{1/2} \cdot \log d/h^{1/2}$, we have

$$\mathbb{E}\left\{\sup_{z\in[0,1]} \max_{j,k\in[d]} \frac{1}{\sqrt{n}} \sum_{i\in[n]} \sqrt{h} \cdot g_{z,jk}(Z_i, X_{ij}, Y_{ik}) \cdot \xi_i\right\} \leqslant C \cdot \sqrt{\frac{\log^3(d/h)}{h}}. \tag{S108}$$

By Borell's inequality (see Proposition A.2.1 in Van Der Vaart and Wellner, 1996), we have

$$\begin{aligned}&P\left[\left|\sup_{z\in[0,1]} \max_{j,k\in[d]} \frac{1}{\sqrt{n}} \sum_{i\in[n]} \sqrt{h} \cdot g_{z,jk}(Z_i, X_{ij}, Y_{ik}) \cdot \xi_i\right| \geqslant C \cdot \sqrt{\frac{\log^3(d/h)}{h}} + \lambda \;\middle|\; \{(Z_i, \boldsymbol{X}_i, \boldsymbol{Y}_i)\}_{i\in[n]}\right]\\ &\leqslant 2 \cdot \exp\left(-\frac{\lambda^2}{2\sigma_X^2}\right).\end{aligned}$$

Picking $\lambda = C \cdot \sqrt{\frac{\log^3(d/h)}{h}}$, we have

$$P\left[\left|\sup_{z\in[0,1]} \max_{j,k\in[d]} \frac{1}{\sqrt{n}} \sum_{i\in[n]} \sqrt{h} \cdot g_{z,jk}(Z_i, X_{ij}, Y_{ik}) \cdot \xi_i\right| \geqslant C \cdot \sqrt{\frac{\log^3(d/h)}{h}} \;\middle|\; \{(Z_i, \boldsymbol{X}_i, \boldsymbol{Y}_i)\}_{i\in[n]}\right] \leqslant \frac{1}{d}. \tag{S109}$$

Thus, by (S105) and (S109), we have

$$\begin{aligned}I_2 &\leqslant C \cdot M \cdot s \cdot \left\{h^2 + \sqrt{\frac{\log(d/h)}{nh}}\right\} \cdot \sqrt{\frac{\log^3(d/h)}{h}}\\ &\leqslant C \cdot s \cdot \sqrt{h^3 \log^3(d/h)} + C \cdot s \cdot \sqrt{\frac{\log^4(d/h)}{nh^2}},\end{aligned} \tag{S110}$$

with probability at least $1 - 1/d$.

**Upper bound for $I_3$:** By an application of Holder's inequality, we have

$$\begin{aligned}I_3 &\leqslant \sup_{z\in[0,1]} \max_{j\in[d]} \left\|\left\{\widehat{\boldsymbol{\Theta}}_j(z)\right\}^T \widehat{\boldsymbol{\Sigma}}(z) - \mathbf{e}_j\right\|_\infty \left\|\widehat{\boldsymbol{\Theta}}_k(z)\right\|_1^2 \cdot \sqrt{nh} \cdot \left|\sup_{z\in[0,1]} \max_{j,k\in[d]} \frac{1}{n} \sum_{i\in[n]} \{K_h(Z_i - z)X_{ij}Y_{ik} - K_h(Z_i - z)\boldsymbol{\Sigma}_{jk}(z)\}\, \xi_i\right|\\ &\leqslant M^3 \cdot C \cdot s \cdot \left\{h^2 + \sqrt{\frac{\log(d/h)}{nh}}\right\} \sqrt{nh} \cdot \left|\sup_{z\in[0,1]} \max_{j,k\in[d]} \frac{1}{n} \sum_{i\in[n]} \{K_h(Z_i - z)X_{ij}Y_{ik} - K_h(Z_i - z)\boldsymbol{\Sigma}_{jk}(z)\}\, \xi_i\right|\\ &\leqslant C \cdot M^3 \cdot s \cdot \left\{h^2 + \sqrt{\frac{\log(d/h)}{nh}}\right\} \cdot \sqrt{\frac{\log^3(d/h)}{h}}\\ &\leqslant C \cdot s \cdot \sqrt{h^3 \log^3(d/h)} + C \cdot s \cdot \sqrt{\frac{\log^4(d/h)}{nh^2}},\end{aligned} \tag{S111}$$

where the second inequality holds by the fact that $\boldsymbol{\Theta}(z) \in \mathcal{U}_{s,M}$ and by an application of Corollary 1, and the third inequality holds by (S103).

Thus, combining (S104), (S110), and (S111), we have

$$|T^B - T_0^B| \leqslant C \cdot s \cdot \sqrt{h^3 \log^3(d/h)} + C \cdot s \cdot \sqrt{\frac{\log^4(d/h)}{nh^2}} \tag{S112}$$

with probability at least $1 - 3/d$.

**Upper bound for $|T_0^B - T_{00}^B|$:** Recall from (S63) that

$$T_{00}^B = \sup_{z\in[0,1]} \max_{(j,k)\in E(z)} \sqrt{nh} \cdot \Bigg| \sum_{i\in[n]} \Big( \{\boldsymbol{\Theta}_j(z)\}^T K_h(Z_i - z) \left\{\boldsymbol{X}_i \boldsymbol{Y}_i^T - \boldsymbol{\Sigma}(z)\right\} \boldsymbol{\Theta}_k(z)/n \\ - \{\boldsymbol{\Theta}_j(z)\}^T \Big[\mathbb{E}\{K_h(Z-z)\boldsymbol{X}\boldsymbol{Y}^T\} - \mathbb{E}\{K_h(Z-z)\}\boldsymbol{\Sigma}(z)\Big] \boldsymbol{\Theta}_k(z)\Big) \cdot \xi_i/n\Bigg|.$$

By the triangle inequality, we have

$$\begin{aligned} |T_0^B - T_{00}^B| &\leqslant \sqrt{nh} \cdot \sup_{z\in[0,1]} \max_{(j,k)\in E(z)} \left| \frac{1}{n} \sum_{i\in[n]} \Big( \{\boldsymbol{\Theta}_j(z)\}^T \Big[\mathbb{E}\left\{K_h(Z_i - z)\boldsymbol{X}_i \boldsymbol{Y}_i^T\right\} - \mathbb{E}\{K_h(Z_i - z)\}\boldsymbol{\Sigma}(z)\Big] \boldsymbol{\Theta}_k(z)\Big) \cdot \xi_i \right| \\ &\leqslant \sqrt{nh} \cdot \sup_{z\in[0,1]} \max_{(j,k)\in E(z)} \left| \{\boldsymbol{\Theta}_j(z)\}^T \Big[\mathbb{E}\left\{K_h(Z-z)\boldsymbol{X}\boldsymbol{Y}^T\right\} - \mathbb{E}\{K_h(Z-z)\}\boldsymbol{\Sigma}(z)\Big] \boldsymbol{\Theta}_k(z)\right| \cdot \left| \frac{1}{n} \sum_{i\in[n]} \xi_i \right| \\ &\leqslant \sqrt{nh} \cdot M^2 \cdot C \cdot h^2 \cdot \left| \frac{1}{n} \sum_{i\in[n]} \xi_i \right|, \end{aligned} \tag{S113}$$

where the last inequality holds by applying Holder's inequality and Lemma S2. Since $\xi_i \overset{\text{i.i.d.}}{\sim} N(0,1)$, by the Gaussian tail inequality, we have

$$P\left( \left| \frac{1}{n} \sum_{i\in[n]} \xi_i \right| > \sqrt{\frac{2\log n}{n}} \right) \leqslant \frac{1}{n}.$$

Thus, substituting the above expression into (S113), we obtain

$$|T_0^B - T_{00}^B| \leqslant \sqrt{nh} \cdot M^2 \cdot C \cdot h^2 \cdot \sqrt{\frac{2\log n}{n}} \leqslant C \cdot \sqrt{h^5 \log n}, \tag{S114}$$

with probability at least $1 - 1/n$.

**Combining the upper bounds:** Combining the upper bounds (S112) and (S114), and applying the union bound, we have

$$
\begin{aligned}
&P\left[\left|T^B - T^B_{00}\right| \geqslant C\cdot s\cdot\sqrt{h^3\log^3(d/h)} + C\cdot s\cdot\sqrt{\frac{\log^4(d/h)}{nh^2}} + C\cdot\sqrt{h^5\log n}\;\middle|\;\{(Z_i,\boldsymbol{X}_i,\boldsymbol{Y}_i)\}_{i\in[n]}\right]\\
&\leqslant P\left[\left|T^B - T^B_{0}\right| + \left|T^B_0 - T^B_{00}\right| \geqslant C\cdot s\cdot\sqrt{h^3\log^3(d/h)} + C\cdot s\cdot\sqrt{\frac{\log^4(d/h)}{nh^2}} + C\cdot\sqrt{h^5\log n}\;\middle|\;\{(Z_i,\boldsymbol{X}_i,\boldsymbol{Y}_i)\}_{i\in[n]}\right]\\
&\leqslant P\left[\left|T^B - T^B_{0}\right| \geqslant C\cdot s\cdot\sqrt{h^3\log^3(d/h)} + C\cdot s\cdot\sqrt{\frac{\log^4(d/h)}{nh^2}}\;\middle|\;\{(Z_i,\boldsymbol{X}_i,\boldsymbol{Y}_i)\}_{i\in[n]}\right]\\
&\quad + P\left[\left|T^B_0 - T^B_{00}\right| \geqslant C\cdot\sqrt{h^5\log n}\;\middle|\;\{(Z_i,\boldsymbol{X}_i,\boldsymbol{Y}_i)\}_{i\in[n]}\right]\\
&\leqslant 2/d + 1/n,
\end{aligned}
$$

as desired.

Web Appendix F.6 *Lower Bound of the Variance*

We aim to show that the variance of $J_{z,jk}$ defined in (S15) is bounded from below.

LEMMA S10: *Under the same conditions of Theorem 2, there exists a constant* $c > 0$ *such that* $\inf_z \min_{j,k} \mathrm{Var}(J_{z,jk}) \geqslant c > 0$.

*Proof.* In this proof, we will apply Isserlis' theorem (Isserlis, 1918). Given $\mathbf{T} \sim N(0, \boldsymbol{\Sigma})$, Isserlis' theorem implies that for any vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^d$,

$$
\begin{aligned}
\mathbb{E}\{(\mathbf{u}^T\mathbf{T}\mathbf{T}^T\mathbf{v})^2\} &= \mathbb{E}\{(\mathbf{u}^T\mathbf{T})^2\}\mathbb{E}\{(\mathbf{v}^T\mathbf{T})^2\} + 2\{\mathbb{E}(\mathbf{u}^T\mathbf{T}\mathbf{v}^T\mathbf{T})\}^2\\
&= (\mathbf{u}^T\boldsymbol{\Sigma}\mathbf{u})(\mathbf{v}^T\boldsymbol{\Sigma}\mathbf{v}) + 2(\mathbf{u}^T\boldsymbol{\Sigma}\mathbf{v})^2 \qquad \text{(S115)}
\end{aligned}
$$

According to the definition of $J_{z,jk}$ in (S15), it can be decomposed into $J_{z,jk}(Z_i, \boldsymbol{X}_i, \boldsymbol{Y}_i) = J^{(1)}_{z,jk}(Z, \boldsymbol{X}_i, \boldsymbol{Y}_i) - J^{(2)}_{z,jk}(Z_i)$. Recall that

$$
J^{(1)}_{z,jk}(Z_i, \boldsymbol{X}_i, \boldsymbol{Y}_i) = \sqrt{h}\cdot\{\boldsymbol{\Theta}_j(z)\}^T\cdot\left[K_h(Z_i - z)\boldsymbol{X}_i\boldsymbol{Y}_i^T - \mathbb{E}\left\{K_h(Z-z)\boldsymbol{X}\boldsymbol{Y}^T\right\}\right]\cdot\boldsymbol{\Theta}_k(z),
$$

and

$$
J^{(2)}_{z,jk}(Z_i) = \sqrt{h}\cdot\{\boldsymbol{\Theta}_j(z)\}^T\cdot\left[K_h(Z_i - z) - \mathbb{E}\left\{K_h(Z-z)\right\}\right]\cdot\boldsymbol{\Sigma}(z)\cdot\boldsymbol{\Theta}_k(z).
$$

We will calculate $\mathrm{Var}\{J^{(2)}_{z,jk}(Z)\}$, $\mathrm{Var}\{J^{(1)}_{z,jk}(Z, \boldsymbol{X}, \boldsymbol{Y})\}$, and $\mathrm{Cov}\{J^{(1)}_{z,jk}(Z, \boldsymbol{X}, \boldsymbol{Y}), J^{(2)}_{z,jk}(Z)\}$ separately.

We first calculate $\mathrm{Var}\{J^{(2)}_{z,jk}(Z)\}$. Following a similar method as the proof of Lemma S2, we have $\mathbb{E}\{K_h(Z-z)\} = f_Z(z) + \mathcal{O}(h^2)$ and $\mathbb{E}\{K_h^2(Z-z)\} = h^{-1} f_Z(z) \int K^2(u)du + \mathcal{O}(1)$. This implies that

$$\mathrm{Var}\{J_{z,jk}^{(2)}(Z)\} = \mathbf{\Theta}_{jk}^2(z) \cdot f_Z(z) \int K^2(u)du + \mathcal{O}(h). \tag{S116}$$

Next, we proceed to calculate the variance of $J_{z,jk}^{(1)}(Z)$. By a change of variable and Taylor's expansion, we obtain

$$\begin{aligned}
&\mathbf{\Theta}_j(z)^T \mathbb{E}\big\{K_h(Z-z)\mathbf{\Sigma}(Z)\big\}\mathbf{\Theta}_k(z) \\
&= \mathbf{\Theta}_j(z)^T \left\{\int K(u)\mathbf{\Sigma}(z+uh) f_Z(z+uh)du\right\} \mathbf{\Theta}_k(z) \\
&= \mathbf{\Theta}_j(z)^T \left[\int K(u)\{\mathbf{\Sigma}(z) + uh\dot{\mathbf{\Sigma}}(z) + u^2h^2\ddot{\mathbf{\Sigma}}(z')\}\{f_Z(z) + uh\dot{f}_Z(z) + u^2h^2\ddot{f}_Z(z)\}du\right] \mathbf{\Theta}_k(z).
\end{aligned} \tag{S117}$$

Note that each term in the integrant that involves $\int uK(u)du$ is equal to zero since $\int uK(u)du = 0$ by assumption. For terms with $\mathbf{\Sigma}(z)$, we have

$$\begin{aligned}
&\mathbf{\Theta}_j(z)^T\mathbf{\Sigma}(z)\mathbf{\Theta}_k(z) \int K(u)\{f_Z(z) + uh\dot{f}_Z(z) + u^2h^2\ddot{f}_Z(z)\}du \\
&= \mathbf{\Theta}_{jk}(z)\{f_Z(z) + \mathcal{O}(h^2)\}.
\end{aligned}$$

For terms that involve $\dot{\mathbf{\Sigma}}(z)$ and $\ddot{\mathbf{\Sigma}}(z')$, we have

$$\mathbf{\Theta}_j(z)^T\dot{\mathbf{\Sigma}}(z)\mathbf{\Theta}_k(z) \leqslant M_\sigma \|\mathbf{\Theta}_j(z)\|_2 \|\mathbf{\Theta}_k(z)\|_2 \leqslant \rho^2 M_\sigma = \mathcal{O}(1),$$

since the maximum eigenvalue of $\mathbf{\Theta}(z)$ is bounded by $\rho$ by assumption. Thus, combining the above into (S117), we have

$$\mathbf{\Theta}_j(z)^T \mathbb{E}\big\{K_h(Z-z)\mathbf{\Sigma}(Z)\big\}\mathbf{\Theta}_k(z) = \mathbf{\Theta}_{jk}(z) f_Z(z) + \mathcal{O}(h^2). \tag{S118}$$

Next, we bound the second moment. By the Isserlis' theorem in (S115), and by taking the conditional expectation, we have

$$\begin{aligned}
&\mathbb{E}\big[K_h^2(Z-z)\{\mathbf{\Theta}_j(z)^T \boldsymbol{X}\boldsymbol{Y}^T \mathbf{\Theta}_k(z)\}^2\big] \\
&= \mathbb{E}\big(K_h^2(Z-z)[\{\mathbf{\Theta}_j(z)^T\mathbf{\Sigma}(Z)\mathbf{\Theta}_j(z)\}\{\mathbf{\Theta}_k(z)^T\mathbf{\Sigma}(Z)\mathbf{\Theta}_k(z)\} + 2\{\mathbf{\Theta}_j(z)^T\mathbf{\Sigma}(Z)\mathbf{\Theta}_k(z)\}^2]\big).
\end{aligned} \tag{S119}$$

Following a similar argument as in (S118), we can derive

$$\mathbb{E}\big[K_h^2(Z-z)\{\boldsymbol{\Theta}_j(z)^T\boldsymbol{X}\boldsymbol{Y}^T\boldsymbol{\Theta}_k(z)\}^2\big] = \{\boldsymbol{\Theta}_{jj}(z)\boldsymbol{\Theta}_{kk}(z)+2\boldsymbol{\Theta}_{jk}^2(z)\}f_Z(z)h^{-1}\int K^2(u)du+\mathcal{O}(1) \tag{S120}$$

Thus, we have

$$\operatorname{Var}\left\{J_{z,jk}^{(1)}(Z)\right\} = \{\boldsymbol{\Theta}_{jj}(z)\boldsymbol{\Theta}_{kk}(z)+2\boldsymbol{\Theta}_{jk}^2(z)\}f_Z(z)\int K^2(u)du+\mathcal{O}(h). \tag{S121}$$

Now we begin to bound the $\operatorname{Cov}\{J_{z,jk}^{(1)}(Z), J_{z,jk}^{(2)}(Z)\}$. By using a similar argument as (S118), we have

$$\mathbb{E}\big[\boldsymbol{\Theta}_{jk}(z)K_h^2(Z-z)\{\boldsymbol{\Theta}_j(z)^T\boldsymbol{X}\boldsymbol{Y}^T\boldsymbol{\Theta}_k(z)\}\big] = \boldsymbol{\Theta}_{jk}^2(z)\cdot h^{-1}f_Z(z)\int K^2(u)du+\mathcal{O}(1), \tag{S122}$$

Combining with (S122) and (S118), and using the covariance formula, we have that

$$\operatorname{Cov}\left\{J_{z,jk}^{(1)}(Z), J_{z,jk}^{(2)}(Z)\right\} = \boldsymbol{\Theta}_{jk}^2(z)f_Z(z)\int K^2(u)du+\mathcal{O}(h). \tag{S123}$$

Using (S116), (S121) and (S123), we have

$$\begin{aligned}\operatorname{Var}\{J_{z,jk}(Z)\} &= \operatorname{Var}(J_{z,jk}^{(1)}(Z)) + \operatorname{Var}\{J_{z,jk}^{(2)}(Z)\} - 2\operatorname{Cov}\left\{J_{z,jk}^{(1)}(Z), J_{z,jk}^{(2)}(Z)\right\} \\ &= \{\boldsymbol{\Theta}_{jj}(z)\boldsymbol{\Theta}_{kk}(z)+\boldsymbol{\Theta}_{jk}^2(z)\}f_Z(z)\int K^2(u)du+\mathcal{O}(h) \geqslant \rho^2\underline{f}_Z,\end{aligned}$$

where the last inequality is because $\rho$ is smaller than the minimum eigenvalue of $\boldsymbol{\Sigma}(z)$ for any $z\in[0,1]$ and $\inf_{z\in[0,1]}\ f_Z(z)\geqslant \underline{f}_Z>0$ by Assumption 1. Since the lower bound above is uniformly true over $z, j, k$, the lemma is proven.

## Web Appendix G. Proof of Theorem S1

In this section, we show that the proposed procedure in Algorithm 1 is able to control the type I error below a pre-specified level $\alpha$. We first define some notation that will be used throughout the proof of Theorem 3. Let $E^*(z)$ be the true edge set at $Z=z$. That is, $E^*(z)$ is the set of edges induced by the true inverse covariance matrix $\boldsymbol{\Theta}(z)$. Recall from

Definition S4 that the critical edge set is defined as

$$\mathcal{C}\{E(z),\mathcal{P}\} = \{e \mid e \notin E(z), \text{ there exists } E'(z) \supseteq E(z) \text{ such that } E'(z) \in \mathscr{P} \text{ and } E'(z)\backslash\{e\} \notin \mathscr{P}\}, \tag{S124}$$

where $\mathscr{P} = \{E \subseteq V \times V \mid \mathcal{P}(G) = 1\}$ is the class of edge sets satisfying the graph property $\mathcal{P}$.

Suppose that Algorithm 1 rejects the null hypothesis at the $T$th iteration. That is, there exists $z_0 \in [0,1]$ such that $E_T(z_0) \in \mathscr{P}$ but $E_{T-1}(z_0) \notin \mathscr{P}$. To prove Theorem 3, we state the following two lemmas on the properties of critical edge set.

LEMMA S11: *Let $E_T(z_0) \in \mathscr{P}$ for some $z_0 \in [0,1]$. Then, at least one rejected edge in $E_T(z_0)$ is in the critical edge set $\mathcal{C}\{E^*(z_0),\mathcal{P}\}$.*

LEMMA S12: *Let $\bar{e} \in \mathcal{C}\{E^*(z_0),\mathcal{P}\}$ be the first rejected edge in the critical edge set $\mathcal{C}\{E^*(z_0),\mathcal{P}\}$. Suppose that $\bar{e}$ is rejected at the $l$th step of Algorithm 1. Then, $\mathcal{C}\{E^*(z),\mathcal{P}\} \subseteq \mathcal{C}\{E_{l-1}(z),\mathcal{P}\}$ for all $z \in [0,1]$.*

The proofs of Lemmas S11 and S12 are deferred to Sections Web Appendix G.2 and Web Appendix G.3, respectively. We now provide the proof of Theorem 3.

## Web Appendix G.1 *Proof of Theorem 3*

Suppose that Algorithm 1 rejects the null hypothesis at the $T$th iteration. That is, $E_T(z_0) \in \mathscr{P}$ and $E_{T-1}(z_0) \notin \mathscr{P}$. By Lemma S11, there is at least one edge in $E_T(z_0)$ that is also in the critical edge set $\mathcal{C}\{E^*(z_0),\mathcal{P}\}$. We denote the first rejected edge in the critical edge set as $\bar{e}$, i.e., $\bar{e} \in \mathcal{C}\{E^*(z_0),\mathcal{P}\}$ and suppose that $\bar{e}$ is rejected at the $l$th iteration of Algorithm 1.

We note that $l$ is not necessarily $T$. Thus, we have

$$\begin{aligned}\sup_{z\in[0,1]}\max_{e\in\mathcal{C}\{E^*(z),\mathcal{P}\}}\sqrt{nh}\cdot\widehat{\mathbf{\Theta}}^{\mathrm{de}}_{e}(z)\cdot\sum_{i\in[n]}K_h(Z_i-z)/n &\geqslant \sqrt{nh}\cdot\widehat{\mathbf{\Theta}}^{\mathrm{de}}_{\bar{e}}(z_0)\cdot\sum_{i\in[n]}K_h(Z_i-z_0)/n\\ &\geqslant c\{1-\alpha,\mathcal{C}(E_{l-1},\mathcal{P})\}\\ &\geqslant c\{1-\alpha,\mathcal{C}(E^*,\mathcal{P})\},\end{aligned}$$

where the first inequality follows by Lemma S11, the second inequality follows from the $l$th step of Algorithm 1, and the last inequality follows directly from Lemma S12.

Under the null hypothesis, $\mathbf{\Theta}_e(z)=0$ for any $e\in\mathcal{C}\{E^*(z),\mathcal{P}\}$. By Theorem 2, we have

$$\begin{aligned}&\lim_{n\to\infty}\sup_{\mathbf{\Theta}(\cdot)\in\mathcal{G}_0}P_{\mathbf{\Theta}(\cdot)}(\psi_\alpha=1)\\ &\leqslant \lim_{n\to\infty}\sup_{\mathbf{\Theta}(\cdot)\in\mathcal{G}_0}P\left[\sup_{z\in[0,1]}\max_{e\in\mathcal{C}\{E^*(z),\mathcal{P}\}}\sqrt{nh}\cdot|\widehat{\mathbf{\Theta}}^{\mathrm{de}}_{jk}(z)|\cdot\sum_{i\in[n]}K_h(Z_i-z)/n\geqslant c\{1-\alpha,\mathcal{C}(E^*,\mathcal{P})\}\right]\\ &\leqslant \alpha,\end{aligned}$$

as desired.

## Web Appendix G.2 *Proof of Lemma S11*

To prove Lemma S11, it suffices to show that the intersection between the two sets $E_T(z_0)$ and $\mathcal{C}\{E^*(z_0),\mathcal{P}\}$ is not an empty set, i.e., $E_T(z_0)\cap\mathcal{C}\{E^*(z_0),\mathcal{P}\}\neq\emptyset$. To this end, we let $F=E_T(z_0)\cup E^*(z_0)$ and let $E_T(z_0)\setminus E^*(z_0)=\{e_1,e_2,\ldots,e_k\}$. We note that the set $E_T(z_0)\setminus E^*(z_0)$ is not an empty set since $E_T(z_0)\in\mathscr{P}$ but $E^*(z_0)\notin\mathscr{P}$.

Using the fact that $\mathcal{P}$ is monotone and that $E_T(z_0)\in\mathscr{P}$, we have $F\in\mathscr{P}$ since adding additional edges to $E_T(z_0)$ does not change the graph property of $E_T(z_0)$. Then, we have

$$E^*(z_0)\subseteq E^*(z_0)\cup\{e_1\}\subseteq E^*(z_0)\cup\{e_1,e_2\}\subseteq\cdots\subseteq E^*(z_0)\cup\{e_1,\ldots,e_k\}=F.$$

Since $E^*(z_0)\notin\mathscr{P}$ and $F\in\mathscr{P}$, there must exists an edge set $\{e_1,\ldots,e_{k_0}\}$ for $k_0\leqslant k$ that changes the graph property of $E^*(z_0)$ from $E^*(z_0)\notin\mathscr{P}$ to $E^*(z_0)\cup\{e_1,\ldots,e_{k_0}\}\in\mathscr{P}$.

Thus, there must exists at least an edge $\bar{e}\in\{e_1,\ldots,e_{k_0}\}$ such that $\bar{e}\in\mathcal{C}\{E^*(z_0),\mathcal{P}\}$ since adding the set of edges $\{e_1, \ldots, e_{k_0}\}$ changes the graph property of $E^*(z_0)$. Also, $\bar{e} \in E_T(z_0)$ by construction. Thus, we conclude that $E_T(z_0) \cap \mathcal{C}\{E^*(z_0), \mathcal{P}\} \neq \emptyset$.

Web Appendix G.3 *Proof of Lemma S12*

Let $\bar{e} \in \mathcal{C}\{E^*(z_0), \mathcal{P}\}$ be the first rejected edge in the critical edge set $\mathcal{C}\{E^*(z_0), \mathcal{P}\}$ for some $z_0 \in [0, 1]$. Suppose that $\bar{e}$ is rejected at the $l$th step of Algorithm 1. We want to show that $\mathcal{C}\{E^*(z), \mathcal{P}\} \subseteq \mathcal{C}\{E_{l-1}(z), \mathcal{P}\}$ for all $z \in [0, 1]$. It suffices to show that $\mathcal{C}\{E^*(z_0), \mathcal{P}\} \subseteq \mathcal{C}\{E_{l-1}(z_0), \mathcal{P}\}$. In other words, we want to prove that for any $e' \in \mathcal{C}\{E^*(z_0), \mathcal{P}\}$, $e' \in \mathcal{C}\{E_{l-1}(z_0), \mathcal{P}\}$. We first note the following fact

$$E_{l-1}(z_0) \cap \mathcal{C}\{E^*(z_0), \mathcal{P}\} = \emptyset \qquad \text{and} \qquad E_{l-1}(z_0) \notin \mathscr{P}. \tag{S125}$$

By the definition of the critical edge set (S124), we construct a set $E'$ such that $E^*(z_0) \supseteq E'$, $E' \in \mathscr{P}$, and $E' \setminus \{e'\} \notin \mathscr{P}$, for any $e' \in \mathcal{C}\{E^*(z_0), \mathcal{P}\}$. By the definition of monotone property, we have $E' \cup E_{l-1}(z_0) \in \mathscr{P}$. Since $\mathcal{C}\{E' \cup E_{l-1}(z_0), \mathcal{P}\} \subseteq \mathcal{C}\{E_{l-1}(z_0), \mathcal{P}\}$, to show that $e' \in \mathcal{C}\{E_{l-1}(z_0), \mathcal{P}\}$, it is equivalent to showing $e' \in \mathcal{C}\{E' \cup E_{l-1}(z_0), \mathcal{P}\}$. That is, we want to show

$$\{E' \cup E_{l-1}(z_0)\} \setminus \{e'\} \notin \mathscr{P}.$$

This is equivalent to showing

$$\{E' \setminus e'\} \cup \{E_{l-1}(z_0) \setminus E'\} \notin \mathscr{P}. \tag{S126}$$

There are two cases: (1) $E_{l-1}(z_0) \setminus E' = \emptyset$ and (2) $E_{l-1}(z_0) \setminus E' \neq \emptyset$. For the first case, (S126) is true by the construction of $E'$. For the second case, we prove by contradiction.

Suppose that $(E' \setminus e') \cup \{E_{l-1}(z_0) \setminus E'\} \in \mathscr{P}$. Let $E_{l-1}(z_0) \setminus E' = \{e'_1, \ldots, '_k\}$. By the definition of monotone property, we have

$$E' \setminus \{e'\} \subseteq (E' \setminus \{e'\}) \cup \{e'_1\} \subseteq \cdots \subseteq (E' \setminus \{e'\}) \cup \{e'_1, e'_2, \ldots, e'_k\} = (E' \setminus \{e'\}) \cup (E_{l-1}(z_0) \setminus E').$$

Since $E' \setminus \{e'\} \notin \mathscr{P}$ by construction, and that $(E' \setminus \{e'\}) \cup (E_{l-1}(z_0) \setminus E') \in \mathscr{P}$, there must exists an edge set $\{e_1, \ldots, e_{k_0}\}$ for $k_0 \leqslant k$ that changes the graph property of $E' \setminus \{e'\} \notin \mathscr{P}$ to $(E' \setminus \{e'\}) \cup \{e'_1, \ldots, e'_{k_0}\} \in \mathscr{P}$.

Since $e'_{k_0} \in E_{l-1}(z_0) \setminus E'$ and that $E^*(z_0) \subseteq E'$ by construction, we have $e'_{k_0} \notin E^*(z_0)$. Thus, $e'_{k_0} \in \mathcal{C}\{E^*(z_0), \mathcal{P}\}$. This contradicts the fact that

$$E_{l-1}(z_0) \cap \mathcal{C}\{E^*(z_0), \mathcal{P}\} = \emptyset.$$

## Web Appendix H. Proof of Theorem S2

By the definition in (S5), if $\mathbf{\Theta}(\cdot) \in \mathcal{G}_1(\theta; \mathcal{P})$, there exists an edge set $E'_0$ and $z_0 \in [0, 1]$ satisfying

$$E'_0 \subseteq E\{\mathbf{\Theta}(z_0)\}, \mathcal{P}(E'_0) = 1 \text{ and } \min_{e \in E'_0} |\mathbf{\Theta}_e(z_0)| > C\sqrt{\log(d/h)/nh}, \tag{S127}$$

and we will determine the magnitude tf constant $C$ later. We aim to show that $\mathcal{P}\{E'_0 \cap \mathcal{C}(\emptyset, \mathcal{P})\} = \mathcal{P}(E'_0) = 1$. First, there exists a subgraph $E''_0 \subset E'_0$ such that $\mathcal{P}(E''_0) = \mathcal{P}(E'_0) = 1$ and for any $\widetilde{E} \subset E''_0$, $\mathcal{P}(\widetilde{E}) = 0$. We can construct such $E''_0$ by deleting edges from $E'_0$ until it is impossible to further deleting any edge such that the property $\mathcal{P}$ is still true. By Definition S4, $E''_0 \subseteq \mathcal{C}(\emptyset, \mathcal{P})$ and therefore $E''_0 \subseteq E'_0 \cap \mathcal{C}(\emptyset, \mathcal{P})$. By monotone property, we have $\mathcal{P}\{E'_0 \cap \mathcal{C}(\emptyset, \mathcal{P})\} = \mathcal{P}(E'_0) = 1$ since $\mathcal{P}(E''_0) = \mathcal{P}(E'_0) = 1$. Consider the following event

$$\mathcal{E}_1 = \Big[\min_{e \in E'_0 \cap \mathcal{C}(\emptyset, \mathcal{P})} \sqrt{nh} |\widehat{\mathbf{\Theta}}^{\text{de}}_e(z_0)| \cdot \sum_{i \in [n]} K_h(Z_i - z_0)/n > c\{1 - \alpha, \mathcal{C}(\emptyset, \mathcal{P})\}\Big].$$

According to Algorithm 1, the rejected set in the first iteration at $z_0$ is

$$E_1(z_0) = \Big[e \in \mathcal{C}(\emptyset, \mathcal{P}) : \sqrt{nh} |\widehat{\mathbf{\Theta}}^{\text{de}}_e(z_0)| \cdot \sum_{i \in [n]} K_h(Z_i - z_0)/n > c\{1 - \alpha, \mathcal{C}(\emptyset, \mathcal{P})\}\Big].$$

Under the event $\mathcal{E}_1$, we have $E'_0 \cap \mathcal{C}(\emptyset, \mathcal{P}) \subseteq E_1(z_0)$ and since $\mathcal{P}\{E'_0 \cap \mathcal{C}(\emptyset, \mathcal{P})\} = \mathcal{P}(E'_0)$, we have $\mathcal{P}\{E_1(z_0)\} = \mathcal{P}(E'_0) = 1$. Therefore,

$$\mathbb{P}(\psi_\alpha = 1) \geqslant \mathbb{P}(\mathcal{E}_1). \tag{S128}$$

It suffices to bound $\mathbb{P}(\mathcal{E}_1)$ then. We consider two events

$$\mathcal{E}_2 = \Big[\min_{e\in E_0'} \sqrt{nh}|\boldsymbol{\Theta}_e(z_0)| \cdot \sum_{i\in[n]} K_h(Z_i - z_0)/n > 2c\{1-\alpha, \mathcal{C}(\emptyset, \mathcal{P})\}\Big];$$

$$\mathcal{E}_3 = \Big[\max_{e\in V\times V} \sqrt{nh}|\widehat{\boldsymbol{\Theta}}_e^{\mathrm{de}}(z_0) - \boldsymbol{\Theta}_e(z_0)| \cdot \sum_{i\in[n]} K_h(Z_i - z_0)/n \leqslant c\{1-\alpha, \mathcal{C}(\emptyset, \mathcal{P})\}\Big].$$

We have $\mathbb{P}(\mathcal{E}_1) \geqslant \mathbb{P}(\mathcal{E}_2 \cup \mathcal{E}_3)$. By Lemmas S2 and S3, we have

$$\mathbb{P}\Big\{\sup\nolimits_z \Big|\sum\nolimits_{i\in[n]} K_h(Z_i - z)/n - f_Z(z)\Big| > \sqrt{\log(d/h)/nh}\Big\} < 3/d. \tag{S129}$$

Combining with (14) in Corollary 1, we have with probability at least $1-6/d$,

$$\sup_z \max_{e\in V\times V} \sqrt{nh}|\widehat{\boldsymbol{\Theta}}_e^{\mathrm{de}}(z) - \boldsymbol{\Theta}_e(z)| \cdot \sum_{i\in[n]} K_h(Z_i - z)/n \leqslant C\sqrt{\log(d/h)/nh} \cdot \sqrt{nh}.$$

For any fixed $\alpha \in (0,1)$ and sufficiently large $d, n$, as $\mathcal{C}(\emptyset, \mathcal{P}) \subseteq V \times V$, we have

$$c\{1-\alpha, \mathcal{C}(\emptyset, \mathcal{P})\} \leqslant c(1-\alpha, V\times V) \leqslant C\sqrt{\log(d/h)/nh} \cdot \sqrt{nh}.$$

Thus $\mathbb{P}(\mathcal{E}_3) > 1 - 6/d$. Similarly, we also have $\mathbb{P}(\mathcal{E}_2) > 1 - 3/d$. By (S128), we have

$$\mathbb{P}(\psi_\alpha = 1) \geqslant \mathbb{P}(\mathcal{E}_1) \geqslant \mathbb{P}(\mathcal{E}_2 \cup \mathcal{E}_3) \geqslant 1 - 9/d.$$

Therefore, we complete the proof of the theorem.

## Web Appendix I. Technical Lemmas on Covering Number

In this section, we present some technical lemmas on the covering number of some function classes. Lemma S13 provides an upper bound on the covering number for the class of function of bounded variation. Lemma S14 provides an upper bound on the covering number of a class of Lipschitz function. Lemma S15 provides an upper bound on the covering numbers for function classes generated from the product and addition of two function classes.

Lemma S13: *(Lemma 3 in Giné and Nickl, 2009) Let $K : \mathbb{R} \to \mathbb{R}$ be a function of bounded variation. Define the function class $\mathcal{F}_h = [K\{(t - \cdot)/h)\} \mid t \in \mathbb{R}]$. Then, there exists $C_K < \infty$ independent of $h$ and of $K$ such that for all $0 < \epsilon < 1$,*

$$\sup_Q N\{\mathcal{F}_h, L_2(Q), \epsilon\} \leqslant \left(\frac{2 \cdot C_K \cdot \|K\|_{\mathrm{TV}}}{\epsilon}\right)^4,$$

*where* $\|K\|_{\mathrm{TV}}$ *is the total variation norm of the function* $K$*.*

LEMMA S14: *Let* $f(l)$ *be a Lipschitz function defined on* $[0,1]$ *such that* $|f(l)-f(l')| \leqslant L_f \cdot |l-l'|$ *for any* $l, l' \in [0,1]$*. We define the constant function class* $\mathcal{F} = \{g_l := f(l) \mid l \in [0,1]\}$*. For any probability measure* $Q$*, the covering number of the function class* $\mathcal{F}$ *satisfies*

$$N\{\mathcal{F}, L_2(Q), \epsilon\} \leqslant \frac{L_f}{\epsilon},$$

*where* $\epsilon \in (0,1)$*.*

*Proof.* Let $\mathcal{N} = \left\{\frac{i\epsilon}{L_f} \mid i = 1, \ldots, \frac{L_f}{\epsilon}\right\}$. By definition of $\mathcal{N}$, for any $l \in [0,1]$, there exists an $l' \in \mathcal{N}$ such that $|l - l'| \leqslant \epsilon/L_f$. Thus, we have

$$|f(l) - f(l')| \leqslant L_f \cdot |l - l'| \leqslant \epsilon.$$

This implies that $\{g_l \mid l \in \mathcal{N}\}$ is an $\epsilon$-cover of the function class $\mathcal{F}$. To complete the proof, we note that the cardinality of the set $|\mathcal{N}| \leqslant L_f/\epsilon$.

LEMMA S15: *Let* $\mathcal{F}_1$ *and* $\mathcal{F}_2$ *be two function classes satisfying*

$$N\{\mathcal{F}_1, L_2(Q), a_1\epsilon\} \leqslant C_1\epsilon^{-v_1} \quad \text{and} \quad N\{\mathcal{F}_2, L_2(Q), a_2\epsilon\} \leqslant C_2\epsilon^{-v_2}$$

*for some* $C_1, C_2, a_1, a_2, v_1, v_2 > 0$ *and any* $0 < \epsilon < 1$*. Define* $\|F_\ell\|_\infty = \sup_{f \in \mathcal{F}_\ell} \|f\|_\infty$ *for* $\ell = 1, 2$ *and* $U = \|\mathcal{F}_1\|_\infty \vee \|\mathcal{F}_2\|_\infty$*. For the function classes* $\mathcal{F}_\times = \{f_1 f_2 \mid f_1 \in \mathcal{F}_1,\ f_2 \in \mathcal{F}_2\}$ *and* $\mathcal{F}_+ = \{f_1 + f_2 \mid f_1 \in \mathcal{F}_1,\ f_2 \in \mathcal{F}_2\}$*, we have for any* $\epsilon \in (0,1)$*,*

$$N\{\mathcal{F}_\times, L_2(Q), \epsilon\} \leqslant C_1 \cdot C_2 \cdot \left(\frac{2a_1 U}{\epsilon}\right)^{v_1} \cdot \left(\frac{2a_2 U}{\epsilon}\right)^{v_2}$$

*and*

$$N\{\mathcal{F}_+, L_2(Q), \epsilon\} \leqslant C_1 \cdot C_2 \cdot \left(\frac{2a_1}{\epsilon}\right)^{v_1} \cdot \left(\frac{2a_2}{\epsilon}\right)^{v_2}.$$

LEMMA S16: *Let* $w_z(u) = K_h(u - z)$*. We define the function classes*

$$\mathcal{K}_1 = \{w_z(\cdot) \mid z \in [0,1]\} \qquad \text{and} \qquad \mathcal{K}_2 = [\mathbb{E}\{w_z(Z)\} \mid z \in [0,1]].$$

*Given Assumptions 1-2, we have for any* $\epsilon \in (0,1)$,

$$\sup_Q N\{\mathcal{K}_1, L_2(Q), \epsilon\} \leqslant \left(\frac{2 \cdot C_K \cdot \|K\|_{\mathrm{TV}}}{h\epsilon}\right)^4$$

*and*

$$\sup_Q N\{\mathcal{K}_2, L_2(Q), \epsilon\} \leqslant \frac{2}{h\epsilon} \cdot \|K\|_{\mathrm{TV}} \cdot \bar{f}_Z.$$

*Moreover, let* $k_z(u) = w_z(u) - \mathbb{E}\{w_z(Z)\}$ *and let* $\mathcal{K} = \{k_z(\cdot) \mid z \in [0,1]\}$. *We have*

$$\sup_Q N\{\mathcal{K}, L_2(Q), \epsilon\} \leqslant \left(\frac{4 \cdot \|K\|_{\mathrm{TV}} \cdot C_K^{4/5} \cdot \bar{f}_Z^{1/5}}{h\epsilon}\right)^5.$$

*Proof.* The covering number for the function class $\mathcal{K}_1$ is obtained by an application of Lemma S13. To obtain the covering number for $\mathcal{K}_2$, we show that the constant function $\mathbb{E}\{w_z(Z)\}$ is Lipschitz. The covering number is obtained by applying Lemma S14. Finally, we note that the function class $\mathcal{K}$ is generated from the addition of the two function classes $\mathcal{K}_1$ and $\mathcal{K}_2$. The covering number can be obtained by an application of Lemma S15. The details are deferred to Web Appendix I.1.

LEMMA S17: *Let* $g_{z,jk}(u, X_{ij}, Y_{ik}) = K_h(u-z)X_{ij}Y_{ik}$. *We define the function classes*

$$\mathcal{G}_{1,jk} = \{g_{z,jk}(\cdot) \mid z \in [0,1]\} \qquad \text{and} \qquad \mathcal{G}_{2,jk} = \left[\mathbb{E}\{g_{z,jk}(Z, X_j, Y_k)\} \mid z \in [0,1]\right].$$

*Given Assumptions 1-2, for all* $\epsilon \in (0,1)$,

$$\sup_Q N\{\mathcal{G}_{1,jk}, L_2(Q), \epsilon\} \leqslant \left(\frac{2 \cdot M_X^2 \cdot \log d \cdot C_K \cdot \|K\|_{\mathrm{TV}}}{h\epsilon}\right)^4$$

*and*

$$\sup_Q N\{\mathcal{G}_{2,jk}, L_2(Q), \epsilon\} \leqslant \frac{2}{h\epsilon} \cdot \|K\|_{\mathrm{TV}} \cdot \bar{f}_Z \cdot M_\sigma,$$

*with probability at least* $1-1/d$. *Moreover, let* $q_{z,jk}(u, X_{ij}, Y_{ik}) = g_{z,jk}(u, X_{ij}, Y_{ik}) - \mathbb{E}\{g_{z,jk}(Z, X_j, Y_k)\}$ *and let* $\mathcal{G}_{jk} = \{q_{z,jk}(\cdot) \mid z \in [0,1]\}$. *We have*

$$\sup_Q N\{\mathcal{G}_{jk}, L_2(Q), \epsilon\} \leqslant \left(\frac{4 \cdot \|K\|_{\mathrm{TV}} \cdot C_K^{4/5} \cdot \bar{f}_Z^{1/5} \cdot M_\sigma^{1/5} \cdot M_X^{8/5} \cdot \log^{4/5} d}{h\epsilon}\right)^5.$$

*with probability at least* $1 - 1/d$.

*Proof.* The proof uses the same set of argument as in the proof of Lemma S16. The probability statement comes from the fact that we upper bound the random variable $X_j$ by $M_X \cdot \sqrt{\log d}$ for some constant $M_X > 0$. The details are deferred to Web Appendix I.2.

Lemma S18: *Let* $J^{(1)}_{z,jk}(u, \boldsymbol{X}_i, \boldsymbol{Y}_i) = \sqrt{h}\cdot\{\boldsymbol{\Theta}_j(z)\}^T \cdot \left[K_h(u-z)\boldsymbol{X}_i\boldsymbol{Y}_i^T - \mathbb{E}\left\{K_h(Z-z)\boldsymbol{X}\boldsymbol{Y}^T\right\}\right]\cdot \boldsymbol{\Theta}_k(z)$ *and let* $\mathcal{J}^{(1)}_{jk} = \{J^{(1)}_{z,jk} \mid z \in [0,1]\}$*. Given Assumptions 1-2, for all* $\epsilon \in (0,1)$

$$\sup_Q N\{\mathcal{J}^{(1)}_{jk}, L_2(Q), \epsilon\} \leqslant C \cdot \left(\frac{d^{5/4}\cdot \log^{3/2} d}{\sqrt{h}\cdot \epsilon}\right)^6,$$

*with probability at least* $1 - 1/d$*, where* $C > 0$ *is a generic constant that does not depend on* $d$*,* $h$*, and* $n$*.*

*Proof.* The proof is deferred to Web Appendix I.3.

Lemma S19: *Let* $J^{(2)}_{z,jk}(u) = \sqrt{h}\cdot \{\boldsymbol{\Theta}_j(z)\}^T \cdot [K_h(u-z) - \mathbb{E}\{K_h(Z-z)\}]\cdot \boldsymbol{\Sigma}(z)\cdot \boldsymbol{\Theta}_k(z)$ *and let* $\mathcal{J}^{(2)}_{jk} = \{J^{(2)}_{z,jk} \mid z \in [0,1]\}$*. Given Assumptions 1-2, for all probability measures* $Q$ *on* $\mathbb{R}$ *and all* $0 < \epsilon < 1$*,*

$$N\{\mathcal{J}^{(2)}_{jk}, L_2(Q), \epsilon\} \leqslant C \cdot \left(\frac{d^{1/6}}{h^{4/3}\cdot \epsilon}\right)^6,$$

*where* $C > 0$ *is a generic constant that does not depend on* $d$*,* $h$*, and* $n$*.*

*Proof.* We first note that $\mathcal{J}^{(2)}_{jk}$ is a function class generated from the product of two function classes $\mathcal{K}$ as in Lemma S16 and $\Theta_{jk} = \{\boldsymbol{\Theta}_{jk}(z) \mid z \in [0,1]\}$. To obtain the covering number of $\Theta_{jk}$, we show that the constant function $\boldsymbol{\Theta}_{jk}(z)$ is Lipschitz and apply Lemma S14. We then apply Lemma S15 to obtain the covering number of $\mathcal{J}^{(2)}_{jk}$. The details are deferred to Web Appendix I.4.

Web Appendix I.1 *Proof of Lemma S16*

Let $w_z(u) = K_h(u-z)$ and that $k_z(u) = w_z(u) - \mathbb{E}\{w_z(Z)\}$. We first obtain the covering number for the function classes

$$\mathcal{K}_1 = \{w_z(\cdot) \mid z \in [0,1]\} \qquad \text{and} \qquad \mathcal{K}_2 = [\mathbb{E}\{w_z(Z)\} \mid z \in [0,1]].$$

Then, we apply Lemma S15 to obtain the covering number of the function class

$$\mathcal{K} = \{k_z(\cdot) \mid z \in [0,1]\}.$$

**Covering number for $\mathcal{K}_1$:** By an application of Lemma S13, the covering number for $\mathcal{K}_1$ is

$$\sup_Q N\{\mathcal{K}_1, L_2(Q), \epsilon\} \leqslant \left(\frac{2 \cdot C_K \cdot \|K\|_{\mathrm{TV}}}{h\epsilon}\right)^4. \tag{S130}$$

**Covering number for $\mathcal{K}_2$:** First, note that $\mathbb{E}\{w_z(Z)\} = \int K_h(z-Z)f_Z(Z)dZ = (K_h * f_Z)(z)$ is a function of $z$ generated by the convolution $(K_h * f_Z)(z)$. By the property of the derivative of a convolution as in (S18), we have

$$\sup_{z_0\in[0,1]} \left|\frac{\partial}{\partial z}\mathbb{E}\{w_z(Z)\}\Big|_{z=z_0}\right| = \sup_{z_0\in[0,1]} \left|\dot{K}_h * f_Z(z_0)\right| = \left\|(\dot{K}_h * f_Z)(z)\right\|_\infty \leqslant \left\|\dot{K}_h\right\|_1 \cdot \|f_Z\|_\infty, \tag{S131}$$

where the last expression is obtained by an application of Young's inequality. The expression in (S131) depends on the quantity $\|\dot{K}_h\|_1$, which is equal to the following expression

$$\left\|\dot{K}_h\right\|_1 = \int \frac{1}{h^2}\left|\dot{K}\left(\frac{Z-z}{h}\right)\right| dZ = \frac{1}{h}\int \left|\dot{K}(u)\right| du = \frac{1}{h}\cdot \|K\|_{\mathrm{TV}}, \tag{S132}$$

where the second inequality holds by a change of variable, and $\|K\|_{\mathrm{TV}}$ is the total variation of the function $K(\cdot)$. Substituting (S132) into (S131) and by Assumption 1, we have

$$\sup_{z_0\in[0,1]} \left|\frac{\partial}{\partial z}\mathbb{E}\{w_z(Z)\}\Big|_{z=z_0}\right| \leqslant \frac{1}{h}\cdot \|K\|_{\mathrm{TV}} \cdot \bar{f}_Z. \tag{S133}$$

Thus, for any $z_1, z_2 \in [0,1]$, we have

$$|\mathbb{E}\{w_{z_1}(Z)\} - \mathbb{E}\{w_{z_2}(Z)\}| \leqslant \frac{1}{h}\cdot \|K\|_{\mathrm{TV}} \cdot \bar{f}_Z \cdot |z_1 - z_2|,$$

implying that $\mathbb{E}\{w_z(Z)\}$ is a Lipschitz continuous function with Lipschitz constant $h^{-1} \cdot \|K\|_{\mathrm{TV}} \cdot \bar{f}_Z$. By Lemma S14, an upper bound for the covering number of $\mathcal{K}_2$ is

$$\sup_Q N\{\mathcal{K}_2, L_2(Q), \epsilon\} \leqslant \frac{2}{h\epsilon}\cdot \|K\|_{\mathrm{TV}} \cdot \bar{f}_Z. \tag{S134}$$

**Covering number of the function class $\mathcal{K}$:** The function class $\mathcal{K}$ can be written as

$$\mathcal{K} = \{f_1 - f_2 \mid f_1 \in \mathcal{K}_1,\ f_2 \in \mathcal{K}_2\}.$$

By an application of Lemma S15 with $C_1 = (2\cdot C_K \cdot \|K\|_{\mathrm{TV}})^4$, $C_2 = 2\cdot \bar{f}_Z \cdot \|K\|_{\mathrm{TV}}$, $a_1 =$

$a_2 = h^{-1}$, $v_1 = 4$, and $v_2 = 1$, along with (S130) and (S134), we obtain

$$\sup_Q N\{\mathcal{K}, L_2(Q), \epsilon\} \leqslant \left( \frac{4 \cdot \|K\|_{\mathrm{TV}} \cdot C_K^{4/5} \cdot \bar{f}_Z^{1/5}}{h\epsilon} \right)^5.$$

Web Appendix I.2 *Proof of Lemma S17*

Throughout the proof, we condition on the event

$$\mathcal{A} = \left\{ \max_{i\in[n]} \max_{j\in[d]} \max(|X_{ij}|, |Y_{ij}|) \leqslant M_X \cdot \sqrt{\log d} \right\}. \tag{S135}$$

Since $\boldsymbol{X}$ and $\boldsymbol{Y}$ conditioned on $Z$ are Gaussian random variables, the event $\mathcal{A}$ occurs with probability at least $1 - 1/d$ for sufficiently large constant $M_X > 0$.

Recall that $g_{z,jk}(u, X_{ij}, Y_{ik}) = K_h(u - z)X_{ij}Y_{ik}$ and that $q_{z,jk}(u, X_{ij}, Y_{ik}) = K_h(u - z)X_{ij}Y_{ik} - \mathbb{E}\{K_h(Z - z)X_jY_k\}$. We first obtain the covering number of the function classes

$$\mathcal{G}_{1,jk} = \{g_{z,jk}(\cdot) \mid z \in [0, 1]\}$$

and

$$\mathcal{G}_{2,jk} = [\mathbb{E}\{g_{z,jk}(Z, X_j, Y_k)\} \mid z \in [0, 1]].$$

Then, we apply Lemma S15 to obtain the covering number of the function class

$$\mathcal{G}_{jk} = \{q_{z,jk}(\cdot) \mid z \in [0, 1],\ j, k \in [d]\}.$$

**Covering number for $\mathcal{G}_{1,jk}$:** Conditioned on the event $\mathcal{A}$ in (S135), we have

$$g_{z,jk}(u, X_{ij}, Y_{ik}) = K_h(u - z)X_{ij}Y_{ik} \leqslant M_X^2 \cdot \log d \cdot K_h(u - z).$$

By an application of Lemma S13, the covering number for $\mathcal{G}_{1,jk}$ is

$$\sup_Q N\{\mathcal{G}_{1,jk}, L_2(Q), \epsilon\} \leqslant \left( \frac{2 \cdot M_X^2 \cdot \log d \cdot C_K \cdot \|K\|_{\mathrm{TV}}}{h\epsilon} \right)^4. \tag{S136}$$

**Covering number for $\mathcal{G}_{2,jk}$:** We now obtain the covering number for $\mathcal{G}_{2,jk}$ by showing that the function $\mathbb{E}\{g_{z,jk}(Z, X_j, Y_k)\}$ is Lipschitz. First, note that

$$\mathbb{E}\{g_{z,jk}(Z, X_j, Y_k)\} = \mathbb{E}\{K_h(Z - z) \cdot \boldsymbol{\Sigma}_{jk}(Z)\} = \int K_h(z - Z) \cdot \varphi_{jk}(Z) dZ = (K_h * \varphi_{jk})(z),$$

where $\varphi_{jk}(Z) = f_Z(Z) \cdot \boldsymbol{\Sigma}_{jk}(Z)$ and $K_h * \varphi_{jk}$ is the convolution between $K_h$ and $\varphi_{jk}$. Similar to (S131)-(S133), we have

$$\begin{aligned}\sup_{z_0\in[0,1]}\max_{j,k\in[d]}\left|\frac{\partial}{\partial z}\mathbb{E}\{g_{z,jk}(Z,X_j,Y_k)\}\Big|_{z=z_0}\right| &= \sup_{z_0\in[0,1]}\max_{j,k\in[d]}\left|(\dot{K}_h*\varphi_{jk})(z_0)\right| \\ &= \max_{j,k\in[d]}\left\|(\dot{K}_h*\varphi_{jk})(z)\right\|_\infty \\ &\leqslant \left\|\dot{K}_h\right\|_1\cdot\max_{j,k\in[d]}\|\varphi_{jk}\|_\infty \\ &\leqslant \frac{1}{h}\cdot\|K\|_{\mathrm{TV}}\cdot\bar{f}_Z\cdot M_\sigma,\end{aligned} \tag{S137}$$

where the first inequality is obtained by an application of Young's inequality, and the last expression is obtained by (S132) and Assumptions 1-2.

Equation S137 implies that for any $z_1, z_2 \in [0,1]$,

$$|\mathbb{E}\{g_{z_1,jk}(Z,X_j,Y_k)\}-\mathbb{E}\{g_{z_2,jk}(Z,X_j,Y_k)\}| \leqslant \frac{1}{h}\cdot\|K\|_{\mathrm{TV}}\cdot\bar{f}_Z\cdot M_\sigma\cdot|z_1-z_2|,$$

implying that $\mathbb{E}\{g_{z,jk}(Z,X_j,Y_k)\}$ is a Lipschitz continuous function with Lipschitz constant $h^{-1}\cdot\|K\|_{\mathrm{TV}}\cdot\bar{f}_Z\cdot M_\sigma$. By an application of Lemma S14, we have

$$\sup_Q N\{\mathcal{G}_{2,jk}, L_2(Q),\epsilon\} \leqslant \frac{2}{h\epsilon}\cdot\|K\|_{\mathrm{TV}}\cdot\bar{f}_Z\cdot M_\sigma. \tag{S138}$$

**Covering number of the function class $\mathcal{G}_{jk}$:** The function class $\mathcal{G}_{jk}$ can be written as

$$\mathcal{G}_{jk} = \{f_{1,jk}-f_{2,jk} \mid f_{1,jk}\in\mathcal{G}_{1,jk},\ f_{2,jk}\in\mathcal{G}_{2,jk},\ j,k\in[d]\}.$$

By an application of Lemma S15 with $C_1 = (2\cdot C_K\cdot\|K\|_{\mathrm{TV}}\cdot M_X^2)^4$, $C_2 = 2\cdot\bar{f}_Z\cdot\|K\|_{\mathrm{TV}}\cdot M_\sigma$, $a_1 = h^{-1}\cdot\log d$, $a_2 = h^{-1}$, $v_1 = 4$, and $v_2 = 1$, along with (S136) and (S138), we obtain

$$\sup_Q N\{\mathcal{G}_{jk}, L_2(Q),\epsilon\} \leqslant \left(\frac{4\cdot\|K\|_{\mathrm{TV}}\cdot C_K^{4/5}\cdot\bar{f}_Z^{1/5}\cdot M_\sigma^{1/5}\cdot M_X^{8/5}\cdot\log^{4/5}d}{h\epsilon}\right)^5, \tag{S139}$$

as desired.

Web Appendix I.3 *Proof of Lemma S18*

Similar to the proof of Lemma S17, we condition on the event

$$\mathcal{A} = \left\{\max_{i\in[n]}\max_{j\in[d]}\max(|X_{ij}|,|Y_{ij}|) \leqslant M_X\cdot\sqrt{\log d}\right\}.$$

The event $\mathcal{A}$ holds with probability at least $1-1/d$.

Recall that $J^{(1)}_{z,jk}(u,\boldsymbol{X}_i,\boldsymbol{Y}_i) = \sqrt{h}\cdot\{\boldsymbol{\Theta}_j(z)\}^T\cdot\left[K_h(u-z)\boldsymbol{X}_i\boldsymbol{Y}_i^T - \mathbb{E}\left\{K_h(Z-z)\boldsymbol{X}\boldsymbol{Y}^T\right\}\right]\cdot$

$\mathbf{\Theta}_k(z)$ and let $\mathcal{J}_{jk}^{(1)} = \left\{ J_{z,jk}^{(1)} \mid z \in [0,1] \right\}$. To obtain the covering number of the function class $\mathcal{J}_{jk}^{(1)}$, we consider bounding the covering number of a larger class of function. To this end, we define $\mathbf{\Phi}_\omega^{(1)}(u, \boldsymbol{X}_i, \boldsymbol{Y}_i) = \sqrt{h} \cdot \left[ K_h(u-z)\boldsymbol{X}_i\boldsymbol{Y}_i^T - \mathbb{E}\left\{ K_h(Z-z)\boldsymbol{X}\boldsymbol{Y}^T \right\} \right]$ to be a $d \times d$ matrix. We denote the $(j,k)$th element of $\Phi_\omega^{(1)}(u, \boldsymbol{X}_i, \boldsymbol{Y}_i)$ as $\Phi_{\omega,jk}^{(1)}(u, X_{ij}, Y_{ik}) = \sqrt{h} \cdot q_{\omega,jk}(u, X_{ij}, Y_{ik})$, where $q_{\omega,jk}(u, X_{ij}, Y_{ik}) = K_h(u-\omega)X_{ij}Y_{ik} - \mathbb{E}\{K_h(Z-\omega)X_jY_k\}$. We aim to obtain an $\epsilon$-cover $\mathcal{N}^{(1')}$ for the following function class

$$\mathcal{J}_{jk}^{(1')} = \left[ \{\mathbf{\Theta}_j(z)\}^T \mathbf{\Phi}_\omega^{(1)}(\cdot) \mathbf{\Theta}_k(z) \mid \omega, z \in [0,1] \right].$$

In other words, we show that for any $(\omega_1, z_1) \in [0,1]^2$, there exists $(\omega_2, z_2) \in \mathcal{N}^{(1')}$ such that

$$\left\| \{\mathbf{\Theta}_j(z)\}^T \mathbf{\Phi}_\omega^{(1)}(u, \boldsymbol{X}_i, \boldsymbol{Y}_i) \mathbf{\Theta}_k(z) - \{\mathbf{\Theta}_j(z')\}^T \mathbf{\Phi}_{\omega'}^{(1)}(u, \boldsymbol{X}_i, \boldsymbol{Y}_i) \mathbf{\Theta}_k(z') \right\|_{L_2(Q)} \leqslant \epsilon.$$

Given any $j, k \in [d]$, $\omega, \omega', z, z' \in [0,1]$, by the triangle inequality, we have

$$\begin{aligned}
&\left\| \{\mathbf{\Theta}_j(z_1)\}^T \mathbf{\Phi}_{\omega_1}^{(1)}(u, \boldsymbol{X}_i, \boldsymbol{Y}_i) \mathbf{\Theta}_k(z_1) - \{\mathbf{\Theta}_j(z_2)\}^T \mathbf{\Phi}_{\omega_2}^{(1)}(u, \boldsymbol{X}_i, \boldsymbol{Y}_i) \mathbf{\Theta}_k(z_2) \right\|_{L_2(Q)} \\
&\leqslant \underbrace{\left\| \{\mathbf{\Theta}_j(z_1) - \mathbf{\Theta}_j(z_2)\}^T \mathbf{\Phi}_{\omega_1}^{(1)}(u, \boldsymbol{X}_i, \boldsymbol{Y}_i) \mathbf{\Theta}_k(z_1) \right\|_{L_2(Q)}}_{I_1} + \underbrace{\left\| \{\mathbf{\Theta}_j(z_2)\}^T \left\{ \mathbf{\Phi}_{\omega_1}^{(1)}(u, \boldsymbol{X}_i, \boldsymbol{Y}_i) - \mathbf{\Phi}_{\omega_2}^{(1)}(u, \boldsymbol{X}_i, \boldsymbol{Y}_i) \right\} \mathbf{\Theta}_k(z_1) \right\|_{L_2(Q)}}_{I_2} \\
&\quad + \underbrace{\left\| \{\mathbf{\Theta}_j(z_2)\}^T \mathbf{\Phi}_{\omega_2}^{(1)}(u, \boldsymbol{X}_i, \boldsymbol{Y}_i) \{\mathbf{\Theta}_k(z_1) - \mathbf{\Theta}_k(z_2)\} \right\|_{L_2(Q)}}_{I_3}.
\end{aligned} \tag{S140}$$

We now obtain the upper bounds for $I_1, I_2$, and $I_3$.

**Upper bound for $I_1$ and $I_3$:** First, we note that by Holder's inequality, we have

$$I_1 \leqslant \|\mathbf{\Theta}_j(z_1) - \mathbf{\Theta}_j(z_2)\|_1 \cdot \max_{j,k \in [d]} \left\| \Phi_{\omega_1,jk}^{(1)}(u, X_{ij}, Y_{ik}) \right\|_{L_2(Q)} \cdot \|\mathbf{\Theta}_k(z_1)\|_1.$$

Since $\mathbf{\Theta}(z) \in \mathcal{U}(s, M, \rho)$, we have

$$\sup_{z \in [0,1]} \max_{j \in [d]} \|\mathbf{\Theta}_j(z)\|_1 \leqslant M. \tag{S141}$$

Moreover, for any $z_1, z_2 \in [0,1]$, we have

$$\begin{aligned}\sup_{j\in[d]} \|\boldsymbol{\Theta}_j(z_1) - \boldsymbol{\Theta}_j(z_2)\|_1 &\leqslant \sqrt{d}\cdot\|\boldsymbol{\Theta}(z_1) - \boldsymbol{\Theta}(z_2)\|_2 \\ &\leqslant \sqrt{d}\cdot\|\boldsymbol{\Theta}(z_1)\|_2\cdot\|\mathbf{I}_d - \boldsymbol{\Sigma}(z_1)\boldsymbol{\Theta}(z_2)\|_2 \\ &\leqslant \sqrt{d}\cdot\|\boldsymbol{\Theta}(z_1)\|_2\cdot\|\boldsymbol{\Theta}(z_2)\|_2\cdot\|\boldsymbol{\Sigma}(z_1) - \boldsymbol{\Sigma}(z_2)\|_2 \\ &\leqslant \sqrt{d}\cdot\rho^2\cdot d\cdot\|\boldsymbol{\Sigma}(z_1) - \boldsymbol{\Sigma}(z_2)\|_{\max} \\ &\leqslant d^{3/2}\cdot\rho^2\cdot M_\sigma\cdot|z_1 - z_2|,\end{aligned} \tag{S142}$$

where the second to the last inequality follows from the fact that $\boldsymbol{\Theta}(z) \in \mathcal{U}(s, M, \rho)$ and the last inequality follows from Assumption 2. Finally, from (S54) and the definition of $\Phi^{(1)}_{\omega_1,jk}(\cdot) = \sqrt{h}\cdot q_{\omega_1,jk}(\cdot)$, we have

$$\max_{j,k\in[d]} \left\|\Phi^{(1)}_{\omega_1,jk}(u, X_{ij}, Y_{ik})\right\|_{L_2(Q)} \leqslant \frac{2}{\sqrt{h}}\cdot M_X^2\cdot\|K\|_\infty\cdot\log d. \tag{S143}$$

Combining (S141)-(S143), we have

$$I_1 \leqslant d^{3/2}\cdot\log d\cdot\rho^2\cdot M_\sigma\cdot M\cdot M_X^2\cdot\|K\|_\infty\cdot\frac{2}{\sqrt{h}}\cdot|z_1 - z_2|. \tag{S144}$$

We note that $I_3$ can be upper bounded the same way as $I_1$.

**Upper bound for $I_2$:** Recall from (S140) that

$$\begin{aligned}I_2 &= \left\|\{\boldsymbol{\Theta}_j(z_2)\}^T\left\{\boldsymbol{\Phi}^{(1)}_{\omega_1}(u, \boldsymbol{X}_i, \boldsymbol{Y}_i) - \boldsymbol{\Phi}^{(1)}_{\omega_2}(u, \boldsymbol{X}_i, \boldsymbol{Y}_i)\right\}\boldsymbol{\Theta}_k(z_1)\right\|_{L_2(Q)} \\ &\leqslant \|\boldsymbol{\Theta}_k(z_1)\|\cdot\|\boldsymbol{\Theta}_j(z_2)\|_1\cdot\max_{j,k\in[d]}\left\|\sqrt{h}\cdot\{q_{\omega_1,jk}(u, X_{ij}, Y_{ik}) - q_{\omega_2,jk}(u, X_{ij}, Y_{ik})\}\right\|_{L_2(Q)},\end{aligned}$$

where the inequality holds by Holder's inequality and the definition of $\boldsymbol{\Phi}^{(1)}_\omega(u, \boldsymbol{X}_i, \boldsymbol{Y}_i)$. Let $\mathit{\Phi}^{(1)}_{jk} = \left\{\sqrt{h}\cdot q_{\omega,jk}(\cdot) \mid \omega\in[0,1]\right\}$ and recall from Lemma S17 that we constructed an $\epsilon$-cover $\mathcal{N}^{(1'')} \subset [0,1]$ for the function class $\mathit{\Phi}^{(1)}_{jk}$ with cardinality $\left|\mathcal{N}^{(1'')}\right| = \left(\frac{4\cdot\|K\|_{\mathrm{TV}}\cdot C_K^{4/5}\cdot\bar{f}_Z^{1/5}\cdot M_\sigma^{1/5}\cdot M_X^{8/5}\cdot\log^{4/5} d}{\sqrt{h}\cdot\epsilon}\right)^5$. Since the construction of the $\epsilon$-cover in Lemma S17 is independent of the indices $j$ and $k$, we have that for any $j,k\in[d]$ and $\omega_1\in[0,1]$, there exists a $\omega_2 \in \mathcal{N}^{(1'')}$ such that

$$\max_{j,k\in[d]}\left\|\sqrt{h}\cdot\{q_{\omega_1,jk}(u, X_{ij}, Y_{ik}) - q_{\omega_2,jk}(u, X_{ij}, Y_{ik})\}\right\|_{L_2(Q)} \leqslant \epsilon. \tag{S145}$$

Thus, by (S141) and (S145), we have

$$I_2 \leqslant M^2\cdot\epsilon. \tag{S146}$$

**Covering number of the function class $\mathcal{J}_{jk}^{(1)}$:** Since $\mathcal{J}_{jk}^{(1)} \subset \mathcal{J}_{jk}^{(1')}$, the covering number of $\mathcal{J}_{jk}^{(1)}$ is upper bounded by the covering number of $\mathcal{J}_{jk}^{(1')}$. It suffices to construct an $\epsilon$-cover of the function class $\mathcal{J}_{jk}^{(1')}$. In the following, we show that $\mathcal{N}^{(1')} = \mathcal{N}^{(1'')} \times \left\{ i \cdot \epsilon \cdot \sqrt{h} \mid i = 1, \ldots, \frac{1}{\epsilon \cdot \sqrt{h}} \right\}$ is an $\epsilon$-cover of $\mathcal{J}_{jk}^{(1')}$. For any $(\omega_1, z_1) \in [0,1]^2$, there exists $(\omega_2, z_2) \in \mathcal{N}^{(1')}$ such that (S145) holds and that $|z_1 - z_2| \leqslant \sqrt{h} \cdot \epsilon$. Thus, combining (S144) and (S146), we have

$$\begin{aligned}
&\left\| \{\boldsymbol{\Theta}_j(z_1)\}^T \boldsymbol{\Phi}_{\omega_1}^{(1)}(u, \boldsymbol{X}_i, \boldsymbol{Y}_i) \boldsymbol{\Theta}_k(z_1) - \{\boldsymbol{\Theta}_j(z_2)\}^T \boldsymbol{\Phi}_{\omega_2}^{(1)}(u, \boldsymbol{X}_i, \boldsymbol{Y}_i) \boldsymbol{\Theta}_k(z_2) \right\|_{L_2(Q)} \\
&\leqslant 2 \cdot d^{3/2} \cdot \log d \cdot \rho^2 \cdot M_\sigma \cdot M \cdot M_X^2 \cdot \|K\|_\infty \cdot \frac{2}{\sqrt{h}} \cdot |z_1 - z_2| + M^2 \cdot \epsilon \\
&\leqslant 4 \cdot d^{3/2} \cdot \log d \cdot \rho^2 \cdot M_\sigma \cdot M \cdot M_X^2 \cdot \|K\|_\infty \cdot \frac{2}{\sqrt{h}} \cdot \epsilon + M^2 \cdot \epsilon \\
&\leqslant C \cdot d^{3/2} \cdot \log d \cdot \epsilon,
\end{aligned} \tag{S147}$$

where $C > 0$ is a generic constant that does not depend on $d$, $h$, and $n$.

Thus, we have

$$N\{\mathcal{J}_{jk}^{(1')}, L_2(Q), C \cdot d^{3/2} \cdot \log d \cdot \epsilon\} \leqslant \left| \mathcal{N}^{(1')} \right| \leqslant \left( \frac{4 \cdot \|K\|_{\mathrm{TV}} \cdot C_K^{4/5} \cdot \bar{f}_Z^{1/5} \cdot M_\sigma^{1/5} \cdot M_X^{8/5} \cdot \log^{4/5} d}{\sqrt{h} \cdot \epsilon} \right)^5 \cdot \frac{1}{\sqrt{h} \cdot \epsilon}.$$

Since $\mathcal{J}_{jk}^{(1)} \subset \mathcal{J}_{jk}^{(1')}$, the above expression implies that

$$N\{\mathcal{J}_{jk}^{(1)}, L_2(Q), \epsilon\} \leqslant N\{\mathcal{J}_{jk}^{(1')}, L_2(Q), \epsilon\} \leqslant C \cdot \left( \frac{d^{5/4} \cdot \log^{3/2} d}{\sqrt{h} \cdot \epsilon} \right)^6, \tag{S148}$$

as desired.

Web Appendix I.4 *Proof of Lemma S19*

First, we note that

$$\begin{aligned}
J_{z,jk}^{(2)}(u) &= \sqrt{h} \cdot \{\boldsymbol{\Theta}_j(z)\}^T \cdot [K_h(u - z) - \mathbb{E}\{K_h(Z - z)\}] \cdot \boldsymbol{\Sigma}(z) \cdot \boldsymbol{\Theta}_k(z) \\
&= \sqrt{h} \cdot k_z(u) \cdot \boldsymbol{\Theta}_{jk}(z),
\end{aligned}$$

where $k_z(u) = K_h(u - z) - \mathbb{E}\{K_h(Z - z)\}$. Let $\mathcal{J}_{jk}^{(2)} = \left\{ J_{z,jk}^{(2)} \mid z \in [0,1] \right\}$. Furthermore, recall that $\mathcal{K} = \{k_z(\cdot) \mid z \in [0,1]\}$ and let $\Theta_{jk} = \{\boldsymbol{\Theta}_{jk}(z) \mid z \in [0,1]\}$. The function class $\mathcal{J}_{jk}^{(2)}$ can be written as

$$\mathcal{J}_{jk}^{(2)} = \{\sqrt{h} \cdot f_1 \cdot f_{2,jk} \mid f_1 \in \mathcal{K},\ f_{2,jk} \in \mathit{\Omega}_{jk}\}.$$

It suffices to obtain the covering number for $\mathcal{K}$ and $\mathit{\Omega}_{jk}$, and apply Lemma S15.

**Covering number of the function class $\mathcal{K}$:** By Lemma S16, we have

$$N\{\mathcal{K}, L_2(Q), \epsilon\} \leqslant \left(\frac{4 \cdot \|K\|_{\mathrm{TV}} \cdot C_K^{4/5} \cdot \bar{f}_Z^{1/5}}{h\epsilon}\right)^5. \tag{S149}$$

**Covering number of the function class $\mathit{\Theta}_{jk}$:** We show that $\mathbf{\Theta}_{jk}(z)$ is Lipschitz, and apply Lemma S14 to obtain the covering number for $\mathit{\Theta}_{jk}$. Similar to (S142), for any $z_1, z_2 \in [0,1]$, we have

$$\begin{aligned} \|\mathbf{\Theta}(z_1) - \mathbf{\Theta}(z_2)\|_{\max} &\leqslant \|\mathbf{\Theta}(z_1)\|_2 \cdot \|\mathbf{\Theta}(z_2) \cdot \{\mathbf{\Sigma}(z_1) - \mathbf{\Sigma}(z_2)\}\|_2 \\ &\leqslant \|\mathbf{\Theta}(z_1)\|_2 \cdot \|\mathbf{\Theta}(z_2)\|_2 \cdot \|\mathbf{\Sigma}(z_1) - \mathbf{\Sigma}(z_2)\|_2 \\ &\leqslant \rho^2 \cdot d \cdot \|\mathbf{\Sigma}(z_1) - \mathbf{\Sigma}(z_2)\|_{\max} \\ &\leqslant \rho^2 \cdot d \cdot M_\sigma \cdot |z_1 - z_2|, \end{aligned}$$

where the last inequality follows from Assumption 2. Since $\mathbf{\Theta}_{jk}(z)$ is $\rho^2 \cdot d \cdot M_\sigma$-Lipschitz, by Lemma S14, we have

$$N\{\mathit{\Theta}_{jk}, L_2(Q), \epsilon\} \leqslant \frac{M_\sigma \cdot \rho^2 \cdot d}{\epsilon}. \tag{S150}$$

**Covering number of the function class $\mathcal{J}_{jk}^{(2)}$:** We now apply Lemma S15 to obtain the covering number of $\mathcal{J}_{jk}^{(2)}$. Applying Lemma S15 with $a_1 = d$, $v_1 = 1$, $C_1 = M_\sigma \cdot \rho^2$, $a_2 = h^{-1}$, $v_2 = 5$, $C_2 = \left(4 \cdot \|K\|_{\mathrm{TV}} \cdot C_K^{4/5} \cdot \bar{f}_Z^{1/5}\right)^5$, and $U = \frac{2}{h} \cdot \|K\|_\infty$, along with (S149) and (S150), we have

$$N\{\mathcal{J}_{jk}^{(2)}, L_2(Q), \sqrt{h} \cdot \epsilon\} \leqslant C \cdot \left(\frac{d^{1/6}}{h^{11/6} \cdot \epsilon}\right)^6,$$

where $C > 0$ is a generic constant that does not depend on $n, d$, and $h$. This implies that

$$N\{\mathcal{J}_{jk}^{(2)}, L_2(Q), \epsilon\} \leqslant C \cdot \left(\frac{d^{1/6}}{h^{4/3} \cdot \epsilon}\right)^6,$$

as desired.

**Web Appendix J. Technical Lemmas on Empirical Process**

In this section, we present some existing tools on empirical process. The following lemma states that the supreme of any empirical process is concentrated near its mean. It follows directly from Theorem 2.3 in Bousquet (2002).

LEMMA S20: *(Theorem A.1 in Van de Geer, 2008) Let $X_1, \ldots, X_n$ be independent random variables and let $\mathcal{F}$ be a function class such that there exists $\eta$ and $\tau^2$ satisfying*

$$\sup_{f \in \mathcal{F}} \|f\|_\infty \leqslant \eta \qquad \text{and} \qquad \sup_{f \in \mathcal{F}} \frac{1}{n} \sum_{i \in n} \operatorname{Var}\{f(X_i)\} \leqslant \tau^2.$$

*Define*

$$Y = \sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i \in [n]} [f(X_i) - \mathbb{E}\{f(X_i)\}] \right|.$$

*Then, for any $t > 0$,*

$$P\left[Y \geqslant \mathbb{E}(Y) + t\sqrt{2\left\{\tau^2 + 2\eta\mathbb{E}(Y)\right\}} + 2t^2\eta/3\right] \leqslant \exp\left(-nt^2\right).$$

The above inequality involves evaluating the expectation of the supreme of the empirical process. The following lemma follows directly from Theorem 3.12 in Koltchinskii (2011). It provides an upper bound on the expectation of the supreme of the empirical process as a function of its covering number.

LEMMA S21: *(Lemma F.1 in Lu et al., 2015) Assume that the functions in $\mathcal{F}$ defined on $\mathcal{X}$ are uniformly bounded by a constant $U$ and $F(\cdot)$ is the envelope of $\mathcal{F}$ such that $|f(x)| \leqslant F(x)$ for all $x \in \mathcal{X}$ and $f \in \mathcal{F}$. Let $\sigma_P^2 = \sup_{f \in \mathcal{F}} \mathbb{E}(f^2)$. Let $X_1, \ldots, X_n$ be* i.i.d. *copies of the random variables $X$. We denote the empirical measure as $\mathbb{P}_n = \frac{1}{n}\sum_{i \in [n]} \delta_{X_i}$. If for some $A, V > 0$ and for all $\epsilon > 0$ and $n \geqslant 1$, the covering entropy satisfies*

$$N\{\mathcal{F}, L_2(\mathbb{P}_n), \epsilon\} \leqslant \left(\frac{A\|F\|_{L_2(\mathbb{P}_n)}}{\epsilon}\right)^V,$$

*then for any* i.i.d. *sub-gaussian mean zero random variables $\xi_1, \ldots, \xi_n$, there exists a universal* *constant* $C$ *such that*

$$\mathbb{E}\left\{\sup_{f\in\mathcal{F}}\frac{1}{n}\left|\sum_{i\in[n]}\xi_i f(X_i)\right|\right\}\leqslant C\left\{\sqrt{\frac{V}{n}}\sigma_P\sqrt{\log\left(\frac{A\|F\|_{L_2(\mathbb{P})}}{\sigma_P}\right)}+\frac{VU}{n}\log\left(\frac{A\|F\|_{L_2(\mathbb{P})}}{\sigma_P}\right)\right\}.$$

*Furthermore, we have*

$$\mathbb{E}\left\{\sup_{f\in\mathcal{F}}\frac{1}{n}\left|\sum_{i\in[n]}[f(X_i)-\mathbb{E}\{f(X_i)\}]\right|\right\}\leqslant C\left\{\sqrt{\frac{V}{n}}\sigma_P\sqrt{\log\left(\frac{A\|F\|_{L_2(\mathbb{P})}}{\sigma_P}\right)}+\frac{VU}{n}\log\left(\frac{A\|F\|_{L_2(\mathbb{P})}}{\sigma_P}\right)\right\}.$$

## References

Bondy, J. A. and Murty, U. S. R. (1976). *Graph theory with applications.* Macmillan London.

Bousquet, O. (2002). A Bennett concentration inequality and its application to suprema of empirical processes. *Comptes Rendus Mathematique* **334,** 495–500.

Chernozhukov, V., Chetverikov, D., and Kato, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *The Annals of Statistics* **41,** 2786–2819.

Chernozhukov, V., Chetverikov, D., and Kato, K. (2014a). Anti-concentration and honest, adaptive confidence bands. *The Annals of Statistics* **42,** 1787–1818.

Chernozhukov, V., Chetverikov, D., and Kato, K. (2014b). Gaussian approximation of suprema of empirical processes. *The Annals of Statistics* **42,** 1564–1597.

Giné, E. and Nickl, R. (2009). An exponential inequality for the distribution function of the kernel density estimator, with applications to adaptive estimation. *Probability Theory and Related Fields* **143,** 569–596.

Isserlis, L. (1918). On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables. *Biometrika* **12,** 134–139.

Koltchinskii, V. (2011). *Oracle Inequalities in Empirical Risk Minimization and Sparse Recovery Problems: Ecole d'Eté de Probabilités de Saint-Flour XXXVIII-2008.* Springer Science & Business Media.

Lovász, L. (2012). *Large networks and graph limits*, volume 60. American Mathematical Soc.

Lu, J., Kolar, M., and Liu, H. (2015). Post-regularization confidence bands for high dimensional nonparametric models with local sparsity. *arXiv preprint arXiv:1503.02978* .

Pagan, A. and Ullah, A. (1999). *Nonparametric econometrics*. Cambridge university press.

Romano, J. P. and Wolf, M. (2005). Exact and approximate stepdown methods for multiple hypothesis testing. *Journal of the American Statistical Association* **100,** 94–108.

Van de Geer, S. A. (2008). High-dimensional generalized linear models and the lasso. *The Annals of Statistics* **36,** 614–645.

Van Der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence*. Springer.

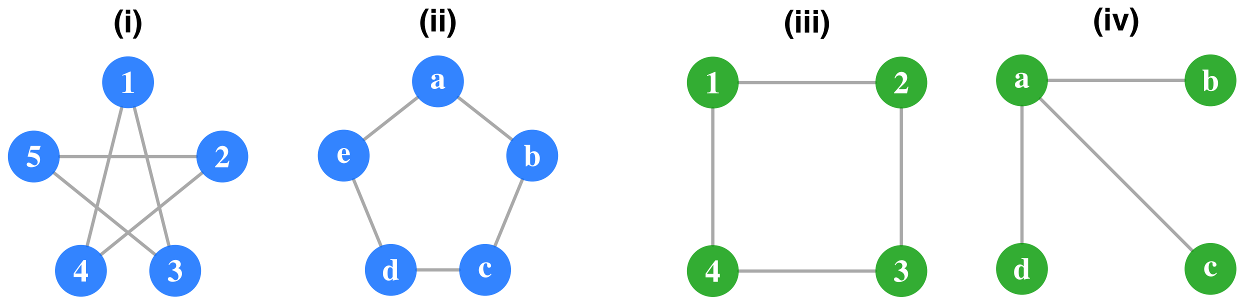


**Figure S1.** Graphs (i) and (ii) are isomorphic. Graphs (iii) and (iv) are not isomorphic.

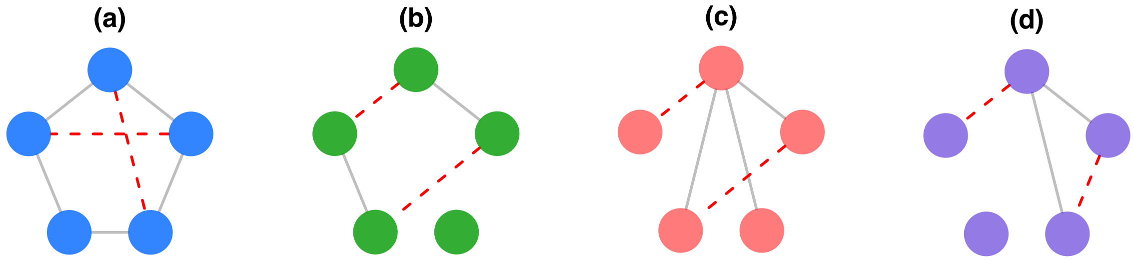


**Figure S2.** Some examples on graph property that are monotone. The gray edges are the original edges and the red dash edges are additional edges added to the existing graph. (a) Graph that is connected. (b) Graph that has no more than three connected components. (c) Graph with maximum degree at least three. (d) Graph with no more than two isolated nodes. Adding the red dash edges to the existing graphs does not change the graph property.

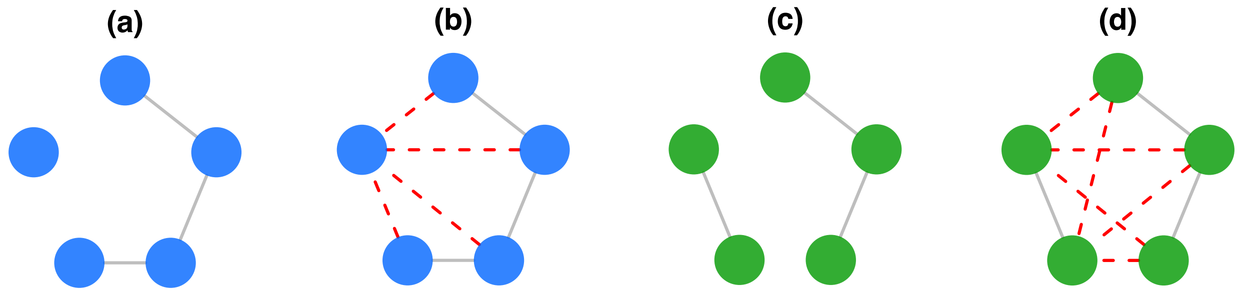


**Figure S3.** Let $\mathcal{P}$ be the graph property of being connected. Gray edges are the original edges of a graph $G$ and the red dash edges are the critical edges that will change the graph property from $\mathcal{P}(G) = 0$ to $\mathcal{P}(G) = 1$. (a) The graph satisfies $\mathcal{P}(G) = 0$. (b) The graph property changes from $\mathcal{P}(G) = 0$ to $\mathcal{P}(G) = 1$ if some red dash edges are added to the graph. (c) The graph satisfies $\mathcal{P}(G) = 0$. (d) The graph property changes from $\mathcal{P}(G) = 0$ to $\mathcal{P}(G) = 1$ if some red dash edges are added to the existing graph.