Supplementary Appendix for “Conditional Functional Graphical Models”

Author A, Author B
Author C, and Author D

S.1 Technical lemmas

We first present two lemmas that connect the operator $\mathfrak{V}^y_{X_iX_j|X_{-(i,j)}}$ with the conditional distribution $(X_i, X_j)^\intercal \mid [X_{-(i,j)}, Y = y]$. For any subvector $\mathsf{A} \subseteq \mathsf{V}$, let $\mathfrak{M}^y_{X_\mathsf{A}|X_{\mathsf{A}^c}} = \left[\mathfrak{V}^y_{X_{\mathsf{A}^c}X_{\mathsf{A}^c}}\right]^\dagger \mathfrak{V}^y_{X_{\mathsf{A}^c}X_\mathsf{A}}$.

Lemma S1 *If Assumptions 1 to 3 hold, then, for any* $(i,j) \in \mathsf{V} \times \mathsf{V}$ *and* $y \in \Omega_Y$,

(i) For any $(f,g) \in \Omega_{X_i} \times \Omega_{X_j}$, $(\langle f, X_i\rangle\, \langle g, X_j\rangle)^\intercal \mid [X_{-(i,j)}, y]$ *follows a bivariate Gaussian distribution, and* $\mathrm{cov}(\langle f, X_i\rangle, \langle g, X_j\rangle \mid X_{-(i,j)}, y) = \langle f, \mathfrak{V}^y_{X_iX_j|X_{-(i,j)}} g\rangle$;

(ii) $\mathfrak{V}^y_{X_iX_j|X_{-(i,j)}} = 0$ *if and only if* $X_i \perp\!\!\!\perp X_j \mid [X_{-(i,j)}, Y = y]$;

(iii) $\mathfrak{V}^y_{X_iX_j|X_{-(i,j)}} = \sum_{a,b\in\mathbb{N}} E(\alpha_i^a \alpha_j^b \mid X_{-(i,j)}, y)\eta_i^a \otimes \eta_j^b$.

Proof. Let $\mathsf{A} = (i,j)$. By Assumption 3, for any $f \in \Omega_{X_\mathsf{A}}$, $g \in \Omega_{X_{\mathsf{A}^c}}$, and $y \in \Omega_Y$,

$$
\begin{aligned}
& E\left(\exp\left\{\iota\left[t_1\left(\langle f, X_\mathsf{A}\rangle - \langle \mathfrak{M}^y_{X_\mathsf{A}|X_{\mathsf{A}^c}} f, X_{\mathsf{A}^c}\rangle\right) + t_2\langle g, X_{\mathsf{A}^c}\rangle\right]\right\} \mid y\right) \\
= \;& t_1^2\langle f, \mathfrak{V}^y_{X_\mathsf{A}X_\mathsf{A}} f\rangle - 2t_1^2\langle f, \mathfrak{V}^y_{X_\mathsf{A}X_{\mathsf{A}^c}}\mathfrak{M}^y_{X_\mathsf{A}|X_{\mathsf{A}^c}} f\rangle + 2t_1t_2\langle f, \mathfrak{V}^y_{X_\mathsf{A}X_{\mathsf{A}^c}} g\rangle \\
& -2t_1t_2\langle g, \mathfrak{V}^y_{X_{\mathsf{A}^c}X_{\mathsf{A}^c}}\mathfrak{M}^y_{X_\mathsf{A}|X_{\mathsf{A}^c}} f\rangle + t_1^2\langle \mathfrak{M}^y_{X_\mathsf{A}|X_{\mathsf{A}^c}} f, \mathfrak{V}^y_{X_{\mathsf{A}^c}X_{\mathsf{A}^c}}\mathfrak{M}^y_{X_\mathsf{A}|X_{\mathsf{A}^c}} f\rangle + t_2^2\langle g, \mathfrak{V}^y_{X_{\mathsf{A}^c}X_{\mathsf{A}^c}} g\rangle \\
= \;& t_1^2\langle f, \mathfrak{V}^y_{X_\mathsf{A}X_\mathsf{A}|X_{\mathsf{A}^c}} f\rangle + t_2^2\langle g, \mathfrak{V}^y_{X_{\mathsf{A}^c}X_{\mathsf{A}^c}} g\rangle.
\end{aligned}
$$

Therefore, $\langle f, X_\mathsf{A}\rangle - \langle \mathfrak{M}^y_{X_\mathsf{A}|X_{\mathsf{A}^c}} f, X_{\mathsf{A}^c}\rangle$ and $\langle g, X_{\mathsf{A}^c}\rangle$ are independent for all $g \in \Omega_{X_{\mathsf{A}^c}}$, which implies assertion (i). Assertion (ii) follows immediately by (i). Assertion (iii) can be shown using an argument similar to the proof of Proposition 4 and (i). This completes the proof. □

Note that Lemma S1 (ii) generalizes the classical result that, under the Gaussian setting, the conditional independence is equivalent to the zero conditional covariance, from the case when X is a vector of random variables to the case when X is a vector of random functions. The next lemma provides an alternative expression of $\mathfrak{V}^y_{X_iX_j|X_{-(i,j)}}$ using $\mathfrak{C}^y_{X_iX_j}$.

Lemma S2 *If Assumptions 1 to 5 hold, then, for any* $(i,j) \in \mathsf{V} \times \mathsf{V}$,

$$\mathfrak{V}^y_{X_iX_j|X_{-(i,j)}} = \mathfrak{V}^y_{X_iX_j} - [\mathfrak{V}^y_{X_iX_i}]^{1/2}\mathfrak{C}^y_{X_iX_{-(i,j)}}[\mathfrak{C}^y_{X_{-(i,j)}X_{-(i,j)}}]^{-1}\mathfrak{C}^y_{X_{-(i,j)}X_j}[\mathfrak{V}^y_{X_jX_j}]^{1/2}.$$

Proof. Denote the right hand side of the above quantity as $\Sigma^y_{X_iX_j|X_{-(i,j)}}$. Then by direct calculation, for any $f \in \Omega_{X_i}$ and $h \in \Omega_{X_{-(i,j)}}$, we have

$$\begin{aligned} E\left[(\langle f, X_i\rangle - \langle h, X_{-(i,j)}\rangle)^2 \mid y\right] = \langle f, \Sigma^y_{X_iX_j|X_{-(i,j)}} f\rangle + \|[\mathfrak{C}^y_{X_{-(i,j)}X_{-(i,j)}}]^{-1/2}\mathfrak{C}^y_{X_{-(i,j)}X_i} \\ \times[\mathfrak{V}^y_{X_iX_i}]^{1/2} f - [\mathfrak{C}^y_{X_{-(i,j)}X_{-(i,j)}}]^{1/2}\mathfrak{D}^y_{X_{-(i,j)}} h\|^2 \equiv V_1(f) + V_2(f,h). \end{aligned}$$

Following a similar proof in Fukumizu et al. (2009, Proposition 2), we have

$$\begin{aligned} V_1(f) &= \inf\left\{E\left[(\langle f, X_i\rangle - \langle h, X_{-(i,j)}\rangle)^2 \mid y\right] : h \in \Omega_{X_{-(i,j)}}\right\} \\ &= E\left[(\langle f, X_i\rangle - \langle h', X_{-(i,j)}\rangle)^2 \mid y\right], \end{aligned}$$

where $h' = \mathfrak{M}^y_{X_i|X_{-(i,j)}} f$, and the last equality holds by Lemma S1(i), and the definition of conditional expectation. This implies that $V_2(f, h') = 0$, and

$$[\mathfrak{C}^y_{X_{-(i,j)}X_{-(i,j)}}]^{-1/2}\mathfrak{C}^y_{X_{-(i,j)}X_i}[\mathfrak{V}^y_{X_iX_i}]^{1/2} f = [\mathfrak{C}^y_{X_{-(i,j)}X_{-(i,j)}}]^{1/2}\mathfrak{D}^y_{X_{-(i,j)}}\mathfrak{M}^y_{X_i|X_{-(i,j)}} f,$$

for all $f \in \Omega_{X_i}$. Denote the above relation as $V_3 f = V_4 f$; therefore,

$$\mathfrak{V}^y_{X_iX_j} + V_3^* V_3 = \Sigma^y_{X_iX_j|X_{-(i,j)}} = \mathfrak{V}^y_{X_iX_j} + V_4^* V_4 = \mathfrak{V}^y_{X_iX_j|X_{-(i,j)}}.$$

This completes the proof. □

We next present a lemma that extends the classical Bernstein inequality to Hilbert spaces. Its proof immediately follows Bosq (2000, Theorem 2.5) and is omitted.

Lemma S3 (Bernstein's inequality in Hilbert space) *Suppose* $U^1, \ldots, U^n$ *are i.i.d. samples from* U *in* Ω_U*, where* U *is a random element with* $E(U) = 0$*, and* Ω_U *is a generic Hilbert space. If, for any* $\ell \in \mathbb{N}$, $E\|U\|^\ell_\Omega \le b^\ell \ell!$*, then, for any* $t > 0$,

$$P\left[\|E_n U\|_{\Omega_U} > t\right] \le 2\exp\left(\frac{-nt^2}{4b^2 + 2bt}\right).$$

Note that the function $f(t) = t^2/(4b^2 + 2bt) > t/(4b)$ if $t > 2b$, and $f(t) \ge t^2/(8b^2)$ if $t \le 2b$. Therefore,

$$P\left[\|E_n U\|_{\Omega_U} > t\right] \leq 2\exp[-cn(t \wedge t^2)], \tag{S1}$$

for some constant $c > 0$. This means, when the moment condition $E\|U\|_\Omega^\ell \leq b^\ell\,\ell!$ holds, the probability of $\{\|E_n U\|_{\Omega_U} > t\}$ behaves as a sub-Gaussian when t is small, and as a sub-Exponential when t is large; see also Hanson and Wright (1971).

The next lemma gives a result on the perturbation of linear operators.

Lemma S4 *Let $\mathfrak{V}$ and $\hat{\mathfrak{V}}$ be the population and sample covariance operators of $U \in \Omega_U$, and $\{(\lambda^a, \eta^a)\}_{a=1}^{\mathbb{N}}$ and $\{(\hat{\lambda}^a, \hat{\eta}^a)\}_{a=1}^{n}$ be their eigenvalue and eigenfunction pairs, with $\lambda^1 > \lambda^2 > \cdots$, and $\hat{\lambda}^1 \geq \hat{\lambda}^2 \geq \cdots \geq \hat{\lambda}^n$. Then $\max_{a=1,\ldots,m} \|\hat{\eta}^a - \eta^a\|_\Omega \leq 4\kappa_m^{-1}\|\hat{\mathfrak{V}} - \mathfrak{V}\|$, where $\kappa_m = \min\{\lambda^a - \lambda^{a+1} : a = 1, \ldots, m+1\}$.*

Proof. Suppose $\{\tilde{\lambda}^1, \ldots, \tilde{\lambda}^n\}$ are the closest members to $\{\lambda^1, \ldots, \lambda^n\}$ in the spectrum of $\hat{\mathfrak{V}}$. Then by Kato (1980, Theorem 4.10), $\max\{|\tilde{\lambda}^a - \lambda^a| : i = 1, \ldots, n\} \leq \|\hat{\mathfrak{V}} - \mathfrak{V}\|$. This implies that $\tilde{\lambda}^a < \lambda^a$, for all $a = m+1, \ldots, n$. Therefore, we have $\tilde{\lambda}^a < \lambda^{m+1} + \|\hat{\mathfrak{V}} - \mathfrak{V}\|$ for all $a = m+1, \ldots, n$. Similarly, we can show that $\tilde{\lambda}^a > \lambda^m - \|\hat{\mathfrak{V}} - \mathfrak{V}\|$ for all $a = 1, \ldots, m$.

When $\kappa_m \leq 2\|\hat{\mathfrak{V}} - \mathfrak{V}\|$, the asserted inequality of this lemma holds.

When $\kappa_m > 2\|\hat{\mathfrak{V}} - \mathfrak{V}\|$, we have $\lambda^m - \|\hat{\mathfrak{V}} - \mathfrak{V}\| > \lambda^{m+1} + \|\hat{\mathfrak{V}} - \mathfrak{V}\|$, which, together with the above bounds for $\tilde{\lambda}^a$, implies that $\min\{\tilde{\lambda}^1, \ldots, \tilde{\lambda}^m\} > \max\{\tilde{\lambda}^{m+1}, \ldots, \tilde{\lambda}^n\}$. Moreover, for any $a = 1, \ldots, m$, $\tilde{\lambda}^{a+1} \leq \lambda^{a+1} + \|\hat{\mathfrak{V}} - \mathfrak{V}\| < \lambda^a - \|\hat{\mathfrak{V}} - \mathfrak{V}\| \leq \tilde{\lambda}^a$, indicating that $\hat{\lambda}^a = \tilde{\lambda}^a$, for all $a = 1, \ldots, m$. Therefore, $\max\{|\hat{\lambda}^a - \lambda^a| : a = 1, \ldots, m\} \leq \|\hat{\mathfrak{V}} - \mathfrak{V}\|$. By Kazdan (1971, Lemma 2), we obtain the asserted inequality, and complete the proof. $\square$

The next lemma shows that the intermediate operator $\mathfrak{P}^y(\epsilon_2)$ and the estimated conditional correlation operator $\hat{\mathfrak{C}}^y_{XX}(d, \epsilon_Y, \epsilon_1)$ are both semi-positive definite.

Lemma S5 *If Assumptions 1 to 5 hold, then $\mathfrak{P}^y(\epsilon_2)$ is semi-positive definite and bounded. Moreover, if $\kappa_Y(y_1, y_2) \geq 0$ for any $y_1, y_2 \in \Omega_Y$, then $\hat{\mathfrak{C}}^y_{XX}(d, \epsilon_Y, \epsilon_1)$ is also semi-positive definite.*

Proof. By Proposition S3, $\mathfrak{C}^y_{XX}$ is positive definite, which implies that $\mathfrak{P}^y(\epsilon_2)$ is semi-positive definite. Moreover, $\mathfrak{P}^y(\epsilon_2)$ is bounded because $\|\mathfrak{P}^y(\epsilon_2)\| \leq \|\mathfrak{P}^y\|$. Also note that, from the proof of Proposition 5, $\hat{\mathfrak{V}}^y_{XX}(d, \epsilon_Y)$ is semi-positive definite. Furthermore, for any $f = (f_1, \ldots, f_p)^\intercal \in \Omega_X$,

$$\left\langle f, \left(\hat{\mathfrak{C}}^y_{XX}(d,\epsilon_Y,\epsilon_1) - \{[\hat{\mathfrak{D}}^y_X(d,\epsilon_Y)]^{\dagger\epsilon_1}\}^{1/2}\hat{\mathfrak{V}}^y_{XX}(d,\epsilon_Y)\{[\hat{\mathfrak{D}}^y_X(d,\epsilon_Y)]^{\dagger\epsilon_1}\}^{1/2}\right) f\right\rangle$$
$$= \epsilon_1 \textstyle\sum_{i=1}^p \langle f_i, [\hat{\mathfrak{V}}^y_{X_iX_i}(d,\epsilon_Y)]^{\dagger\epsilon_1} f_i\rangle \geq 0,$$

where $\hat{\mathfrak{D}}^y_X(d,\epsilon_Y)$ is a block diagonal matrix with $[\hat{\mathfrak{D}}^y_X(d,\epsilon_Y)]_{ii} = \hat{\mathfrak{V}}^y_{X_iX_i}(d,\epsilon_Y)$, for $i \in \mathsf{V}$. This implies $\hat{\mathfrak{C}}^y_{XX}(d,\epsilon_Y,\epsilon_1)$ is semi-positive definite. □

S.2 Proofs

Proof of Proposition 1: For $\mathfrak{V}_{YY}$, note that for any $h_1, h_2 \in \mathcal{H}_Y$, $E\left[h_1(Y)h_2(Y)\right] \leq \|h_1\|\,\|h_2\|\, E\|\kappa_Y(\cdot,Y)\|^2$, which is bounded by $M_Y\|h_1\|\,\|h_2\|$ by Assumption 1. Then the existence and uniqueness of $\mathfrak{V}_{YY}$ are ensured by the Riesz representation theorem.

For $\mathfrak{V}_{X_iX_j}$, because $E\|X_i\|^2_{\Omega_{X_i}}$ is finite by Assumption 1, the existence and uniqueness of $\mathfrak{V}_{X_iX_j}$ can be proved by a similar argument as that for $\mathfrak{V}_{YY}$.

For $\mathfrak{V}_{YX_{ij}}$, note that the expectation $E[\langle X_i, f\rangle\,\langle X_j, g\rangle\, h(Y)]$ is bounded by

$$E|\langle X_i, f\rangle\,\langle X_j, g\rangle\, h(Y)| \leq E^{1/2}\langle h, \kappa_Y(\cdot,Y)\rangle^2 E^{1/2}\left[E^2(|\langle X_i, f\rangle\,\langle X_j, g\rangle| \mid Y)\right]. \tag{S2}$$

Moreover, the conditional expectation $E(|\langle X_i, f\rangle\,\langle X_j, g\rangle| \mid Y) \leq \|f\|\,\|g\|\, E^{1/2}(\|X_i\|^2 \mid Y)\, E^{1/2}(\|X_i\|^2 \mid Y) < M_0^2\|f\|\,\|g\|$ by Assumption 1. This implies the right hand side of (S2) is bounded by $\|h\|\,\|f\|\,\|g\|\, M_0^2\, E^{1/2}\kappa_Y(Y,Y)$. This completes the proof. □

Proof of Proposition 2: By definition, for any $(f,g) \in \Omega_{X_i} \times \Omega_{X_j}$, and $h \in \mathcal{H}_Y$,

$$E[\langle f, X_i\rangle\,\langle g, X_j\rangle\, h(Y)] = \langle \mathfrak{V}_{YY}h, \mathfrak{M}_{X_{ij}|Y}(f\otimes g)\rangle = E[\mathfrak{M}_{X_{ij}|Y}(f\otimes g)\circ(Y)\, h(Y)],$$

implying $E\{[\langle f, X_i\rangle\,\langle g, X_j\rangle - \mathfrak{M}_{X_{ij}|Y}(f\otimes g)\circ(Y)]h(Y)\} = 0$, for $h \in \mathcal{H}_Y$. By the definition of conditional expectation and the fact that $\mathcal{H}_Y$ is dense in $L_2(P_Y)$, the proof is completed. □

Proof of Proposition 4: By Definition 2, we have $\mathrm{range}(\mathfrak{V}^y_{X_iX_j}) \subseteq \mathrm{span}(\{\eta_i^a\}_{a=1}^\infty)$, and $\ker(\mathfrak{V}^y_{X_iX_j})^\perp \supseteq \mathrm{span}(\{\eta_j^b\}_{b=1}^\infty)^\perp$. Therefore,

$$\langle f, \mathfrak{V}^y_{X_iX_j}g\rangle = \textstyle\sum_{a,b\in\mathbb{N}} E(\alpha_i^a\alpha_j^b \mid y)\langle f, \eta_i^a\rangle\,\langle g, \eta_j^b\rangle = \sum_{a,b\in\mathbb{N}} E(\alpha_i^a\alpha_j^b \mid y)\langle f, (\eta_i^a\otimes\eta_j^b)g\rangle,$$

for any $f \in \mathrm{span}(\{\eta_i^a\}_{a=1}^\infty)$ and $g \in \mathrm{span}(\{\eta_j^b\}_{b=1}^\infty)$, where the last equality is by the definition of tensor product. This completes the proof. □

Proof of Theorem 1: Let $\mathfrak{C}^y_{ij|-(i,j)} = \mathfrak{C}^y_{X_i X_{-(i,j)}}[\mathfrak{C}^y_{X_{-(i,j)}X_{-(i,j)}}]^{-1}\mathfrak{C}^y_{X_{-(i,j)}X_j}$. By Lemma S2, we have $\mathfrak{V}^y_{X_iX_j|X_{-(i,j)}} = [\mathfrak{V}^y_{X_iX_i}]^{1/2}[\mathfrak{C}^y_{X_iX_j} - \mathfrak{C}^y_{ij|-(i,j)}][\mathfrak{V}^y_{X_jX_j}]^{1/2}$, implying that

$$\mathfrak{V}^y_{X_iX_j|X_{-(i,j)}} = 0 \quad \Leftrightarrow \quad \mathfrak{C}^y_{X_iX_j} - \mathfrak{C}^y_{ij|-(i,j)} = 0. \tag{S3}$$

It suffices to show that $\mathfrak{C}^y_{X_iX_j} - \mathfrak{C}^y_{ij|-(i,j)} = 0$ if and only if $[\mathfrak{P}^y]_{i,j} = 0$, which, together with (S3) and Lemma S1, imply that $X_i \perp\!\!\!\perp X_j \mid (X_{-(i,j)}, y)$ if and only if $[\mathfrak{P}^y]_{i,j} = 0$.

Let $\mathsf{A} = (i,j)$. Then by rules of matrix inversion,

$$[([\mathfrak{P}^y]_{i,j\in\mathsf{A}})^{-1}]_{1,2} = [([\mathfrak{P}^y]_{i,j\in\mathsf{A}})^{-1}]_{1,1}[\mathfrak{P}^y]_{i,j}([\mathfrak{P}^y]_{j,j})^{-1}.$$

Because $[([\mathfrak{P}^y]_{i,j\in\mathsf{A}})^{-1}]_{1,1} = I - \mathfrak{C}_{ii|-(i,j)} \geq I - \mathfrak{C}_{ii|-i} = ([\mathfrak{P}^y]_{1,1})^{-1}$, which is invertible, we have $\mathfrak{C}^y_{X_iX_j} - \mathfrak{C}^y_{ij|-(i,j)} = 0$ if and only if $[\mathfrak{P}^y]_{i,j} = 0$. This completes the proof. □

Proof of Theorem 2: This is a direct result of Lemma S1 and Proposition 3. □

Proof of Proposition 5: By definition, $\|[\hat{\mathfrak{C}}^y_{XX}(d,\epsilon_Y,\epsilon_1)]_{i,j}\| = \sup\{\|[\hat{\mathfrak{C}}^y_{XX}(d,\epsilon_Y,\epsilon_1)]_{i,j} f\| : f \in \Omega_{X_i}, \|f\| = 1\} = \sup\{\langle g, [\hat{\mathfrak{C}}^y_{XX}(d,\epsilon_Y,\epsilon_1)]_{i,j} f\rangle : f \in \Omega_{X_i}, g \in \Omega_{X_j}, \|f\| = \|g\| = 1\}$, which, by (8), can be computed by solving

$$\begin{aligned} &\max_{f\in\Omega_{X_i}, g\in\Omega_{X_j}} \langle g, \hat{\mathfrak{V}}^y_{X_iX_j}(d,\epsilon_Y) f\rangle \\ &\quad\text{subject to}\quad \langle f, (\hat{\mathfrak{V}}^y_{X_iX_i}(d,\epsilon_Y) + \epsilon_1 I) f\rangle = \langle g, [\hat{\mathfrak{V}}^y_{X_jX_j}(d,\epsilon_Y) + \epsilon_1 I] g\rangle = 1. \end{aligned} \tag{S4}$$

Note that, by (6) and (7), for all $(i,j) \in \mathsf{V}\times\mathsf{V}$, and $a, b = 1, \ldots, d$,

$$[\hat{\mathfrak{M}}_{X_{ij}|Y}(\epsilon_Y)](\hat{\eta}^a_i \otimes \hat{\eta}^b_j) = (\hat{\mathfrak{V}}_{YY} + \epsilon_Y)^{-1} E_n\left[\langle X_i \otimes X_j, \hat{\eta}^a_i \otimes \hat{\eta}^b_j\rangle \kappa_Y(\cdot, Y)\right],$$

which equals to $E_n[\hat{\alpha}^a_i \hat{\alpha}^b_j c(\cdot)]$ with $c(\cdot) = (\hat{\mathfrak{V}}_{YY} + \epsilon_Y)^{-1}\kappa(\cdot, Y)$. Therefore, for any $f \in \text{span}(\{\hat{\eta}^a_i\}^d_{a=1})$ and $g \in \text{span}(\{\hat{\eta}^b_j\}^d_{b=1})$ the three inner products in (S4) are:

$$\begin{aligned} \langle f, \hat{\mathfrak{V}}^y_{X_iX_j}(d,\epsilon_Y) g\rangle &= f_i^\intercal \mathbf{A} D_y \mathbf{B}^\intercal g_j \\ \langle f, (\hat{\mathfrak{V}}^y_{X_iX_i}(d,\epsilon_Y) + \epsilon_1 I) f\rangle &= f_i^\intercal \mathbf{A} D_y \mathbf{A}^\intercal f_i + \epsilon_1 \|f\|^2 \\ \langle g, (\hat{\mathfrak{V}}^y_{X_jX_j}(d,\epsilon_Y) + \epsilon_1 I) g\rangle &= g_j^\intercal \mathbf{B} D_y \mathbf{B}^\intercal g_j + \epsilon_1 \|g\|^2, \end{aligned}$$

where D_y is diagonal with $[D_y]_{kk} = \langle c(\cdot), \kappa_Y(\cdot, Y^k)\rangle$, $k = 1, \ldots, n$, $f_i = (f^1_i, \ldots, f^d_i)^\intercal \in \mathbb{R}^d$ with $f^a_i = \langle f, \hat{\eta}^a_i\rangle$ for $a = 1, \ldots, d$, and $\mathbf{A} \in \mathbb{R}^{d\times n}$ with $[\mathbf{A}]_{s,t} = \hat{\alpha}^{t,s}_i$. Similarly, we define $g_j = (g_j^1, \ldots, g_j^d)^\intercal$, with $g_j^b = \langle g, \hat{\eta}_j^b \rangle$ for $b = 1, \ldots, d$ and $[\mathbf{B}]_{s,t} = \hat{\alpha}_j^{t,s}$. Then by Cauchy-Schwarz inequality, $f_i^\intercal \mathbf{A} D_y \mathbf{B}^\intercal g_j \leq (f_i^\intercal \mathbf{A} D_y \mathbf{A}^\intercal f_i)^{1/2} (g_j^\intercal \mathbf{B} D_y \mathbf{B}^\intercal g_j)^{1/2}$, and thus

$$\langle g, \hat{\mathfrak{V}}^y_{X_i X_j}(d, \epsilon_Y) f \rangle \leq [\langle f, (\hat{\mathfrak{V}}^y_{X_i X_i}(d, \epsilon_Y) + \epsilon_1 I) f \rangle]^{1/2} [\langle g, (\hat{\mathfrak{V}}^y_{X_j X_j}(d, \epsilon_Y) + \epsilon_1 I) g \rangle],$$

which is no greater than 1 by the constraints in (S4). This completes the proof. □

Proof of Proposition 6: The representation of $\hat{\mathfrak{V}}_{YY}$ can be derived following Fukumizu et al. (2009). The representation of $\hat{\mathfrak{V}}_{X_i X_i}$ is

$$\begin{aligned} \lfloor \hat{\mathfrak{V}}_{X_i X_i} \rfloor &= \Big(\lfloor \hat{\mathfrak{V}}_{X_i X_i} \mathcal{B}_1(\cdot) \rfloor, \ldots, \lfloor \hat{\mathfrak{V}}_{X_i X_i} \mathcal{B}_m(\cdot) \rfloor \Big) \\ &= n^{-1} \textstyle\sum_{k=1}^n \left(\langle X_i^k, \mathcal{B}_1(\cdot) \rangle \lfloor X_i \rfloor, \ldots, \langle X_i^k, \mathcal{B}_m(\cdot) \rangle \lfloor X_i \rfloor \right) = E_n \lfloor X_i \rfloor \lfloor X_i \rfloor^\intercal, \end{aligned}$$

where the last equality is because $\langle X_i^k, \mathcal{B}_n(\cdot) \rangle = \lfloor X_i^k \rfloor^\intercal e_j$, where e_j is a vector of size m whose jth element is one and zero otherwise. Similarly we can derive $\lfloor \hat{\mathfrak{V}}_{Y X_{ij}}(f \otimes g) \rfloor$, for $(f, g) \in \Omega^N \times \Omega^N$. This completes the proof. □

Proof of Proposition 7: It suffices to show (14), from which the representations in (15) follow immediately. First, by (6) and Proposition 6, the coordinates of $\hat{\mathfrak{M}}_{X_{ij}|Y}(f \otimes g)$, for all $(f, g) \in \Omega^N \times \Omega^N$, can be written as

$$\lfloor \hat{\mathfrak{M}}_{X_{ij}|Y}(f \otimes g) \rfloor = (K_Y + \epsilon_Y I_n)^{-1} \left[(\lfloor f \rfloor^\intercal \lfloor X_i^1 \rfloor \, \lfloor g \rfloor^\intercal \lfloor X_j^1 \rfloor), \ldots, (\lfloor f \rfloor^\intercal \lfloor X_i^n \rfloor \, \lfloor g \rfloor^\intercal \lfloor X_j^n \rfloor) \right]^\intercal,$$

for each $(i, j) \in \mathsf{V} \times \mathsf{V}$. Therefore, by (13), $[\hat{\mathfrak{M}}_{X_{ij}|Y}(\epsilon_Y)(\hat{\eta}_i^a \otimes \hat{\eta}_j^b)] \circ (y)$ is equal to

$$\mathcal{B}_Y^\intercal(y) \lfloor \hat{\mathfrak{M}}_{X_{ij}|Y}(\epsilon_Y)(\hat{\eta}_i^a \otimes \hat{\eta}_j^b) \rfloor = (\mathbf{a}_i^a)^\intercal \left(\operatorname{diag}\left[(K_Y + \epsilon_Y I_n)^{-1} \mathcal{B}_Y(y) \right] \right) \mathbf{a}_j^b,$$

for any $a, b = 1, \ldots, d$. Then by (7), the proof is completed. □

For notational simplicity, for two sets A, B and two integers A, B, we use $\sum_{a,b}^{\mathsf{A},\mathsf{B}}$ and $\sum_{a,b}^{A,B}$ to abbreviate the double sums $\sum_{a \in \mathsf{A}} \sum_{b \in \mathsf{B}}$ and $\sum_{a=1}^{A} \sum_{b=1}^{B}$, respectively.

Proof of Theorem 3: For (i), by definition, we have,

$$\hat{\mathfrak{V}}_{X_i X_j} - \mathfrak{V}_{X_i X_j} = E_n[X_i \otimes X_j - E(X_i \otimes X_j)].$$

Therefore, by Lemma S3 and (S1), we have the asserted bound if there exists $c > 0$ such that, for all $\ell \in \mathbb{N}$,

$$E\|X_i \otimes X_j - E(X_i \otimes X_j)\|_{\mathrm{HS}}^{\ell} \leq c^{\ell}\, \ell!. \tag{S5}$$

To bound (S5), we note that,

$$\begin{aligned}&E\|X_i \otimes X_j - E(X_i \otimes X_j)\|_{\mathrm{HS}}^{\ell}\\ \leq\, & 2^{\ell-1}\left\{E[E(\|X_i \otimes X_j\|_{\mathrm{HS}}^{\ell} \mid Y)] + E\|E(X_i \otimes X_j \mid Y)\|_{\mathrm{HS}}^{\ell}\right\} \equiv 2^{\ell-1}[E\Lambda_1(Y) + E\Lambda_2(Y)].\end{aligned}$$

Next we show that both $\Lambda_1(y)$ and $\Lambda_2(y)$ can be uniformly bounded for all $y \in \Omega_Y$.

For $\Lambda_1(y)$, when $\ell > 1$, it is equal to

$$E[(\textstyle\sum_{a,b}^{\mathbb{N}_i,\mathbb{N}_j}\langle X_i \otimes X_j, \eta_i^a \otimes \eta_j^b\rangle_{\mathrm{HS}}^2)^{\ell/2} \mid y] \leq M_0^{\ell}\left\{\textstyle\sum_{a,b}^{\mathbb{N}_i,\mathbb{N}_j}(\lambda_i^a\lambda_j^b/M_0^2)E[(\xi_i^a)^2(\xi_j^b)^2 \mid y]^{\ell/2}\right\},$$

where the inequality is by Jensen's inequality, $\mathbb{N}_i = \{a \in \mathbb{N} : \lambda_i^a \neq 0\}$, and $\xi_i^a = (\lambda_i^a)^{-1/2}\alpha_i^a$, for all a, b. Furthermore, when conditioning on Y, $\xi_i^a = (\lambda_i^a)^{-1/2}\alpha_i^a$ is standard normal by Assumption 3. Therefore, due to that $\sum_{a,b}^{\mathbb{N}_i,\mathbb{N}_j}\lambda_i^a\lambda_j^b \leq (\sum_a^{\mathbb{N}}\lambda_i^a)^2$, which equals $E^2\|X_i\|_{\Omega_{X_i}}^2$, and that for all $i \in \mathsf{V}$, $E\|X_i\|_{\Omega_{X_i}}^2 \leq M_0$ by Assumption 1, we have $\Lambda_1(y) \leq (2M_0)^{\ell}\,\ell!$. When $\ell = 1$, for any $y \in \Omega_Y$,

$$\Lambda_1(y) \leq E^{1/2}(\|X_i \otimes X_j\|_{\mathrm{HS}}^2 \mid y) = \{\textstyle\sum_{a,b}^{\mathbb{N}_i,\mathbb{N}_j}\lambda_i^a\lambda_j^b E[(\xi_i^a)^2(\xi_j^b)^2 \mid y]\}^{1/2} \leq M_0.$$

Combining the above bounds, we have, for any $\ell \in \mathbb{N}$ and $y \in \Omega_Y$, $\Lambda_1(y) \leq (2M_0)^{\ell}\,\ell!$.

For $\Lambda_2(y)$, we have

$$\Lambda_2(y) = (\textstyle\sum_{a,b}^{\mathbb{N},\mathbb{N}} E\langle X_i \otimes X_j, \eta_i^a \otimes \eta_j^b\rangle_{\mathrm{HS}}^2)^{\ell/2} \leq \left(\textstyle\sum_{a,b}^{\mathbb{N}_i,\mathbb{N}_j}\lambda_i^a\lambda_j^b E[(\xi_i^a)^2(\xi_j^b)^2 \mid y]\right)^{\ell/2} \leq M_0^{\ell},$$

Combining the bounds for $\Lambda_1(y)$ and $\Lambda_2(y)$ leads to (S5), with $c = 4M_0$.

For (ii), again by definition, we have,

$$\hat{\mathfrak{V}}_{YX_{ij}} - \mathfrak{V}_{YX_{ij}} = E_n[\kappa_Y(\cdot, Y) \otimes X_i \otimes X_j - E(\kappa_Y(\cdot, Y) \otimes X_i \otimes X_j)],$$

which implies that, for any $\ell \in \mathbb{N}$,

$$E\|\kappa_Y(\cdot, Y) \otimes X_i \otimes X_j\|_{\mathrm{HS}}^{\ell} = E[\kappa_Y^{1/2}(Y, Y)\,\|X_i \otimes X_j\|_{\mathrm{HS}}]^{\ell} \leq M_Y^{\ell/2} E\|X_i \otimes X_j\|_{\mathrm{HS}}^{\ell},$$

where the inequality is by Assumption 1. Moreover, by Jensen's inequality,

$$\|E[\kappa_Y(\cdot, Y) \otimes X_i \otimes X_j]\|_{\mathrm{HS}}^{\ell} \leq E^{\ell/2}\|\kappa_Y(\cdot, Y) \otimes X_i \otimes X_j\|_{\mathrm{HS}}^2 \leq M_Y^{\ell/2} E^{\ell/2}\|X_i \otimes X_j\|_{\mathrm{HS}}^2.$$

Then following a similar argument as in the proof of (i), we have

$$E\|\kappa_Y(\cdot,Y)\otimes X_i\otimes X_j - E(\kappa_Y(\cdot,Y)\otimes X_i\otimes X_j)\|_{\mathrm{HS}}^{\ell} \leq (4M_0\sqrt{M_Y})^{\ell}\,\ell!,$$

which, again by Lemma S3, leads to the asserted bound.

For (iii), it can be proved by following the proof in Bach (2009, Proposition 13).

For (iv), we have, for any $t \geq 0$,

$$P\left(\max_{i,j\in\mathsf{V}}\|\hat{\mathfrak{V}}_{YX_{ij}} - \mathfrak{V}_{YX_{ij}}\| > t\right) \leq \sum_{i,j}^{\mathsf{V},\mathsf{V}} P(\|\hat{\mathfrak{V}}_{YX_{ij}} - \mathfrak{V}_{YX_{ij}}\| > t)$$
$$\leq C_3 p^2 \exp[-C_4 n(t\wedge t^2)],$$

where the last inequality is by assertion (i). Hence, by the condition $\log p/n \to 0$ as $n \to 0$, we have,

$$\max_{i,j\in\mathsf{V}}\|\hat{\mathfrak{V}}_{YX_{ij}} - \mathfrak{V}_{YX_{ij}}\| = O_P[(\log p/n)^{1/2}].$$

By similar arguments, we can prove (v) and (vi). □

Proof of Theorem 4: First note that,

$$\max_{i,j\in\mathsf{V}}\|\hat{\mathfrak{V}}^y_{X_iX_j}(d,\epsilon_Y) - \mathfrak{V}^y_{X_iX_j}(d)\|_{\mathrm{HS}} \leq \Lambda_3^y + \Lambda_4^y,$$

where

$$\Lambda_3^y = \max_{i,j\in\mathsf{V}}\|\sum_{a,b}^{d,d}\{[\hat{\mathfrak{M}}_{X_{ij}|Y}(\epsilon_Y)](\hat\eta_i^a\otimes\hat\eta_j^b)\circ(y) - E(\alpha_i^a\alpha_j^b\mid y)\}\,\hat\eta_i^a\otimes\hat\eta_j^b\|_{\mathrm{HS}},$$
$$\Lambda_4^y = \max_{i,j\in\mathsf{V}}\|\sum_{a,b}^{d,d}E(\alpha_i^a\alpha_j^b\mid y)\,[(\hat\eta_i^a-\eta_i^a)\otimes\hat\eta_j^b + \eta_i^a\otimes(\hat\eta_j^b-\eta_j^b)]\|_{\mathrm{HS}}.$$

We next derive the bounds of Λ_3^y and Λ_4^y, respectively.

For Λ_3^y, by Proposition 2, it is further bounded by $\sum_{k=1}^4\Lambda_{3,k}^y$, where

$$\Lambda_{3,1}^y = \max_{i,j\in\mathsf{V}}\|\sum_{a,b}^{d,d}\{[\hat{\mathfrak{M}}_{X_{ij}|Y}(\epsilon_Y) - \mathfrak{M}_{X_{ij}|Y}](\eta_i^a\otimes\eta_j^b)\}\circ(y)\,\hat\eta_i^a\otimes\hat\eta_j^b\|_{\mathrm{HS}},$$
$$\Lambda_{3,2}^y = \max_{i,j\in\mathsf{V}}\|\sum_{a,b}^{d,d}\{[\hat{\mathfrak{M}}_{X_{ij}|Y}(\epsilon_Y)][(\hat\eta_i^a-\eta_i^a)\otimes(\hat\eta_j^b-\eta_j^b)]\}\circ(y)\,\hat\eta_i^a\otimes\hat\eta_j^b\|_{\mathrm{HS}},$$
$$\Lambda_{3,3}^y = \max_{i,j\in\mathsf{V}}\|\sum_{a,b}^{d,d}\{[\hat{\mathfrak{M}}_{X_{ij}|Y}(\epsilon_Y)][\eta_i^a\otimes(\hat\eta_j^b-\eta_j^b)]\}\circ(y)\,\hat\eta_i^a\otimes\hat\eta_j^b\|_{\mathrm{HS}},$$
$$\Lambda_{3,4}^y = \max_{i,j\in\mathsf{V}}\|\sum_{,a,b}^{d,d}\{[\hat{\mathfrak{M}}_{X_{ij}|Y}(\epsilon_Y)][(\hat\eta_i^a-\eta_i^a)\otimes\eta_j^b]\}\circ(y)\,\hat\eta_i^a\otimes\hat\eta_j^b\|_{\mathrm{HS}}.$$

We next derive the bounds of $\Lambda_{3,k}^y$, $k=1,2,3,4$, respectively. For $\Lambda_{3,1}^y$,

$$\Lambda^y_{3,1} \le \max\left\{\textstyle\sum_{a,b}^{d,d} |\langle [\hat{\mathfrak{M}}_{X_{ij}|Y}(\epsilon_Y) - \mathfrak{M}_{X_{ij}|Y}](\eta_i^a \otimes \eta_j^b), \kappa_Y(\cdot, y)\rangle| : i,j \in \mathsf{V}\right\}$$
$$\le d^2 M_Y^{1/2} \times \max_{i,j\in\mathsf{V}} \|\hat{\mathfrak{M}}_{X_{ij}|Y}(\epsilon_Y) - \mathfrak{M}_{X_{ij}|Y}\|,$$

where the first inequality is because $\|\hat{\eta}_i^a\|\,\|\hat{\eta}_i^b\| = 1$, for all i, j, a, b, and the last inequality is by Assumption 1, and that $\|\eta_i^a \otimes \eta_j^b\|_{\mathrm{HS}} = 1$. Moreover,

$$\max_{i,j\in\mathsf{V}} \|\hat{\mathfrak{M}}_{X_{ij}|Y}(\epsilon_Y) - \mathfrak{M}_{X_{ij}|Y}\| \le \max_{i,j\in\mathsf{V}} \|\hat{\mathfrak{V}}_{YY}^{\dagger\epsilon_Y}(\hat{\mathfrak{V}}_{YX_{ij}} - \mathfrak{V}_{YX_{ij}})\| + \max_{i,j\in\mathsf{V}} \|(\hat{\mathfrak{V}}_{YY}^{\dagger\epsilon_Y} - \mathfrak{V}_{YY}^{\dagger\epsilon_Y})\mathfrak{V}_{YX_{ij}}\| + \max_{i,j\in\mathsf{V}} \|\mathfrak{V}_{YY}^{\dagger\epsilon_Y}\mathfrak{V}_{YX_{ij}} - \mathfrak{M}_{X_{ij}|Y}\|. \tag{S6}$$

The first term on the right-hand-side of S6) is no greater than $\epsilon_Y^{-1}\max_{i,j\in\mathsf{V}}\|\hat{\mathfrak{V}}_{YX_{ij}} - \mathfrak{V}_{YX_{ij}}\|$ because $\|\hat{\mathfrak{V}}_{YY}^{\dagger\epsilon_Y}\| \le \epsilon_Y^{-1}$. The second term is bounded by

$$\max_{i,j\in\mathsf{V}} \|(\hat{\mathfrak{V}}_{YY}^{\dagger\epsilon_Y} - \mathfrak{V}_{YY}^{\dagger\epsilon_Y})\mathfrak{V}_{YX_{ij}}\| \le \|\hat{\mathfrak{V}}_{YY}^{\dagger\epsilon_Y}\|\,\|\hat{\mathfrak{V}}_{YY} - \mathfrak{V}_{YY}\| \times \max_{i,j\in\mathsf{V}} \|\mathfrak{V}_{YY}^{\dagger\epsilon_Y}\mathfrak{V}_{YX_{ij}}\|$$
$$\le \epsilon_Y^{-1}\|\hat{\mathfrak{V}}_{YY} - \mathfrak{V}_{YY}\| \times \max_{i,j\in\mathsf{V}} \|\mathfrak{M}_{X_{ij}|Y}\| \le cM_Y^{\beta}\,\epsilon_Y^{-1}\|\hat{\mathfrak{V}}_{YY} - \mathfrak{V}_{YY}\|,$$

where the second inequality is because $\mathfrak{V}_{YX_{ij}} = \mathfrak{V}_{YY}\mathfrak{M}_{X_{ij}|Y}$, and the last inequality is because, by Assumption 6, $\|\mathfrak{M}_{X_{ij}|Y}\| \le \|\mathfrak{V}_{YY}^{\beta}\|\,\|\mathfrak{M}_{ij}^0\|$, and that $\|\mathfrak{V}_{YY}^{\beta}\| \le \|\mathfrak{V}_{YY}\|_{\mathrm{HS}}^{\beta} \le E^{\beta}\|\kappa(\cdot, Y) \otimes \kappa(\cdot, Y)\|_{\mathrm{HS}} \le M_Y^{\beta}$. The last term, by Assumption 6, is bounded by

$$\max_{i,j\in\mathsf{V}} \|\mathfrak{V}_{YY}^{\dagger\epsilon}\mathfrak{V}_{YX_{ij}} - \mathfrak{M}_{X_{ij}|Y}\| = \epsilon_Y \times \max_{i,j\in\mathsf{V}} \|\mathfrak{V}_{YY}^{\dagger\epsilon}\mathfrak{M}_{X_{ij}|Y}\| \le c\,\epsilon_Y^{\beta},$$

where the inequality is because $\|\mathfrak{V}_{YY}^{\dagger\epsilon}\mathfrak{V}_{YY}^{\beta}\| \le \epsilon_Y^{\beta-1}$.

Similarly, for $\Lambda^y_{3,2}$, by Lemma S4, and that $\max_{i,j\in\mathsf{V}}\|\mathfrak{V}_{YX_{ij}}\| \le 2M_0M_y^{1/2}$,

$$\Lambda^y_{3,2} \le \epsilon_Y^{-1}M_Y^{1/2} \times \max_{i,j\in\mathsf{V}}\left\{\textstyle\sum_{a,b}^{d,d}\|\hat{\mathfrak{V}}_{YX_{ij}}[(\hat{\eta}_i^a - \eta_i^a)\otimes(\hat{\eta}_j^b - \eta_j^b)]\|\right\}$$
$$\le 16\epsilon_Y^{-1}d^2\kappa_d^{-2}M_Y^{1/2} \times \max_{i\in\mathsf{V}}\|\hat{\mathfrak{V}}_{X_iX_i} - \mathfrak{V}_{X_iX_i}\| \times \max_{j\in\mathsf{V}}\|\hat{\mathfrak{V}}_{X_jX_j} - \mathfrak{V}_{X_jX_j}\|$$
$$\times\,(\max_{i,j\in\mathsf{V}}\|\hat{\mathfrak{V}}_{YX_{ij}} - \mathfrak{V}_{YX_{ij}}\| + 2M_0M_Y^{1/2}).$$

Furthermore, both $\Lambda^y_{3,3}$ and $\Lambda^y_{3,4}$ are bounded by

$$4\epsilon_Y^{-1}d^2\kappa_d^{-1}M_Y^{1/2} \times \max_{j\in\mathsf{V}}\|\hat{\mathfrak{V}}_{X_jX_j} - \mathfrak{V}_{X_jX_j}\| \times (\max_{i,j\in\mathsf{V}}\|\hat{\mathfrak{V}}_{YX_{ij}} - \mathfrak{V}_{YX_{ij}}\| + 2M_0M_Y^{1/2}).$$

For Λ^y_4, by Proposition 2 and Assumption 6,

$$
\begin{aligned}
\Lambda_4^y &\le 4d^2\kappa_d^{-1}M_Y^{1/2} \times \max_{i,j\in\mathsf{V}}[\|\mathfrak{M}_{X_{ij|Y}}\|(\|\hat{\mathfrak{V}}_{X_iX_i} - \mathfrak{V}_{X_iX_i}\| + \|\hat{\mathfrak{V}}_{X_jX_j} - \mathfrak{V}_{X_jX_j}\|)] \\
&\le 4cM_Y^{1/2+\beta}d^2\kappa_d^{-1}(\max_{i\in\mathsf{V}}\|\hat{\mathfrak{V}}_{X_iX_i} - \mathfrak{V}_{X_iX_i}\| + \max_{j\in\mathsf{V}}\|\hat{\mathfrak{V}}_{X_jX_j} - \mathfrak{V}_{X_jX_j}\|).
\end{aligned}
$$

Combining the bounds of $\Lambda_{3,1}^y$ to $\Lambda_{3,4}^y$, and Λ_4^y, and applying the results in Theorem 3(iv) to (vi), we have

$$
\begin{aligned}
&\max_{i,j\in\mathsf{V}}\|\hat{\mathfrak{V}}_{X_iX_j}^y(d,\epsilon_Y) - \mathfrak{V}_{X_iX_j}^y(d)\|_{\mathrm{HS}} \\
&\quad = O_P\{d^2[\epsilon_Y^{-1}(\log p/n)^{1/2} + \epsilon_Y^{-1}n^{-1/2} + \epsilon_Y^{\beta}]\} + O_P\{d^2\epsilon_Y^{-1}\kappa_d^{-2}[(\log p/n)^{3/2} + \log p/n]\} \\
&\qquad + O_p\{d^2\epsilon_Y^{-1}\kappa_d^{-1}[\log p/n + (\log p/n)^{1/2}]\} + O_P[d^2\kappa_d^{-1}(\log p/n)^{1/2}].
\end{aligned}
$$

By the conditions $\epsilon_Y \prec 1$, $(\log p/n) \prec \kappa_d^2$, we can eliminate the terms with smaller order. This then completes the proof. $\square$

Proof of Lemma 1: We first note that,

$$
\max_{i,j\in\mathsf{V}}\|[\hat{\mathfrak{C}}_{XX}^y(d,\epsilon_Y,\epsilon_1)]_{i,j} - [\mathfrak{C}_{XX}^y(d,\epsilon_1)]_{i,j}\|_{\mathrm{HS}} \le \Lambda_5^y + \Lambda_6^y + \Lambda_7^y,
$$

where

$$
\begin{aligned}
\Lambda_5^y &= \max_{i,j\in\mathsf{V}}\|[(\hat{\Sigma}_1^{\dagger\epsilon_1})^{1/2} - (\Sigma_1^{\dagger\epsilon_1})^{1/2}]\hat{\Sigma}_2(\hat{\Sigma}_3^{\dagger\epsilon_1})^{1/2}\|_{\mathrm{HS}}, \\
\Lambda_6^y &= \max_{i,j\in\mathsf{V}}\|(\Sigma_1^{\dagger\epsilon_1})^{1/2}(\hat{\Sigma}_2 - \Sigma_2)(\hat{\Sigma}_3^{\dagger\epsilon_1})^{1/2}\|_{\mathrm{HS}}, \\
\Lambda_7^y &= \max_{i,j\in\mathsf{V}}\|(\Sigma_1^{\dagger\epsilon_1})^{1/2}\Sigma_2[(\hat{\Sigma}_3^{\dagger\epsilon_1})^{1/2} - (\Sigma_3^{\dagger\epsilon_1})^{1/2}]\|_{\mathrm{HS}},
\end{aligned}
$$

$\Sigma_1 = \mathfrak{V}_{X_iX_i}^y(d), \hat{\Sigma}_1 = \hat{\mathfrak{V}}_{X_iX_i}^y(d,\epsilon_Y), \Sigma_2 = \mathfrak{V}_{X_iX_j}^y(d), \hat{\Sigma}_2 = \hat{\mathfrak{V}}_{X_iX_j}^y(d,\epsilon_Y), \Sigma_3 = \mathfrak{V}_{X_jX_j}^y(d)$, and $\hat{\Sigma}_3 = \hat{\mathfrak{V}}_{X_jX_j}^y(d,\epsilon_Y)$. We next derive the bounds of Λ_5^y, Λ_6^y, and Λ_7^y, respectively.

For Λ_5^y, we have that, for any $(i,j) \in \mathsf{V}\times\mathsf{V}$,

$$
\begin{aligned}
\|[(\hat{\Sigma}_1^{\dagger\epsilon_1})^{1/2} - (\Sigma_1^{\dagger\epsilon_1})^{1/2}]\hat{\Sigma}_2(\hat{\Sigma}_3^{\dagger\epsilon_1})^{1/2}\|_{\mathrm{HS}} \le \Big\{&\|(\Sigma_1^{\dagger\epsilon_1})^{1/2}[(\Sigma_1^{\dagger\epsilon_1})^{-3/2} - (\hat{\Sigma}_1^{\dagger\epsilon_1})^{-3/2}]\hat{\Sigma}_1^{\dagger\epsilon_1}\|_{\mathrm{HS}} \\
&+ \|(\Sigma_1 - \hat{\Sigma}_1)\hat{\Sigma}_1^{\dagger\epsilon_1}\|_{\mathrm{HS}}\Big\} \times \|(\hat{\Sigma}_1^{\dagger\epsilon_1})^{1/2}\hat{\Sigma}_2(\hat{\Sigma}_3^{\dagger\epsilon_1})^{1/2}\|.
\end{aligned} \tag{S7}
$$

There are three norms in (S7). The first norm is bounded by $3\epsilon_1^{-3/2}(\|\hat{\Sigma}_1 - \Sigma_1\| + M_0 + \epsilon_1)\|\hat{\Sigma}_1 - \Sigma_1\|_{\mathrm{HS}}$, because $\|(\Sigma_1^{\dagger\epsilon_1})^{-1/2}\| + \|(\hat{\Sigma}_1^{\dagger\epsilon_1})^{-1/2}\| \le [\|\hat{\Sigma}_1 - \Sigma_1\| + \|(\Sigma_1^{\dagger\epsilon_1})^{-1}\|]^{1/2}$, and that $\|(\Sigma_1^{\dagger\epsilon_1})^{-1}\| = \|\mathfrak{V}_{X_iX_i}^y(d)\| + \epsilon_1 \le \mathrm{tr}(\mathfrak{V}_{X_iX_i}^y) + \epsilon_1 \le M_0 + \epsilon_1$ by Assumption 1. The second norm is in a smaller order than the first norm, and thus can be ignored. The third norm is bounded by 1 by Proposition 5. Therefore,

$$\Lambda_5^y \preceq \epsilon_1^{-3/2}[\max_{i\in\mathsf{V}} \|\hat{\mathfrak{V}}^y_{X_iX_i}(d,\epsilon_Y) - \mathfrak{V}^y_{X_iX_i}(d)\|^2_{\mathrm{HS}} + \max_{i\in\mathsf{V}} \|\hat{\mathfrak{V}}^y_{X_iX_i}(d,\epsilon_Y) - \mathfrak{V}^y_{X_iX_i}(d)\|_{\mathrm{HS}}],$$

whose order of magnitude is $O_P[\epsilon_1^{-3/2}\{d^2\epsilon_Y^{-1}\kappa_d^{-1}(\log p/n)^{1/2} + d^2\epsilon_Y^{\beta}\}]$ by the condition that $d^2\epsilon_Y^{-1}\kappa_d^{-1}(\log p/n)^{1/2} + d^2\epsilon_Y^{\beta} \prec 1$.

For Λ_6^y and Λ_7^y, following a similar argument, we have

$$\Lambda_6^y + \Lambda_7^y = O_P\left[\epsilon_1^{-3/2}\{d^2\epsilon_Y^{-1}\kappa_d^{-1}(\log p/n)^{1/2} + d^2\epsilon_Y^{\beta}\}\right].$$

Combining the bounds of Λ_5^y, Λ_6^y, and Λ_7^y leads to the asserted rate. □

Proof of Lemma 2: Let $[\mathfrak{C}^y_{XX}(\epsilon_1)]_{i,j} = ([\mathfrak{D}^y_X]_{i,i})^{\dagger\epsilon_1}\mathfrak{V}^y_{X_iX_j}([\mathfrak{D}^y_X]_{j,j})^{\dagger\epsilon_1}$ for $(i,j)\in\mathsf{V}\times\mathsf{V}$ with $i\neq j$, and $\max_{i,j\in\mathsf{V}^0}$ be the maximum over all $(i,j)\in\mathsf{V}\times\mathsf{V}$ with $i\neq j$. Then,

$$\max_{i,j\in\mathsf{V}^0}\|[\mathfrak{C}^y_{XX}(d,\epsilon_1)]_{i,j} - [\mathfrak{C}^y_{XX}]_{i,j}\|_{\mathrm{HS}} \le \Lambda_8^y + \Lambda_9^y,$$

where

$$\Lambda_8^y = \max_{i,j\in\mathsf{V}^0}\|[\mathfrak{C}^y_{XX}(d,\epsilon_1)]_{i,j} - [\mathfrak{C}^y_{XX}(\epsilon_1)]_{ij}\|_{\mathrm{HS}},$$
$$\Lambda_9^y = \max_{i,j\in\mathsf{V}^0}\|[\mathfrak{C}^y_{XX}(\epsilon_1)]_{i,j} - [\mathfrak{C}^y_{XX}]_{ij}\|_{\mathrm{HS}}.$$

We next derive the bounds of Λ_8^y and Λ_9^y, respectively.

For Λ_8^y, following a similar argument as the proof in Lemma 1, we have,

$$\Lambda_8^y \preceq \epsilon_1^{-2/3}[\max_{i\in\mathsf{V}}\|\mathfrak{V}^y_{X_iX_i}(d) - \mathfrak{V}^y_{X_iX_i}\|_{\mathrm{HS}} + \max_{j\in\mathsf{V}}\|\mathfrak{V}^y_{X_jX_j}(d) - \mathfrak{V}^y_{X_jX_j}\|_{\mathrm{HS}}] + \epsilon_1^{-1}\max_{i,j\in\mathsf{V}^0}\|\mathfrak{V}^y_{X_iX_j}(d) - \mathfrak{V}^y_{X_iX_j}\|_{\mathrm{HS}}.$$

Moreover, by Assumption 7, we have,

$$\max_{i,j\in\mathsf{V}^0}\|\mathfrak{V}^y_{X_iX_j}(d) - \mathfrak{V}^y_{X_iX_j}\|_{\mathrm{HS}} = \max_{i,j\in\mathsf{V}^0}\|\textstyle\sum_{a,b}^{\mathbb{N}\setminus d,\mathbb{N}\setminus d} E(\alpha_i^a\alpha_j^b \mid y)(\eta_i^a\otimes\eta_j^b)\|_{\mathrm{HS}}$$
$$\le \max_{i,j\in\mathsf{V}^0}\{\textstyle\sum_{a,b}^{\mathbb{N}\setminus d,\mathbb{N}\setminus d} E[(\alpha_i^a)^2\mid y]\,E[(\alpha_j^b)^2\mid y]\}^{1/2} = O(d^{-\gamma_y}).$$

Therefore, $\Lambda_8^y = O(\epsilon_1^{-3/2}d^{-\gamma_y})$.

For Λ_9^y, by direct calculation, we have,

$$(\Lambda_9^y)^2 = \max_{i,j\in\mathsf{V}^0}\|[\mathfrak{C}^y_{XX}(\epsilon_1)]_{i,j} - [\mathfrak{C}^y_{XX}]_{ij}\|^2_{\mathrm{HS}} \le \max_{i,j\in\mathsf{V}^0}\textstyle\sum_{a,b\in\mathbb{N}}[\Delta_{i,j}^{y,a,b}(\rho_{i,j}^{y,a,b})^2],$$

where $\Delta_{i,j}^{y,a,b} = 2\lambda_i^{y,a}\lambda_j^{y,b} + \epsilon_1(\lambda_i^{y,a}+\lambda_j^{y,b}) + \epsilon_1^2 - 2[(\lambda_i^{y,a}\lambda_j^{y,b})(\lambda_i^{y,a}\lambda_j^{y,b} + \epsilon_1(\lambda_i^{y,a}+\lambda_j^{y,b}) + \epsilon_1^2)]^{1/2}$.

It is further bounded by $2\epsilon_1(\max_{i\in\mathsf{V}}\|\mathfrak{V}^y_{X_iX_i}\|)+\epsilon_1^2$. Therefore, by Assumption 4, we have $\Lambda_9^y=O(\epsilon^{1/2})$.

The proof is completed by combining the bounds of Λ_8^y and Λ_9^y. □

Proof of Lemma 3: By Lemma S5, $\|\hat{\mathfrak{P}}^y(d,\epsilon_Y,\epsilon_1,\epsilon_2)-\mathfrak{P}^y(\epsilon_2)\|_{\mathrm{HS}}$ is bounded by

$$\begin{aligned}&\|\hat{\mathfrak{C}}^y_{XX}(d,\epsilon_Y,\epsilon_1)-\mathfrak{C}^y_{XX}\|_{\mathrm{HS}}\,\|\mathfrak{P}^y(\epsilon_2)\|\\&\le\epsilon_2^{-1}\,\|\hat{\mathfrak{C}}^y_{XX}(d,\epsilon_Y,\epsilon_1)-\mathfrak{C}^y_{XX}\|_{\mathrm{HS}}\le\epsilon_2^{-1}p\times\max_{i,j\in\mathsf{V}}\|[\hat{\mathfrak{C}}^y_{XX}(d,\epsilon_Y,\epsilon_1)]_{i,j}-\mathfrak{C}^y_{X_iX_j}\|_{\mathrm{HS}}.\end{aligned}$$

By Theorem 5, we then have the asserted rate. □

Proof of Lemma 4: Note that $[\mathfrak{P}^y]_{\mathsf{A},\mathsf{A}}=(\mathfrak{C}^y_{X_\mathsf{A}X_\mathsf{A}|X_{\mathsf{A}^c}})^{-1}$ with $\mathsf{A}=\{i,j\}$. By the rule of matrix inversion, for any distinct pair $(i,j)\in\mathsf{V}\times\mathsf{V}$,

$$[\mathfrak{P}^y]_{i,j}=-(\mathfrak{C}^y_{X_iX_i|X_{-(i,j)}})^{-1}\mathfrak{C}^y_{X_iX_j|X_{-(i,j)}}(\mathfrak{C}^y_{X_jX_j|X_{-j}})^{-1}\ \equiv-\Psi_1^{-1}\Psi_2\Psi_3^{-1}.$$

Let $\mathfrak{C}^y_{X_iX_j|X_{-(i,j)}}(\epsilon_2)=\mathfrak{C}^y_{X_iX_j}+\epsilon_2\delta_{ij}I-\mathfrak{C}^y_{X_iX_{-(i,j)}}(\mathfrak{C}^y_{X_{-(i,j)}X_{-(i,j)}}+\epsilon_2 I)^{-1}\mathfrak{C}^y_{X_{-(i,j)}X_j}$. Then,

$$\begin{aligned}[\mathfrak{P}^y(\epsilon_2)]_{i,j}&=-[\mathfrak{C}^y_{X_iX_i|X_{-(i,j)}}(\epsilon_2)]^{-1}\mathfrak{C}^y_{X_iX_j|X_{-(i,j)}}(\epsilon_2)[\mathfrak{C}^y_{X_jX_j|X_{-j}}(\epsilon_2)]^{-1}\\&\equiv[\Psi_1(\epsilon_2)]^{-1}[\Psi_2(\epsilon_2)][\Psi_3(\epsilon_2)]^{-1}.\end{aligned}$$

This further implies that

$$\max_{i,j\in\mathsf{V}^0}\|[\mathfrak{P}^y(\epsilon_2)]_{i,j}-\mathfrak{P}^y_{i,j}\|_{\mathrm{HS}}\le\Delta_1+\Delta_2+\Delta_3,$$

where

$$\begin{aligned}\Delta_1&=\max_{i,j\in\mathsf{V}^0}\|\{[\Psi_1(\epsilon_2)]^{-1}-\Psi_1^{-1}\}\Psi_2(\epsilon_2)[\Psi_3(\epsilon_2)]^{-1}\|_{\mathrm{HS}},\\\Delta_2&=\max_{i,j\in\mathsf{V}^0}\|\Psi_1^{-1}[\Psi_2(\epsilon_2)-\Psi_2][\Psi_3(\epsilon_2)]^{-1}\|_{\mathrm{HS}},\\\Delta_3&=\max_{i,j\in\mathsf{V}^0}\|\Psi_1^{-1}\Psi_2\{[\Psi_3(\epsilon_2)]^{-1}-\Psi_3^{-1}\}\|_{\mathrm{HS}}.\end{aligned}$$

We next derive the bounds of Δ_1, Δ_2, and Δ_3, respectively.

For Δ_1, first note that $\|[\Psi_1(\epsilon_2)]^{-1}-\Psi_1^{-1}\|$ is bounded by

$$\begin{aligned}&\|[\Psi_1(\epsilon_2)]^{-1}-\Psi_1^{-1}\|\le\|\Psi_1(\epsilon_2)\|^{-1}\,\|\Psi_1\|^{-1}\\&\quad\times\|\epsilon_2 I-\mathfrak{C}^y_{X_iX_{-(i,j)}}\{[\mathfrak{C}^y_{X_{-(i,j)}X_{-(i,j)}}+\epsilon_2 I]^{-1}-[\mathfrak{C}^y_{X_{-(i,j)}X_{-(i,j)}}]^{-1}\}\mathfrak{C}^y_{X_{-(i,j)}X_i}\|.\end{aligned}\tag{S8}$$

Moreover, because $\mathfrak{C}^y_{X_iX_i|X_{-i}}\le\Psi_1(\epsilon_2)$ and $\|(\mathfrak{C}^y_{X_iX_i|X_{-i}})^{-1}\|=\|[\mathfrak{P}^y]_{i,i}\|\le\|\mathfrak{P}^y\|<\infty$, the first two norms on the right of (S8) are bounded. The third norm is in a smaller order than ϵ_2 because $(\mathfrak{C}^y_{X_{-(i,j)}X_{-(i,j)}})^{-1} - (\mathfrak{C}^y_{X_{-(i,j)}X_{-(i,j)}} + \epsilon_2 I)^{-1} \leq c_{\min}\epsilon(\mathfrak{C}^y_{X_{-(i,j)}X_{-(i,j)}})^{-1}$ by Proposition S3, and that $\mathfrak{C}^y_{X_i X_{-(i,j)}}(\mathfrak{C}^y_{X_{-(i,j)}X_{-(i,j)}})^{-1}\mathfrak{C}^y_{X_{-(i,j)}X_i} \leq I$. Therefore, we have $\max_{i,j\in\mathsf{V}^0}\|[\Psi_1(\epsilon_2)]^{-1} - \Psi_1^{-1}\| \preceq \epsilon_2$. Because $\Psi_3(\epsilon_2) \geq \mathfrak{C}^y_{X_j X_j|X_{-j}}$, $\|\Phi_3(\epsilon_2)\|$ is also bounded. Moreover,

$$\max_{i,j\in\mathsf{V}^0}\|\mathfrak{C}^y_{X_i X_{-(i,j)}}[\mathfrak{C}^y_{X_{-(i,j)}X_{-(i,j)}} + \epsilon_2 I]^{-1}\mathfrak{C}^y_{X_{-(i,j)}X_j}\|_{\text{HS}}$$
$$\leq \max_{i,j\in\mathsf{V}^0}\|\mathfrak{C}^y_{X_i X_{-(i,j)}}[\mathfrak{C}^y_{X_{-(i,j)}X_{-(i,j)}}]^{-1}\mathfrak{C}^y_{X_{-(i,j)}X_j}\|_{\text{HS}},$$

which is finite by Proposition S2 and Assumption 8. This implies $\|\Psi_2(\epsilon_2)\|_{\text{HS}}$ is uniformly bounded. Therefore, we have $\Delta_1 \preceq \epsilon_2$.

For Δ_2, it suffices to show that $\|\Psi_2(\epsilon_2) - \Psi_2\|_{\text{HS}}$ is uniformly bounded. Following a similar argument as that for Δ_1, we obtain that,

$$\max_{i,j\in\mathsf{V}^0}\|\Psi_2(\epsilon_2) - \Psi_2\|_{\text{HS}} \leq c_{\min}\epsilon_2 \times \max_{i,j\in\mathsf{V}^0}\|\mathfrak{C}^y_{X_i X_{-(i,j)}}[\mathfrak{C}^y_{X_{-(i,j)}X_{-(i,j)}}]^{-1}\mathfrak{C}^y_{X_{-(i,j)}X_j}\|_{\text{HS}} \preceq \epsilon_2,$$

which implies that $\Delta_2 \preceq \epsilon_2$.

For Δ_3, we can similarly show that $\Delta_3 \preceq \epsilon_2$.

The proof is completed by combining the orders of Δ_1, Δ_2, and Δ_3. □

Proof of Theorem 6: The first assertion follows Lemmas 3 and 4. For the second assertion, we have, by Lemma S1, $\mathsf{E}^y = \{(i,j) : i \neq j,\ \|[\mathfrak{P}^y]_{i,j}\|_{\text{HS}} > 0\}$. Furthermore,

$$\begin{aligned}
&P\left[\hat{\mathsf{E}}^y_{\text{CPO}}(d, \epsilon_Y, \epsilon_1, \epsilon_2, \rho_{\text{CPO}}) \neq \mathsf{E}^y\right] \\
&\leq P\left\{\|[\hat{\mathfrak{P}}^y(d, \epsilon_Y, \epsilon_1, \epsilon_2)]_{i,j} > \rho_{\text{CPO}} \text{ and } [\mathfrak{P}^y]_{i,j} = 0, \text{ for some } i, j \in \mathsf{V}\right\} \\
&\quad + P\left\{\|[\hat{\mathfrak{P}}^y(d, \epsilon_Y, \epsilon_1, \epsilon_2)]_{i,j} \leq \rho_{\text{CPO}} \text{ and } [\mathfrak{P}^y]_{i,j} \neq 0, \text{ for some } i, j \in \mathsf{V}\right\},
\end{aligned}$$

where both terms are bounded by $P\{\max_{i,j\in\mathsf{V}^0}\|[\hat{\mathfrak{P}}^y(d, \epsilon_Y, \epsilon_1, \epsilon_2)]_{i,j} - [\mathfrak{P}^y]_{i,j}\|_{\text{HS}} \geq \rho_{\text{CPO}}\}$, which tends to zero as $n \to \infty$ by the condition that $\rho_{\text{CPO}} \succ [\epsilon_2 + \epsilon_2^{-1} p\delta_y]$. This completes the proof. □

S.3 Discussion of Assumptions 4 and 5

We discuss Assumptions 4 and 5 in more detail. Assumption 4 characterizes the level of smoothness for the underlying distributions of the random functions. We first note that the quantity $\sum_{a,b}^{\mathbb{N}_i^y,\mathbb{N}_j^y}(\rho_{i,j}^{y,a,b})^2$ in (5) is zero if and only if $\rho_{i,j}^{y,a,b}=0$, for $a\in\mathbb{N}_i^y$ and $b\in\mathbb{N}_j^y$, which is equivalent to the conditional independence between X_i and X_j given Y. We next provide an equivalent condition of Assumption 4. Its proof immediately follows by the definition of conditional correlation operator and is omitted.

Proposition S1 *Suppose Assumptions 1 and 2 hold. For each* $(i,j)\in\mathsf{V}\times\mathsf{V}$, $i\neq j$, *and* $y\in\Omega_Y$, *if* $\mathfrak{C}_{i,j}^{y,0}=\sum_{a,b\in\mathbb{N}}\rho^{y,a,b}(\eta_i^{y,a}\otimes\eta_j^{y,b})$, *then*

(i) $\mathfrak{C}_{X_iX_j}^y=(\mathfrak{V}_{X_iX_i}^y)^{1/2}\mathfrak{C}_{i,j}^{y,0}(\mathfrak{V}_{X_jX_j}^y)^{1/2}$;

(ii) (5) holds if and only if $\max_{i,j\in\mathsf{V},i\neq j}\|\mathfrak{C}_{i,j}^{y,0}\|_{\mathrm{HS}}^2\leq c_1$.

We note that, in the context of unconditional functional graphical model, Li and Solea (2018, Assumption 4) has introduced a similar condition,

$$\mathfrak{C}_{X_iX_j}=(\mathfrak{V}_{X_iX_i})^{1/2+\beta}\mathfrak{C}_{i,j}^{0}(\mathfrak{V}_{X_jX_j})^{1/2+\beta},\tag{S9}$$

where $\beta>0$, and $\mathfrak{C}_{X_iX_j}$, $\mathfrak{V}_{X_iX_i}$, $\mathfrak{C}_{i,j}^0$, and $\mathfrak{V}_{X_jX_j}$ are the unconditional counterparts of $\mathfrak{C}_{X_iX_j}^y$, $\mathfrak{V}_{X_iX_i}^y$, $\mathfrak{C}_{i,j}^{y,0}$, and $\mathfrak{V}_{X_jX_j}^y$, respectively. Proposition S1(i) and condition (S9) are imposed in a similar fashion. However, Assumption 4 is more transparent than (S9), because it is based on the variances and covariances of the eigenfunctions of the conditional covariance operators.

To provide some further insight to Assumption 4, let $\tau_{i,j}^{y,a,b}=\mathrm{var}^{-1/2}(\langle\eta_i^{y,a},X_i\rangle\mid y)\,\mathrm{cov}(\langle\eta_i^{y,a},X_i\rangle,\langle\eta_j^{y,b},X_j\rangle\mid y)\,\mathrm{var}^{-1/2}(\langle\eta_i^{y,a},X_i\rangle\mid y)$, which is the correlation between $\langle\eta_i^{y,a},X_i\rangle$ and $\langle\eta_j^{y,a},X_j\rangle$ given $Y=y$. Assumption 4 then implies that both quantities $\sum_{a,b\in\mathbb{N}}(\tau_{i,j}^{y,a,b})^2(\lambda_i^{y,a})^{-1}(\lambda_j^{y,a})^{-1}$ and $\sum_{a\in\mathbb{N}}\lambda_i^{y,a}$ need to be uniformly bounded. For example, if $\lambda_i^{y,a}\asymp a^{-\alpha}$ and $\tau_{i,j}^{y,a,b}\asymp(ab)^{-(\beta+\alpha)/2}$ with $\alpha,\beta>1$, then Assumption 4 holds. Note that because $\lambda_i^{y,a}$ vanishes fast, $\tau_{i,j}^{y,a,b}$ needs to vanish faster, which implies that the conditional dependency between X_i and X_j given Y needs to be adequately concentrated on the leading eigenfunctions of $\mathfrak{V}_{X_iX_i}^y$ and $\mathfrak{V}_{X_jX_j}^y$.

Under Assumption 4, the next proposition shows that $\mathfrak{C}_{X_iX_j}^y$ is Hilbert-Schmidt, and thus it is compact.

Proposition S2 *If Assumptions 1 to 4 hold, then there exists a constant* c_{HS}, *such that* $\max_{i,j\in\mathsf{V},i\neq j}\|\mathfrak{C}_{X_iX_j}^y\|_{\mathrm{HS}}\leq c_{\mathrm{HS}}$.

PROOF. By definition, $\|\mathfrak{C}^y_{X_iX_j}\|^2_{\rm HS}$ is equal to

$$\sum_{a,b}^{\mathbb{N}_i^y,\mathbb{N}_j^y}\langle\tilde\eta_i^{y,a},\mathfrak{V}^y_{X_iX_j}\tilde\eta_j^{y,b}\rangle^2=\sum_{a,b}^{\mathbb{N}_i^y,\mathbb{N}_j^y}E^2(\langle\tilde\eta_i^{y,a},X_i\rangle\,\langle\tilde\eta_j^{y,b},X_j\rangle\mid y),$$

where $\tilde\eta_i^{y,a}=\eta_i^{y,a}/(\lambda_i^{y,a})^{1/2}$ for any $y\in\Omega_Y$ and $(i,j)\in\mathsf{V}\times\mathsf{V}$. By Assumptions 3, the right-hand side of the above quantity is further bounded by

$$\sum_{a,b}^{\mathbb{N}_i^y,\mathbb{N}_j^y}\mathrm{cov}^2(\langle\tilde\eta_i^{y,a},X_i\rangle\,\langle\tilde\eta_j^{y,b},X_j\rangle\mid y)=\sum_{a,b}^{\mathbb{N}_i^y,\mathbb{N}_j^y}\left[\lambda_i^{y,a}\lambda_j^{y,b}(\rho_{i,j}^{y,a,b})^2\right]\le c_1(\textstyle\sum_{a,b\in\mathbb{N}}\lambda_i^{y,a}\lambda_j^{y,b}),$$

where $(\rho_{i,j}^{y,a,b})^2$ is defined in (5), and the last inequality is by Assumption 4. Note that the last term in the above relation is $c_1(\sum_{a\in\mathbb{N}}\lambda_i^{y,a})^2=c_1\mathrm{tr}^2(\mathfrak{V}^y_{X_iX_i})$ which is no greater than $c_1M_0^2$ by Assumption 1. □

Assumption 5 is to prevent the existence of a constant function consisting of a linear combination of non-constant functions. To see this, for $f=(f_1,\ldots,f_p)^\intercal\in\Omega_X$, we have that, $\ker(\mathfrak{V}^y_{XX})=\{f\in\Omega_X:E[(\sum_{i=1}^p\langle f_i,X_i\rangle_{\Omega_{X_i}})^2\mid Y]=0\}$, which is further equal to $\{f\in\Omega_X:\sum_{i=1}^p\langle f_i,X_i\rangle_{\Omega_{X_i}}=0$ almost surely$\}$.

Under Assumptions 4 and 5, the next proposition shows that $\mathfrak{C}^y_{XX}$ is lower bounded by a strictly positive constant, which immediately implies that $\mathfrak{C}^y_{XX}$ is invertible.

Proposition S3 *If Assumptions 1 to 5 hold, then there exists $c_{\min}>0$, such that $c_{\min}I\le\mathfrak{C}^y_{XX}$, where I is the identity mapping.*

PROOF. Note that $\mathfrak{C}^y_{XX}$ can be expressed as $\mathfrak{C}^y_{XX}=I+\mathfrak{C}'$, with $\mathfrak{C}'$ being a compact operator. Therefore, by Bach (2008), if $\mathfrak{C}^y_{XX}$ is invertible, then there exists $c>0$ such that $\mathfrak{C}^y_{XX}$ must be bounded below by cI. Moreover, by Assumption 5, $\mathfrak{V}^y_{XX}$ is invertible, which implies that $\mathfrak{C}^y_{XX}$ is also invertible. This completes the proof. □

S.4 Estimation via conditional partial correlation operator

As an alternative approach, we briefly discuss how to estimate the graph E^y via the conditional partial correlation operator $\mathfrak{R}^y_{X_iX_j|X_{-(i,j)}}$. First, we estimate $\mathfrak{V}^y_{X_iX_j|X_{-(i,j)}}$ by, for each $(i,j)\in\mathsf{V}\times\mathsf{V}$,

$$\begin{aligned}\hat{\mathfrak{V}}^y_{X_iX_j|X_{-(i,j)}}(d,\epsilon_Y,\epsilon_3)=&\hat{\mathfrak{V}}^y_{X_iX_j}(d,\epsilon_Y)-\hat{\mathfrak{V}}^y_{X_iX_{-(i,j)}}(d,\epsilon_Y)\\&\times\left[\hat{\mathfrak{V}}^y_{X_{-(i,j)}X_{-(i,j)}}(d,\epsilon_Y)+\epsilon_3I\right]^{-1}\hat{\mathfrak{V}}^y_{X_{-(i,j)}X_j}(d,\epsilon_Y),\end{aligned}$$

where $\epsilon_3 > 0$ is the ridge parameter. By Proposition 3, we estimate $\mathfrak{R}^y_{X_iX_j|X_{-(i,j)}}$ by

$$\hat{\mathfrak{R}}^y_{X_iX_j|X_{-(i,j)}}(d,\epsilon_Y,\epsilon_3,\epsilon_4) = (\hat{\mathfrak{V}}^y_{X_iX_i|X_{-(i,j)}}(d,\epsilon_Y,\epsilon_3) + \epsilon_4 I)^{-1/2} \times \hat{\mathfrak{V}}^y_{X_iX_j|X_{-(i,j)}}(d,\epsilon_Y,\epsilon_3)(\hat{\mathfrak{V}}^y_{X_iX_i|X_{-(i,j)}}(d,\epsilon_Y,\epsilon_3) + \epsilon_4 I)^{-1/2},$$

where $\epsilon_4 > 0$ is a ridge parameter to enable the inversion of $\hat{\mathfrak{V}}^y_{X_iX_i|X_{-(i,j)}}(d,\epsilon_Y,\epsilon_3)$ and $\hat{\mathfrak{V}}^y_{X_jX_j|X_{-(i,j)}}(d,\epsilon_Y,\epsilon_3)$. Then, for each $y \in \Omega_Y$, we can estimate the graph E^y by

$$\hat{\mathsf{E}}^y_{\text{CPCO}}(d,\epsilon_Y,\epsilon_3,\epsilon_4,\rho_{\text{CPCO}}) = \{(i,j) \in \mathsf{V}\times\mathsf{V} \colon \|\hat{\mathfrak{R}}^y_{X_iX_j|X_{-(i,j)}}(d,\epsilon_Y,\epsilon_3,\epsilon_4)\|_{\text{HS}} > \rho_{\text{CPCO}}, i \neq j\},$$

where $\rho_{\text{CPCO}} > 0$ is the thresholding parameter.

At the sample level, we estimate the coordinates of $\hat{\mathfrak{R}}_{X_iX_j|X_{-(i,j)}|Y}$ as

$$\begin{aligned}
{}_{\mathcal{B}_i^*}\lfloor \hat{\mathfrak{V}}^y_{X_iX_j|X_{-(i,j)}}(d,\epsilon_Y,\epsilon_3)\rfloor_{\mathcal{B}_j^*} &= [M(y)]_{i,j} - [M(y)]_{i,-(i,j)} \times \\
&\quad \big([M(y)]_{-(i,j),-(i,j)} + \epsilon_3 I_{(p-2)d\times(p-2)d}\big)^{-1}[M(y)]_{-(i,j),j} \equiv N_{i,j|-(i,j)}(y), \\
{}_{\mathcal{B}_i^*}\lfloor \hat{\mathfrak{R}}^y_{X_iX_j|X_{-(i,j)}}(d,\epsilon_Y,\epsilon_3,\epsilon_4)\rfloor_{\mathcal{B}_j^*} &= [N_{i,i|-(i,j)}(y) + \epsilon_4 I_{d\times d}]^{-1/2}\, N_{i,j|-(i,j)}(y) \times \\
&\quad [N_{j,j|-(i,j)}(y) + \epsilon_4 I_{d\times d}]^{-1/2},
\end{aligned}$$

where $M(y) = {}_{\mathcal{B}^*}\lfloor \hat{\mathfrak{V}}^y_{XX}(d,\epsilon_Y)\rfloor_{\mathcal{B}^*}$, and its (i,j)th block $[M(y)]_{i,j}$ is of dimension $d \times d$.

S.5 Additional sparsity structure

Recall in Theorem 6, our CPO estimator depends on the rate of $p(\log p/n^{1-\pi-\pi'})^{1/2}$. This means that the graph dimension p can only grow at a polynomial rate of the sample size n. This is partly because we did *not* impose any sparsity structure, but only required the threshold ρ_{CPO} approaches zero at the same rate as the minimum signal strength. Next, we consider two explicit sparsity structures, one on the CPO and the other on the CCO. We show that, with such additional sparsity assumptions and some regularized estimation such as hard thresholding, we can further improve the rate in Theorem 6, so that p can grow at an exponential rate of n.

The first sparsity structure we consider is explicitly placed on the CPO, by restricting the number of nonzero elements on the off-diagonal elements of the CPO.

Assumption S1 *For $y \in \Omega_Y$, there exists $s_y \in \mathbb{N}$ such that*

$$\text{card}\big(\{(i,j) : [\mathfrak{P}^y]_{i,j} \neq 0, i \neq j\}\big) = s_y.$$

Although Assumption S1 imposes the sparsity explicitly on the CPO, it indirectly restricts the number of nonzero elements of its inverse, i.e., the CCO, as shown by the next proposition.

Proposition S4 *If Assumptions 1, 3, 4, and S1 hold, then, for each $y \in \Omega_Y$,*

$$\operatorname{card}\big(\{(i,j) : [\mathfrak{C}^y_{XX}]_{i,j} \neq 0, i \neq j\}\big) \leq s_y(s_y - 1).$$

PROOF. For $y \in \Omega_Y$, by Assumption S1 there are at most s^y columns in $\mathfrak{P}^y$ having at least one nonzero elements, which also implies the remaining columns all have zeros on their off-diagonal elements. Let $\mathsf{A} = \{i \in \mathsf{V} : [\mathfrak{P}^y]_{i,-i} = 0\}$ index those remaining columns. Then by Definition 3 and the matrix inversion rule, for $i \in \mathsf{A}$,

$$[\mathfrak{C}^y_{XX}]_{i,-i} = -[\mathfrak{P}^y]_{i,-i}\{[\mathfrak{P}^y]_{-i,-i}\}^{-1} = 0.$$

Moreover, for distinct $i, j \in \mathsf{V} \times \mathsf{V}$, $[\mathfrak{C}^y_{XX}]_{i,j} = 0 \Leftrightarrow X_i \perp\!\!\!\perp X_j$. This implies there are at least $p - s^y$ random functions, each independent with the rest of the functions. Therefore, there are at most $s^y(s^y - 1)$ nonzero off-diagonal elements in $\mathfrak{C}^y_{XX}$. □

Proposition S4 suggests the number of nonzero off-diagonal elements in the CCO is of the order s_y^2. This means, when $s_y \ll p$, the majority of the off-diagonal elements in $\mathfrak{C}^y_{XX}$ are zero. To take advantage of this sparsity structure, we consider a hard thresholding regularization to estimate the CCO, then the CPO and the graph,

$$\begin{aligned}
[\check{\mathfrak{C}}^y_{XX}(d,\epsilon_Y,\epsilon_1,\zeta)]_{i,j} &= [\hat{\mathfrak{C}}^y_{XX}(d,\epsilon_Y,\epsilon_1)]_{i,j}\, 1\big(\|[\hat{\mathfrak{C}}^y_{XX}(d,\epsilon_Y,\epsilon_1)]_{i,j}\|_{\mathrm{HS}} \geq \zeta\big),\\
\check{\mathfrak{P}}^y(d,\epsilon_Y,\epsilon_1,\epsilon_2,\zeta) &= \{\check{\mathfrak{C}}^y_{XX}(d,\epsilon_Y,\epsilon_1,\zeta) + \epsilon_2 I\}^{-1}, \qquad\qquad \text{(S10)}\\
\check{\mathsf{E}}^y_{\mathrm{CPO}}(d,\epsilon_Y,\epsilon_1,\epsilon_2,\zeta,\rho_{\mathrm{CPO}}) &= \{(i,j) \in \mathsf{V}\times\mathsf{V} : \|[\check{\mathfrak{P}}^y(d,\epsilon_Y,\epsilon_1,\epsilon_2,\zeta)]_{i,j}\|_{\mathrm{HS}} > \rho_{\mathrm{CPO}}, i \neq j\}.
\end{aligned}$$

where ζ is the thresholding parameter. The next theorem establishes the consistency of the estimators in (S10).

Theorem S1 *If Assumptions 1 to 8, and S1 hold, $\epsilon_Y, \epsilon_1 \prec 1$, $d^2\epsilon_Y^{-1}\kappa_d^{-1}\,(\log p/n)^{1/2} + d^2\epsilon_Y^{\beta} \prec 1$, $\delta_y \prec \zeta \prec 1$, and $(\epsilon_2^{-1}\zeta s_y^2 + \epsilon_2) \prec \rho_{\mathrm{CPO}}$, then, for any $y \in \Omega_Y$,*

$$\begin{gathered}
\|\check{\mathfrak{C}}^y_{XX}(d,\epsilon_Y,\epsilon_1,\zeta) - \mathfrak{C}^y_{XX}\|_{\mathrm{HS}} = O_p(\zeta s_y^2);\\
\max_{i,j\in\mathsf{V}, i\neq j} \|[\check{\mathfrak{P}}^y(d,\epsilon_Y,\epsilon_1,\epsilon_2,\zeta)]_{i,j} - [\mathfrak{P}^y]_{i,j}\|_{\mathrm{HS}} = O_p(\epsilon_2^{-1}\zeta s_y^2 + \epsilon_2);\\
P[\check{\mathsf{E}}^y_{\mathrm{CPO}}(d,\epsilon_Y,\epsilon_1,\epsilon_2,\zeta,\rho_{\mathrm{CPO}}) = \mathsf{E}^y] \to 1, \text{ as } n \to \infty.
\end{gathered}$$

PROOF. For any $\zeta > 0$, define $\mathfrak{C}^{y,\zeta}_{XX} : \Omega_X \to \Omega_X$ as,

$$(\mathfrak{C}^{y,\zeta}_{XX})_{i,j} = [\mathfrak{C}^y_{XX}]_{i,j}\, 1(\|[\mathfrak{C}^y_{XX}]_{i,j}\|_{\mathrm{HS}} \geq \zeta), \quad (i,j) \in \mathsf{V} \times \mathsf{V}.$$

Note that $\mathfrak{C}^{y,\zeta}_{XX}$ is an intermediate operator between $\check{\mathfrak{C}}^y_{XX}(d, \epsilon_Y, \epsilon_1, \zeta)$ and $\mathfrak{C}^y_{XX}$. For simplicity, write $\hat{\mathfrak{C}}^y_{XX}(d, \epsilon_Y, \epsilon_1)$ as $\hat{\mathfrak{C}}^y_{XX}$, $\hat{\mathfrak{C}}^y_{X_iX_j}(d, \epsilon_Y, \epsilon_1)$ as $\hat{\mathfrak{C}}^y_{X_iX_j}$, and $\check{\mathfrak{C}}^y_{XX}(d, \epsilon_Y, \epsilon_1, \zeta)$ as $\check{\mathfrak{C}}^y_{XX}$. By the triangular inequality

$$\|\check{\mathfrak{C}}^y_{XX} - \mathfrak{C}^y_{XX}\|_{\mathrm{HS}} \leq \Delta_4 + \Delta_5,$$

where $\Delta_4 = \|\check{\mathfrak{C}}^y_{XX} - \mathfrak{C}^{y,\zeta}_{XX}\|_{\mathrm{HS}}$, and $\Delta_5 = \|\mathfrak{C}^{y,\zeta}_{XX} - \mathfrak{C}^y_{XX}\|_{\mathrm{HS}}$. Next, we derive the orders of magnitude of Δ_4 and Δ_5, respectively.

For Δ_5, we have,

$$\Delta_5 \leq \textstyle\sum_{i,j \in \mathsf{V}^0} \|\mathfrak{C}^y_{X_iX_j}\|_{\mathrm{HS}}\, 1(\|\mathfrak{C}^y_{X_iX_j}\|_{\mathrm{HS}} < \zeta) \leq \zeta \textstyle\sum_{i,j \in \mathsf{V}^0} 1(\|\mathfrak{C}^y_{X_iX_j}\|_{\mathrm{HS}} \neq 0),$$

which by Proposition S4, is further bounded by $\zeta(s^y)^2$.

For Δ_4, we have

$$\begin{aligned}\Delta_4 \leq &\textstyle\sum_{i,j \in \mathsf{V}^0} \|\hat{\mathfrak{C}}^y_{X_iX_j} - \mathfrak{C}^y_{X_iX_j}\|_{\mathrm{HS}}\, 1(\|\hat{\mathfrak{C}}^y_{X_iX_j}\|_{\mathrm{HS}} \geq \zeta, \|\mathfrak{C}^y_{X_iX_j}\|_{\mathrm{HS}} \geq \zeta) \\ &+ \textstyle\sum_{i,j \in \mathsf{V}^0} \|\hat{\mathfrak{C}}^y_{X_iX_j}\|_{\mathrm{HS}}\, 1(\|\hat{\mathfrak{C}}^y_{X_iX_j}\|_{\mathrm{HS}} \geq \zeta, \|\mathfrak{C}^y_{X_iX_j}\|_{\mathrm{HS}} < \zeta) \\ &+ \textstyle\sum_{i,j \in \mathsf{V}^0} \|\mathfrak{C}^y_{X_iX_j}\|_{\mathrm{HS}}\, 1(\|\hat{\mathfrak{C}}^y_{X_iX_j}\|_{\mathrm{HS}} < \zeta, \|\mathfrak{C}^y_{X_iX_j}\|_{\mathrm{HS}} \geq \zeta) = \Delta_{41} + \Delta_{42} + \Delta_{43}.\end{aligned}$$

We next find the orders of Δ_{41} to Δ_{43}.

For Δ_{41}, we have,

$$\begin{aligned}\Delta_{41} \leq &\max_{i,j \in \mathsf{V}^0} \|\hat{\mathfrak{C}}^y_{X_iX_j} - \mathfrak{C}^y_{X_iX_j}\|_{\mathrm{HS}} \,[\textstyle\sum_{i,j \in \mathsf{V}^0} 1(\|\hat{\mathfrak{C}}^y_{X_iX_j}\|_{\mathrm{HS}} \geq \zeta, \|\mathfrak{C}^y_{X_iX_j}\|_{\mathrm{HS}} \geq \zeta)] \\ \leq &\max_{i,j \in \mathsf{V}^0} \|\hat{\mathfrak{C}}^y_{X_iX_j} - \mathfrak{C}^y_{X_iX_j}\|_{\mathrm{HS}} \,[\textstyle\sum_{i,j \in \mathsf{V}^0} 1(\|\mathfrak{C}^y_{X_iX_j}\|_{\mathrm{HS}} \neq 0)],\end{aligned}$$

whose order of magnitude is $O_P[\delta_(y)(s^y)^2]$, by Theorem 5 and Proposition S4.

For Δ_{43}, we have,

$$\begin{aligned}\Delta_{43} \leq &\textstyle\sum_{i,j \in \mathsf{V}^0} \|\hat{\mathfrak{C}}^y_{X_iX_j} - \mathfrak{C}^y_{X_iX_j}\|_{\mathrm{HS}}\, 1(\|\hat{\mathfrak{C}}^y_{X_iX_j}\|_{\mathrm{HS}} < \zeta, \|\mathfrak{C}^y_{X_iX_j}\|_{\mathrm{HS}} \geq \zeta) \\ &+ \textstyle\sum_{i,j \in \mathsf{V}^0} \|\hat{\mathfrak{C}}^y_{X_iX_j}\|_{\mathrm{HS}}\, 1(\|\hat{\mathfrak{C}}^y_{X_iX_j}\|_{\mathrm{HS}} < \zeta, \|\mathfrak{C}^y_{X_iX_j}\|_{\mathrm{HS}} \geq \zeta).\end{aligned}$$

Following a similar argument for the order of Δ_{41}, we can show that the orders of two terms on the right above are $O_P[\delta_n(y)(s^y)^2]$ and $O_P[\zeta(s^y)^2]$. Therefore, $\Delta_{13} =$

$O_P[\zeta(s^y)^2]$.

For Δ_{42}, we have,

$$\Delta_{42} \leq \sum_{i,j\in \mathsf{V}^0} \|\hat{\mathfrak{C}}^y_{X_iX_j} - \mathfrak{C}^y_{X_iX_j}\|_{\mathrm{HS}}\, 1(\|\hat{\mathfrak{C}}^y_{X_iX_j}\|_{\mathrm{HS}} \geq \zeta, \|\mathfrak{C}^y_{X_iX_j}\|_{\mathrm{HS}} < \zeta) \\ + \sum_{i,j\in \mathsf{V}^0} \|\mathfrak{C}^y_{X_iX_j}\|_{\mathrm{HS}}\, 1(\|\hat{\mathfrak{C}}^y_{X_iX_j}\|_{\mathrm{HS}} \geq \zeta, \|\mathfrak{C}^y_{X_iX_j}\|_{\mathrm{HS}} < \zeta) \equiv \Delta_{421} + \Delta_{422}.$$

Note that $\Delta_{422} \leq \Delta_2 = O_P[\zeta(s^y)^2]$. Moreover, given $c \in (0,1)$, $\Delta_{421} \leq \Delta_{4211} + \Delta_{4212}$, where

$$\Delta_{4211} = \sum_{i,j\in \mathsf{V}^0} \|\hat{\mathfrak{C}}^y_{X_iX_j} - \mathfrak{C}^y_{X_iX_j}\|_{\mathrm{HS}}\, 1(\|\hat{\mathfrak{C}}^y_{X_iX_j}\|_{\mathrm{HS}} \geq \zeta, c\zeta < \|\mathfrak{C}^y_{X_iX_j}\|_{\mathrm{HS}} < \zeta),$$
$$\Delta_{4212} = \sum_{i,j\in \mathsf{V}^0} \|\hat{\mathfrak{C}}^y_{X_iX_j} - \mathfrak{C}^y_{X_iX_j}\|_{\mathrm{HS}}\, 1(\|\hat{\mathfrak{C}}^y_{X_iX_j}\|_{\mathrm{HS}} \geq \zeta, \|\mathfrak{C}^y_{X_iX_j}\|_{\mathrm{HS}} \leq c\zeta).$$

For Δ_{4211}, by Theorem 5 and the fact that,

$$\Delta_{4211} \leq \max_{i,j\in \mathsf{V}^0} \|\hat{\mathfrak{C}}^y_{X_iX_j} - \mathfrak{C}^y_{X_iX_j}\|_{\mathrm{HS}} \left[\sum_{i,j\in \mathsf{V}^0} 1(c\zeta < \|\mathfrak{C}^y_{X_iX_j}\|_{\mathrm{HS}} < \zeta)\right],$$

the order of magnitude of Δ_{4211} is $O_P[\delta_n(y)(s^y)^2]$.

For Δ_{4212}, we have,

$$\Delta_{4212} \leq \max_{i,j\in \mathsf{V}^0} \|\hat{\mathfrak{C}}^y_{X_iX_j} - \mathfrak{C}^y_{X_iX_j}\|_{\mathrm{HS}} \left\{\sum_{i,j\in \mathsf{V}^0} 1[\|\hat{\mathfrak{C}}^y_{X_iX_j} - \mathfrak{C}^y_{X_iX_j}\|_{\mathrm{HS}} \geq (1-c)\zeta]\right\},$$

because $\|\hat{\mathfrak{C}}^y_{X_iX_j} - \mathfrak{C}^y_{X_iX_j}\|_{\mathrm{HS}} \geq \|\hat{\mathfrak{C}}^y_{X_iX_j}\|_{\mathrm{HS}} - \|\mathfrak{C}^y_{X_iX_j}\|_{\mathrm{HS}} \geq (1-c)\zeta$. In addition,

$$P\left(\left\{\sum_{i,j\in \mathsf{V}^0} 1[\|\hat{\mathfrak{C}}^y_{X_iX_j} - \mathfrak{C}^y_{X_iX_j}\|_{\mathrm{HS}} \geq (1-c)\zeta]\right\} > 0\right) \leq \\ P\left[\max_{i,j\in \mathsf{V}^0} \|\hat{\mathfrak{C}}^y_{X_iX_j} - \mathfrak{C}^y_{X_iX_j}\|_{\mathrm{HS}} \geq (1-c)\zeta\right],$$

which tends to 0 by the condition $\delta_n(y) \prec \zeta$. Therefore, $\sum_{i,j\in \mathsf{V}^0} 1[\|\hat{\mathfrak{C}}^y_{X_iX_j} - \mathfrak{C}^y_{X_iX_j}\|_{\mathrm{HS}} \geq (1-c)\zeta] = o_P(1)$, which implies $\Delta_{1212} = o_P[\delta_n(y)]$.

Combining the orders of Δ_{41}, Δ_{4211}, Δ_{4212}, Δ_{422}, Δ_{43}, and Δ_5, we obtain the convergence rate of $\|\check{\mathfrak{C}}^y_{XX}(d, \epsilon_Y, \epsilon_1, \zeta) - \mathfrak{C}^y_{XX}\|_{\mathrm{HS}}$.

Following a similar argument as that of Theorem 6, we can show the convergence of $\max_{i,j\in \mathsf{V}, i\neq j} \|[\check{\mathfrak{P}}^y(d, \epsilon_Y, \epsilon_1, \epsilon_2, \zeta)]_{i,j} - [\mathfrak{P}^y]_{i,j}\|_{\mathrm{HS}}$ and $P[\check{\mathsf{E}}^y_{\mathrm{CPO}}(d, \epsilon_Y, \epsilon_1, \epsilon_2, \zeta, \rho_{\mathrm{CPO}}) = \mathsf{E}^y]$. This completes the proof. □

Theorem S1 suggests that the uniform convergence rate of $\check{\mathfrak{P}}^y(d, \epsilon_Y, \epsilon_1, \epsilon_2, \zeta)$ depends on $s_y^2(\log p/n)^{(1-c)/2}$, which indicates that p can diverge at an exponential rate of n.

The second sparsity structure we consider is explicitly placed on the CCO, by directly restricting the number of nonzero elements on the off-diagonal elements of the CCO this time, instead of the CPO as in Assumption S1.

Assumption S2 *For $y \in \Omega_Y$, there exists $s_y \in \mathbb{N}$ such that*

$$\operatorname{card}\big(\{(i,j) : [\mathfrak{C}^y_{XX}]_{i,j} \neq 0, i \neq j\}\big) = s_y.$$

We consider the hard thresholding estimators in (S10), and show in the next theorem that, when replacing Assumption S1 with Assumption S2, we can further improve the order of magnitude from $s_y^2\zeta$ to $s_y\zeta$. The proof of the theorem is similar to that of Theorem S1 and is omitted.

Theorem S2 *If Assumptions 1 to 8, and S2 hold, $\epsilon_Y, \epsilon_1 \prec 1$, $d^2\epsilon_Y^{-1}\kappa_d^{-1}$ $(\log p/n)^{1/2} + d^2\epsilon_Y^{\beta} \prec 1$, $\delta_y \prec \zeta \prec 1$, and $(\epsilon_2^{-1}\zeta s_y + \epsilon_2) \prec \rho_{\text{CPO}}$, then, for any $y \in \Omega_Y$,*

$$\|\check{\mathfrak{C}}^y_{XX}(d, \epsilon_Y, \epsilon_1, \zeta) - \mathfrak{C}^y_{XX}\|_{\text{HS}} = O_p(\zeta s_y);$$
$$\max_{i,j\in\mathsf{V}, i\neq j} \|[\check{\mathfrak{P}}^y(d, \epsilon_Y, \epsilon_1, \epsilon_2, \zeta)]_{i,j} - [\mathfrak{P}^y]_{i,j}\|_{\text{HS}} = O_p(\epsilon_2^{-1}\zeta s_y + \epsilon_2);$$
$$P[\check{\mathsf{E}}^y_{\text{CPO}}(d, \epsilon_Y, \epsilon_1, \epsilon_2, \zeta, \rho_{\text{CPO}}) = \mathsf{E}^y] \to 1, \text{ as } n \to \infty.$$

Finally, we remark that, for both sparsity structures, we allow s_y to grow at the polynomial order of n. Besides, we have only considered the estimation by hard thresholding. Other regularization approaches such as the ℓ_1 penalty (Rothman et al., 2008; Cai et al., 2011) can also be used to encourage the sparsity.

S.6 Effect of tuning parameters and kernel functions

Our method involves a number of tuning parameters. We investigate the effect of those parameters on the proposed graph estimator. Overall, we have found our method is robust to the tuning parameters as long as they are within a reasonable range.

We first examine the effect of the parameters m, ϵ_T and γ_T that govern the construction of the coordinates. Recall that m is the number of basis of Ω^N, ϵ_T is the ridge parameter in (12), and γ_T is the bandwidth in the radial basis function kernel. They all control the level of smoothness on the estimation of X_i^k. Figure S1 in the Appendix reports the area under the ROC curve with varying values of m, ϵ_T and γ_T. We see that the estimated graph is robust to the choice of these parameters.

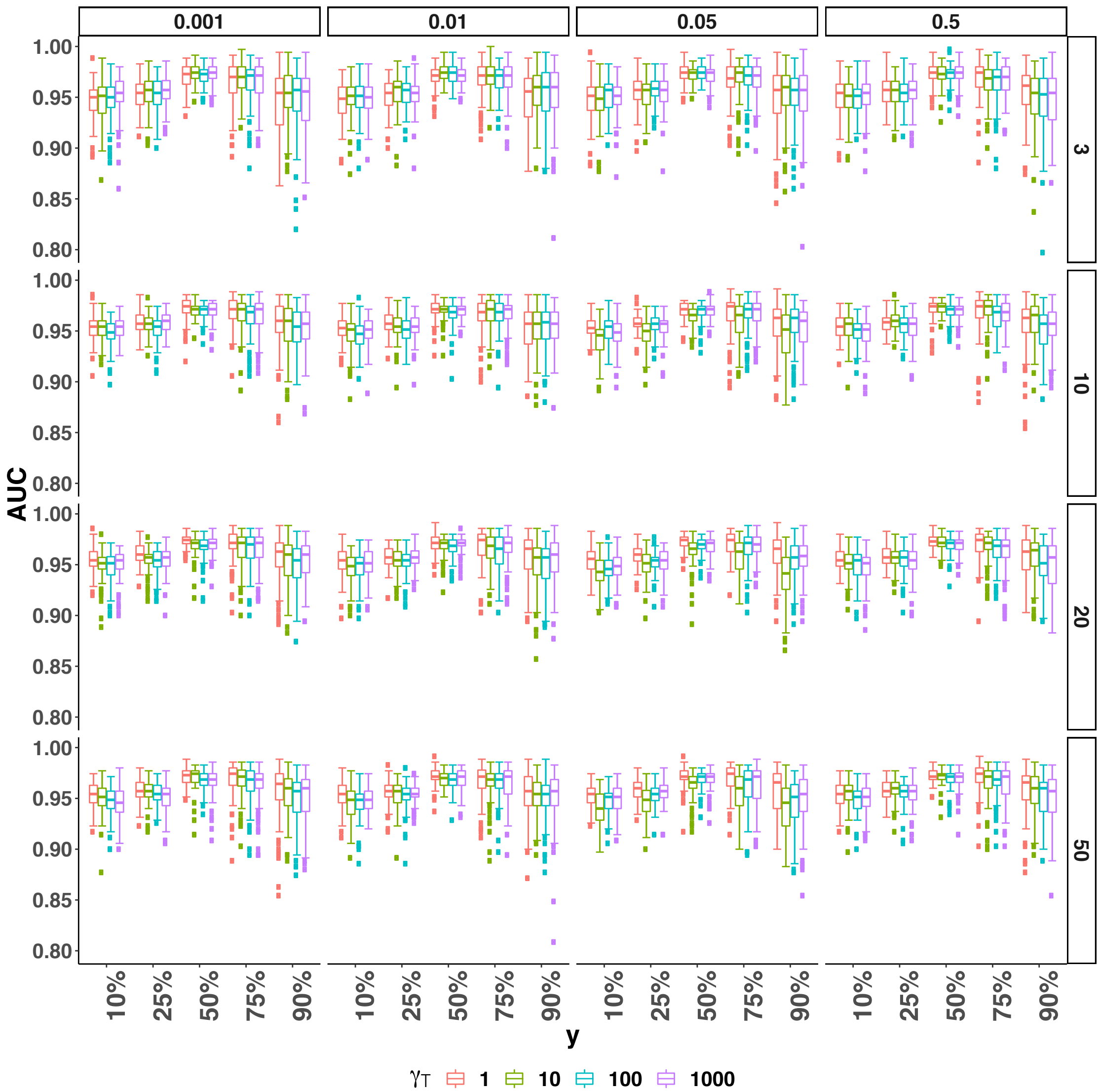

Figure S1: Area under the ROC curve for the reconstructed graph, with varying parameters ϵ_T (upper axis), m (right axis), and γ_T (colored boxes).

We next investigate the effect of d, the number of leading K-L coefficients used to approximate X_i^k. Figure S2 reports the area under the ROC curve with varying values of d from 2 to 10. We see that the performance remains about the same after d reaches 5.

We then study the effect of the ridge parameter ϵ_Y, which controls the smoothness of the estimated regression operator $\hat{\mathfrak{M}}_{X_{ij}|Y}(\epsilon_Y)$, and the two ridge parameters ϵ_1 and ϵ_2, which control the smoothness of the estimators of the conditional correlation

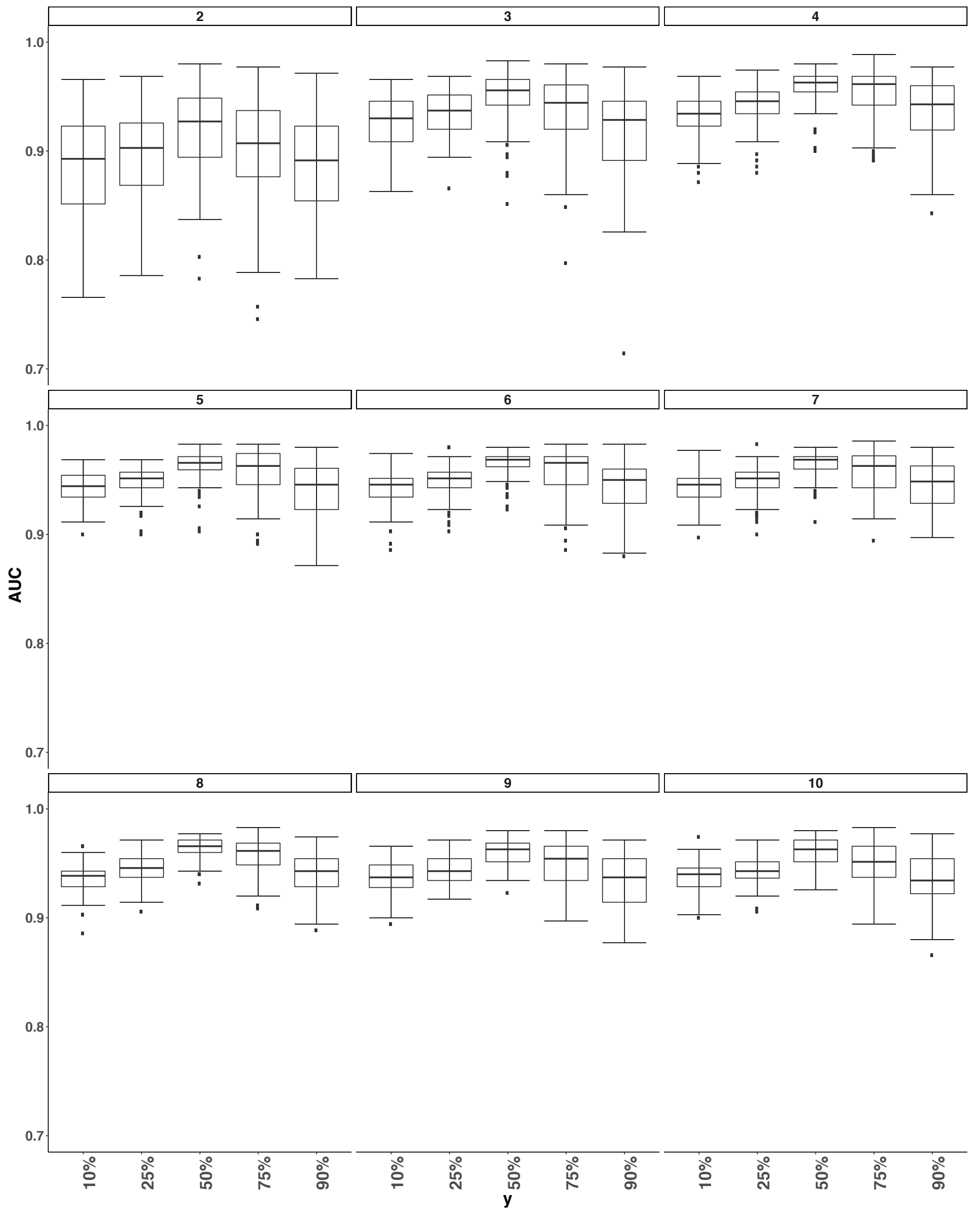

Figure S2: Area under the ROC curve for the reconstructed graph, with the varying parameter d.

operator $\hat{\mathfrak{C}}^{y}_{XX}(d, \epsilon_Y, \epsilon_1)$ and the conditional precision operator $\hat{\mathfrak{P}}^{y}(d, \epsilon_Y, \epsilon_1, \epsilon_2)$. Figure S3 reports the H-S norm of the CPO estimator under varying values of ϵ_Y. We see that, within a reasonable range of ϵ_Y, the CPO estimate is relatively robust. Figure S4 reports the H-S norm of the CPO estimator under varying values of ϵ_1 and ϵ_2 for the random graph. We observe that, a different value of ϵ_1 or ϵ_2 leads to a change of

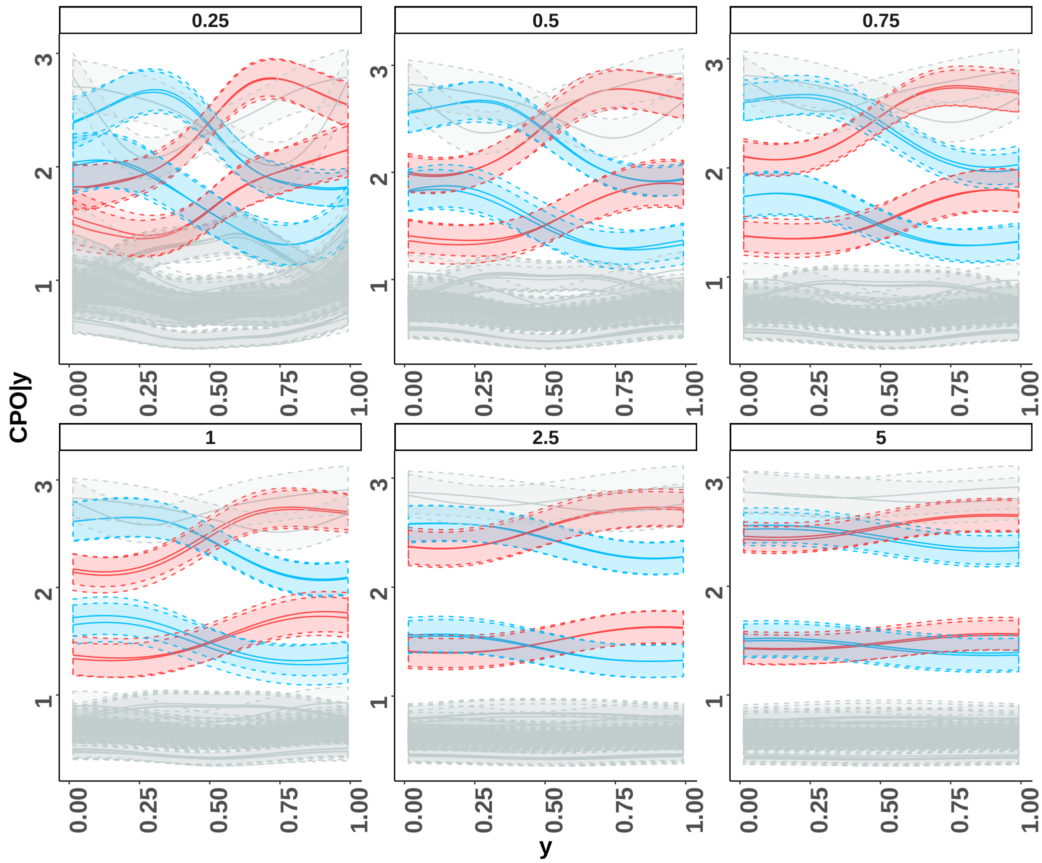

Figure S3: The Hilbert-Schmidt norm of the CPO estimator, with the varying parameter ϵ_Y

the scale of the H-S norm of the CPO, but the overall pattern does not change. This suggests that our CPO estimator is relatively robust with respect to ϵ_1 and ϵ_2 too.

Finally, we study the performance of the CPO with different choices of the kernel function. Specifically, we generate the error function $\varepsilon(t)$ using a Brownian or Gaussian kernel as the basis function. We then choose a Brownian or Gaussian kernel for κ_T, and a Laplacian, Student t or Gaussian kernel for κ_Y. This leads to 12 combinations of $(\varepsilon(t), \kappa_T, \kappa_Y)$. Figure S5 reports the area under the ROC curve for the graph estimated by our CPO method for those different combinations. It is seen that the performance of our method is consistent across all combinations.

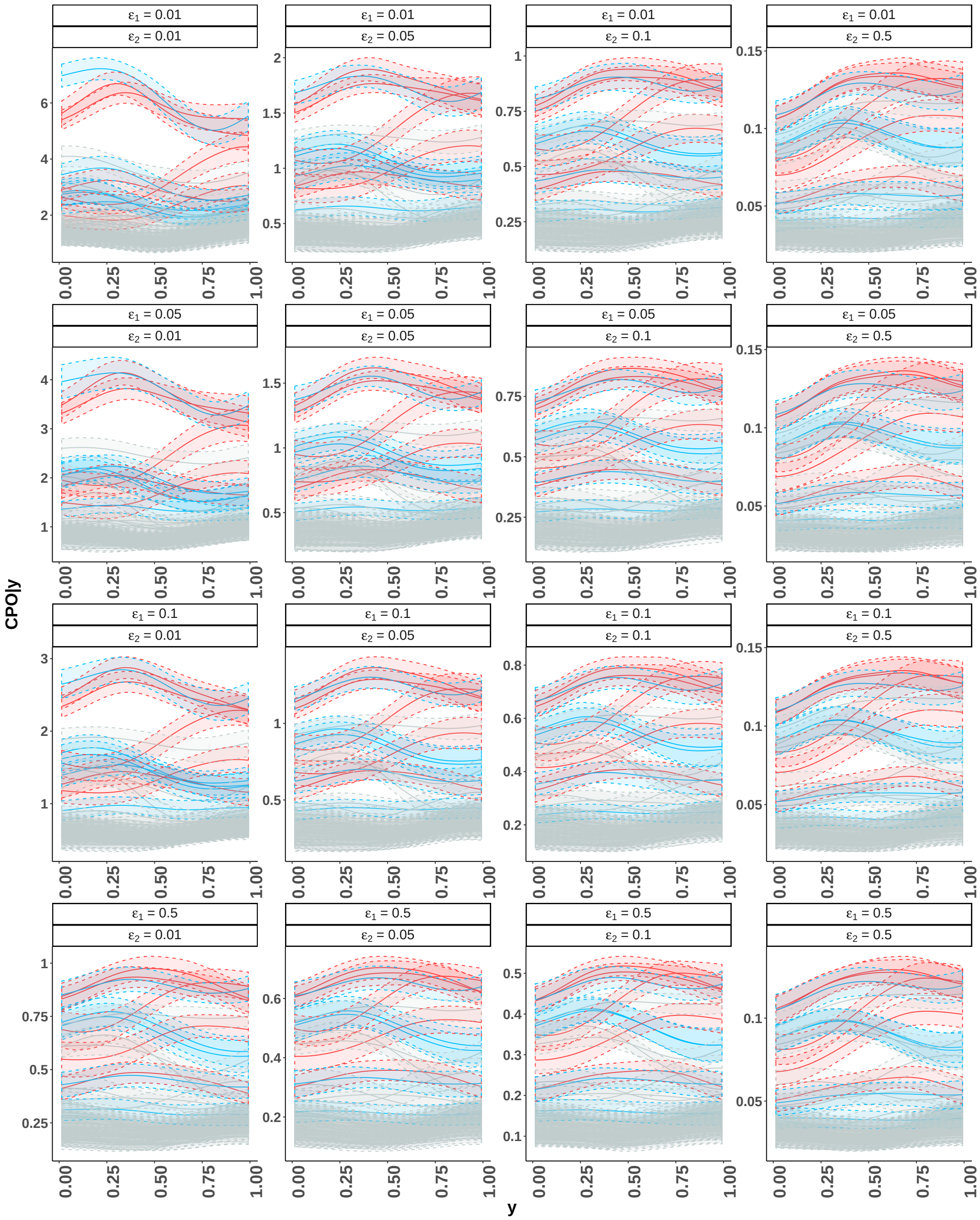

Figure S4: Norm of the CPO estimator, with the varying parameters ϵ_1 (upper axis) and ϵ_2 (right axis).

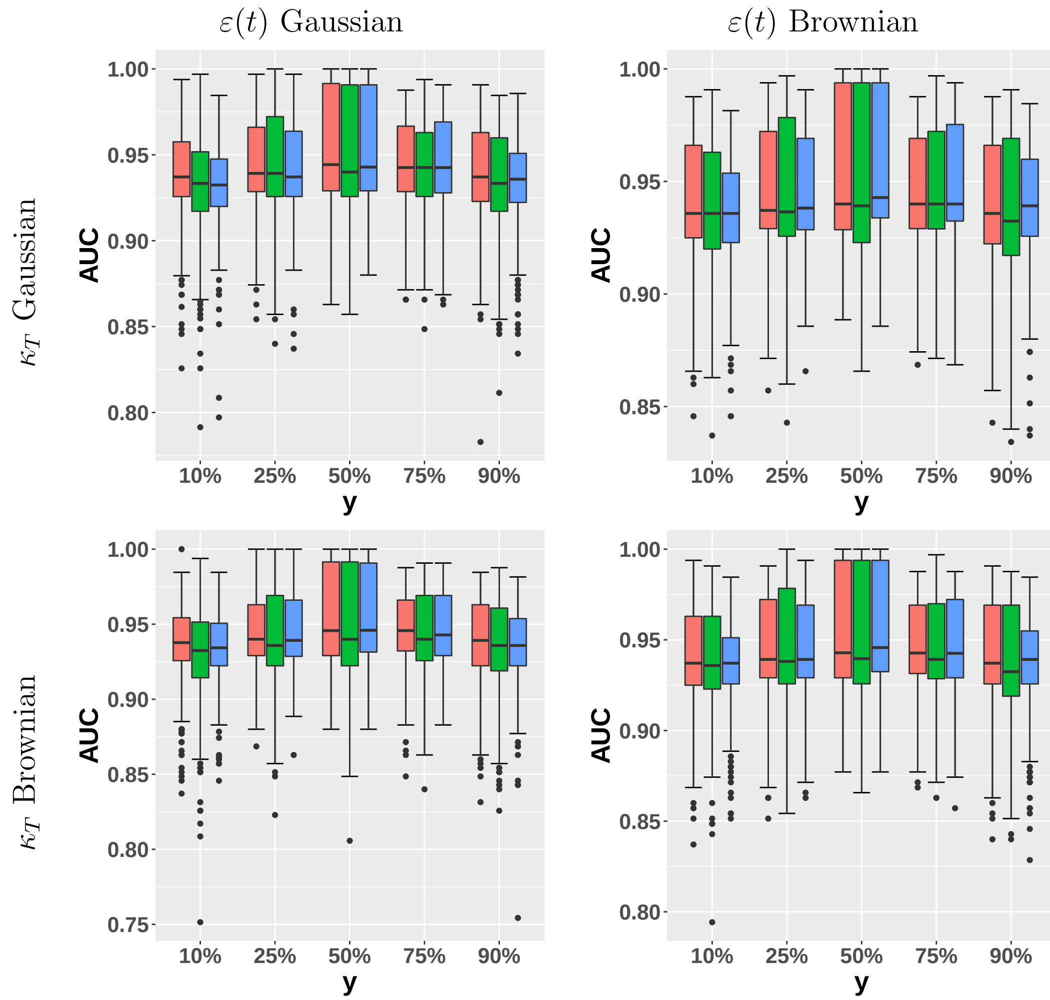

Figure S5: Area under the ROC curve of the estimated graph by the CPO, with respect to the external variable Y, under three kernel functions for κ_Y: Gaussian (red), Laplacian (green) and Student t (blue), two kernel functions for κ_T: Gaussian (top) and Brownian (bottom), and two kernel functions for $\varepsilon(t)$: Gaussian (left) and Brownian (right).

S.7 Brain connectivity validation analysis

We report the analysis result of an independent validation dataset of 828 subjects from HCP. Figure S6 reports the changes of the identified significant edges, with respect to the intelligence score at $7, 11, 15, 19, 23$, for the medial frontal module from the new dataset. The finding is similar to that reported in the paper.

References

Bach, F. (2009). High-dimensional non-linear variable selection through hierarchical kernel learning. *arXiv preprint arXiv:0909.0844*.

Bach, F. R. (2008). Consistency of the group lasso and multiple kernel learning. *Journal of Machine Learning Research*, 9:1179–1225.

Bosq, D. (2000). *Linear Processes in Function Spaces*. Springer, New York.

Cai, T., Liu, W., and Luo, X. (2011). A Constrained ℓ1Minimization Approach to Sparse Precision Matrix Estimation. *Journal of the American Statistical Association*, 106(494):594–607.

Fukumizu, K., Bach, F. R., and Jordan, M. I. (2009). Kernel dimension reduction in regression. *The Annals of Statistics*, 37(5):1871–1905.

Hanson, D. L. and Wright, F. T. (1971). A bound on tail probabilities for quadratic forms in independent random variables. *Ann. Math. Statist.*, 42(3):1079–1083.

Kato, T. (1980). *Perturbation Theory of Linear Operators*. Springer.

Kazdan, J. L. (1971). Perturbation of complete orthonormal sets and eigen-function expansions. *Proceedings of the American Mathematical Society*, 27:506–510.

Li, B. and Solea, E. (2018). A nonparametric graphical model for functional data with application to brain networks based on fmri. *Journal of the American Statistical Association*, 113(524):1637–1655.

Rothman, A. J., Bickel, P. J., Levina, E., and Zhu, J. (2008). Sparse permutation invariant covariance estimation. *Electronic Journal of Statistics*, 2(0):494–515.

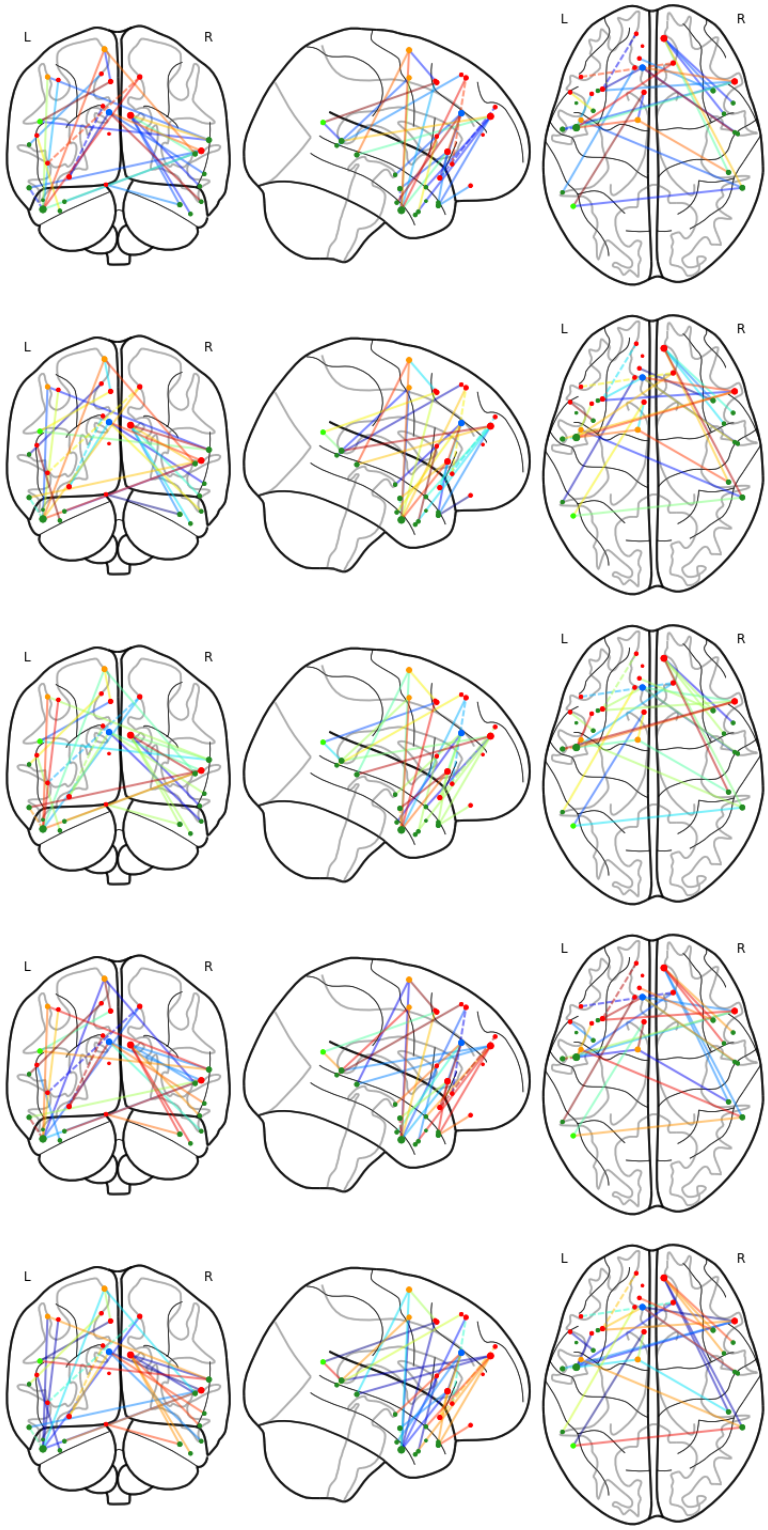

Figure S6: Medial frontal network changes, with respect to the intelligence score at $7, 11, 15, 19, 23$, based on an independent validation dataset. Blue color represents the small H-S norm value of CPO, green the medium norm value, and red the high H-S norm value.