Supplement to "Network Functional Linear Regression"

Xingyu Yan[1] and Yanyuan Ma[2]

Jiangsu Normal University and Pennsylvania State University

This supplement contains the proofs of all the main results presented in the article, along with supporting lemmas and additional simulation results, including

S1 Extension to network partially functional linear regression . . . 2

S2 Some useful lemmas . . . 4

S3 Proof of theory . . . 9

S3.1 Verifications of several claims . . . 9

S3.2 Proof of Lemma S2.2 . . . 12

S3.3 Proof of Lemma S2.3 . . . 16

S3.4 Proof of Theorem 1 . . . 18

S3.5 Proof of Lemma S2.4 . . . 29

S3.6 Proof of Lemma S2.5 . . . 36

S3.7 Proof of Lemma S2.6 . . . 37

S3.8 Proof of Lemma S2.8 . . . 55

S3.9 Proof of Lemma S2.9 . . . 55

S3.10 Proof of Theorem 2 . . . 57

S3.11 Proof of (S.36) . . . 104

S3.12 Proof of (S.43) . . . 107

S3.13 Proof of $E(\mathbf{F}^{\mathrm{T}}\ddot{\mathbf{F}}) = 0$ . . . 109

S3.14 Proof of Lemma S2.10 . . . 112

S3.15 Proof of Theorem 3 . . . 118

S3.16 Proof of Theorem 4 . . . 120

S4 Additional simulation results . . . 123

S1 Extension to network partially functional linear regression

Inspired by an anonymous reviewer, we discuss a natural extension in this section. Our proposed method can be extended to handle the simultaneous presence of multiple functional predictors and scalar covariates. Specifically, for $i = 1, \dots, n$, let Y_i be a scalar continuous response, $\{Z_{i\ell}(\cdot), \ell = 1, \dots, J\}$ are J functional predictors, $\mathbf{X}_i = (X_{i1}, \dots, X_{ip})^{\mathrm{T}}$ be a p-dimensional vector of scalar covariates. Without loss of generality, we assume that the response, the functional predictors and the scalar covariates have been centred to have mean zero. Then, the network partially functional linear regression is defined as

$$Y_i \;=\; \sum_{\ell=1}^{J} \int_{\mathcal{I}} Z_{i\ell}(t)\beta_\ell(t)dt + \mathbf{X}_i^{\mathrm{T}}\boldsymbol{\vartheta} + \rho \sum_{j=1}^{n} w_{ij} Y_j + \varepsilon_i, \quad i = 1, \dots, n, \tag{S.1}$$

where $\{\beta_{j\ell}(\cdot), \ell = 1, \dots, J\}$ are square-integrable regression functions, $\boldsymbol{\vartheta} = (\vartheta_1, \dots, \vartheta_p)^{\mathrm{T}} \in \mathbb{R}^p$ represents the regression coefficients of nonfunctional covariates. The term ρ is the network autocorrelation coefficient, and $(w_{ij})_{i=1,\dots,n;j=1,\dots,n}$ is row-normalized adjacency matrix. The i.i.d. ε_i is the random error. The detailed explanation of the above terms is similar to that in (2.2).

Based on model (S.1) and motivated by the proposed method in Section 2.2, we present a modified three-step algorithm as follows.

Step 1: We perform FPCA separately to each functional predictor and use the pooled FPCA scores as predictors. The estimator of the FPC scores of $Z_{i\ell}(\cdot)$ is denoted by $\{\widehat{\xi}_{i\ell,1}, \dots, \widehat{\xi}_{i\ell,K_\ell}\}$ for $\ell = 1, \dots, J$. Here, $\widehat{\xi}_{i\ell,k}$ is computed as $\int_{\mathcal{I}} Z_{i\ell}(t)\widehat{\phi}_{\ell,k}(t)dt$, where $\widehat{\phi}_{\ell,k}(t)$ denotes the kth estimated eigenfunction of $Z_{i\ell}(\cdot)$ for $k = 1, \dots, K_\ell$. The number of principal components, K_ℓ, is selected according to the procedure outlined in Section 2.3. With these notations, we can approximate (S.1) as

$$Y_i \;=\; \sum_{\ell=1}^{J}\sum_{k=1}^{K}\widehat{\xi}_{i\ell,k}b_{\ell k} + \mathbf{X}_i^{\mathrm{T}}\boldsymbol{\vartheta} + \rho\sum_{j=1}^{n}w_{ij}Y_j + \varepsilon_i,$$

where $b_{\ell k}$ represents the projection of $\beta_\ell(\cdot)$ onto the kth eigenfunction of $Z_{i\ell}(\cdot)$.

Step 2: For $\ell = 1,\ldots,L$, let the estimated scores matrix $\widehat{\mathbf{A}}_\ell = (\widehat{\boldsymbol{\xi}}_{1\ell},\ldots,\widehat{\boldsymbol{\xi}}_{n\ell})^{\mathrm{T}} \in \mathbb{R}^{n\times K}$ serve as the predictor variables, where $\widehat{\boldsymbol{\xi}}_{i\ell} = (\widehat{\xi}_{i\ell,1},\ldots,\widehat{\xi}_{i\ell,K_\ell})^{\mathrm{T}}$ represents the estimated FPC scores for the ℓth functional predictor. Let $\mathbf{b}_\ell = (b_{\ell,1},\ldots,b_{\ell,K})^{\mathrm{T}} \in \mathbb{R}^{K_\ell}$ denote the corresponding coefficient vector. At any given ρ, similar to (2.11), the estimator $\widehat{\boldsymbol{\vartheta}}(\rho)$ and $\{\widehat{\mathbf{b}}_\ell(\rho), \ell = 1,\ldots,K_\ell\}$ are obtained by solving the following least square type optimization problem:

$$\begin{aligned}
(\widehat{\boldsymbol{\vartheta}}(\rho),\widehat{\mathbf{b}}_1(\rho),\ldots,\widehat{\mathbf{b}}_J(\rho)) \;&=\; \underset{\mathbf{b}_1,\ldots,\mathbf{b}_\ell,\boldsymbol{\vartheta}}{\operatorname{argmin}}\frac{1}{n}\left\|\mathbf{S}(\rho)\mathbf{Y} - \sum_{\ell=1}^{J}\widehat{\mathbf{A}}_\ell\mathbf{b}_\ell - \mathbf{X}^{\mathrm{T}}\boldsymbol{\vartheta}\right\|^2\\
&=\; \underset{\mathbf{b}_1,\ldots,\mathbf{b}_\ell,\boldsymbol{\vartheta}}{\operatorname{argmin}}\frac{1}{n}\left\|\mathbf{S}(\rho)\mathbf{Y} - \widehat{\boldsymbol{\eta}}\widetilde{\boldsymbol{\theta}}\right\|^2,
\end{aligned}$$

where $\mathbf{Y} = (Y_1,\ldots,Y_n)^{\mathrm{T}} \in \mathbb{R}^n$, $\mathbf{X}^{\mathrm{T}} = (\mathbf{X}_1^{\mathrm{T}},\ldots,\mathbf{X}_n^{\mathrm{T}}) \in \mathbb{R}^{n\times p}$, $\widehat{\boldsymbol{\eta}} = (\widehat{\mathbf{A}}_1,\ldots,\widehat{\mathbf{A}}_J,\mathbf{X}^{\mathrm{T}}) \in \mathbb{R}^{n\times(JK+p)}$, and $\widetilde{\boldsymbol{\theta}} = (\mathbf{b}_1^{\mathrm{T}},\ldots,\mathbf{b}_J^{\mathrm{T}},\boldsymbol{\vartheta}^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{R}^{JK+p}$.

Step 3: Obtain the estimator $\widehat{\rho}$ using the following composite least squares objective function:

$$\widehat{\rho} \;=\; \underset{\rho}{\operatorname{argmin}}\frac{1}{n}\left\|\{\mathrm{diag}\mathbf{M}(\rho)\}^{-1}\mathbf{S}(\rho)^{\mathrm{T}}\left\{\mathbf{S}(\rho)\mathbf{Y} - \sum_{\ell=1}^{J}\widehat{\mathbf{A}}_\ell\widehat{\mathbf{b}}_\ell(\rho) - \mathbf{X}_i^{\mathrm{T}}\widehat{\boldsymbol{\vartheta}}(\rho)\right\}\right\|^2.$$

On the theoretical side, similar asymptotic results for the functional regression coefficients $\beta_\ell(t), \ell = 1,\ldots,L$ and the network autoregression coefficient ρ can be established by following the proofs of Theorem 1 and Theorem 2, with the involved eigencomponents replaced by pooled eigencomponents. It would be interesting to also investigate the asymptotic normality for the parameter regression coefficients $\boldsymbol{\vartheta}$ estimation, while we leave this for future research.

S2 Some useful lemmas

This section presents additional lemmas essential for the proofs of Theorems 1-4. Detailed proofs and further explanations can be found in the Supplementary Material S3.

The following Lemmas S2.1-S2.3 will be used to prove Theorem 1, which contains the asymptotic results for the functional regression coefficient. The proof of Lemma S2.1 can be found in Hall and Horowitz (2007), the proof of Lemma S2.2 is presented in Supplementary Material S3.2, and the proof of Lemma S2.3 is provided in Supplementary Material S3.3.

Lemma S2.1. *Suppose that Conditions (C1)-(C2) hold, Then as $n \to \infty$, if $k = o(n^{\frac{1}{\alpha+2\tau}})$, the following results hold:*

(a) $\sup_{k\geqslant 1} |\widehat{\lambda}_k - \lambda_k| = O_p(n^{-1/2})$.

(b)

$$\begin{aligned} E\|\widehat{G} - G\|^2 &= O_(n^{-1}), \\ E\left\|\int \left(\widehat{G}(s,t) - G(s,t)\right)\phi_k(t)dt\right\|^2 &= O(n^{-1}), \\ E\left\{\iint \left(\widehat{G}(s,t) - G(s,t)\right)\phi_k(s)\phi_k(t)dsdt\right\}^2 &= O(n^{-1}\lambda_k^2). \end{aligned}$$

(c) $\|\widehat{\phi}_k(t) - \phi_k(t)\| = O_p(kn^{-1/2})$.

Lemma S2.2. *Suppose that Conditions (C1)-(C3) hold, the following results hold:*

(a) $\sum_{\ell:\ell\neq k}^{\infty} \ell^{-2(\alpha+\tau)}(\lambda_k - \lambda_\ell)^{-4} = O(k^{2\alpha-2\tau+4} + \log k)$.

(b) $\sum_{\ell:\ell\neq k}^{\infty} \ell^{-(\frac{3}{2}\alpha+\tau)} |\lambda_k - \lambda_\ell|^{-1} = O(1)$.

(c) $\sum_{\ell:\ell\neq k}^{\infty} \ell^{-(\alpha+\tau)}|\lambda_k - \lambda_\ell|^{-1} = O(1)$.

Define the estimator

$$\begin{aligned}\widehat{g}(t) &= n^{-1}\sum_{i=1}^{n}\left(Y_i - \rho\sum_{j=1}^{n} w_{ij}Y_j\right)Z_i(t)\\ &= n^{-1}\sum_{i=1}^{n} Y_i^* Z_i(t).\end{aligned}$$

The L^2 convergence rate of the estimator is given below.

Lemma S2.3. *Assume that (C1) holds, then* $\|g\|$ *is bounded and* $\|\widehat{g}(t) - g(t)\| = O_p(n^{-1/2})$.

Below, we present several lemmas that will be instrumental in proving the asymptotic results for the network autoregression coefficient in Theorem 2.

Lemma S2.4. *(a) Let* $\delta_k = \min_{1\leqslant \ell \leqslant k}(\lambda_\ell - \lambda_{\ell+1})$, *then*

$$\sup_{k\geqslant 1}|\widehat{\lambda}_k - \lambda_k| \leqslant \|\widehat{G} - G\|, \qquad \sup_{k\geqslant 1}\delta_k\|\widehat{\phi}_k - \phi_k\| \leqslant 8^{1/2}\|\widehat{G} - G\|.$$

(b) Under the Conditions (C1)-(C2), for $k = o(n^{\frac{1}{2\alpha+2}})$,

$$\widehat{\lambda}_k - \lambda_k = \frac{1}{n}\sum_{i=1}^{n}(\xi_{ik}^2 - \lambda_k) + O_p(k/n).$$

$$\widehat{\xi}_{ik} - \xi_{ik} = \sum_{\ell\neq k}(\lambda_k - \lambda_\ell)^{-1}\xi_{ik}\left\langle\left(\widehat{G} - G\right)\phi_\ell, \phi_k\right\rangle \times \{1 + o_p(1)\}.$$

In addition, if $\alpha > 2$, *then*

$$\widehat{\phi}_k(t) - \phi_k(t) = \sum_{\ell\neq k}(\lambda_k - \lambda_\ell)^{-1}\phi_\ell\left\langle\left(\widehat{G} - G\right)\phi_\ell, \phi_k\right\rangle + O_p\left(n^{-1}k^{\alpha+2}logk\right).$$

(c) Under the Conditions (C1), (C2), assume $E(\xi_k^2 - \lambda_k)^2 < C\lambda_k^2$*, and* $k = o(n^{\frac{1}{2\alpha+2}})$*. Then*

$$\widehat{\lambda}_k - \lambda_k = O_p(n^{-1/2}\lambda_k) \quad \text{and} \quad \widehat{\xi}_{ik} - \xi_{ik} = O_p(k^{1-\alpha/2}n^{-1/2})$$

uniformly in k*.*

Note that the conditions $E(\xi_k^2 - \lambda_k)^2 < C\lambda_k^2$ and $k = o(n^{\frac{1}{2\alpha+2}})$ in Lemma S2.4 were used in Wong et al. (2019). The proof of Lemma S2.4 is provided in Supplementary Material S3.10. The following Lemma S2.5 and Lemma S2.6 are devised for measuring the approximation error of scores caused by truncation as well as the statistical estimation error. Their proofs are deferred to Supplementary Material S3.6 and Supplementary Material S3.7, respectively.

Lemma S2.5. *Under the Conditions (C1), (C2) and (C5), assuming* $E(\xi_k^2 - \lambda_k)^2 < C\lambda_k^2$ *and* $k = o(n^{\frac{1}{2\alpha+2}})$*, if* $\alpha > 2$*, then*

$$\lambda_{\max}\left\{\frac{1}{\sqrt{n}}\left(\widehat{\mathbf{A}}^{\mathrm{T}} - \mathbf{A}^{\mathrm{T}}\right)\left(\widehat{\mathbf{A}} - \mathbf{A}\right)\right\} = o_p(1).$$

Lemma S2.6. *Assume that the assumptions in Lemma S2.5 hold. (a) Under the Conditions (C3),(C7)-(C9), we have*

$$\begin{aligned} E\left[\frac{1}{\sqrt{n}}\sum_{i=1}^n S_i\{\widehat{\mathbf{A}};\mathbf{b}(\rho),\rho\} - \frac{1}{\sqrt{n}}\sum_{i=1}^n S_i\{\mathbf{A};\mathbf{b}(\rho),\rho\}\right] &= o(1), \\ var\left[\frac{1}{\sqrt{n}}\sum_{i=1}^n S_i\{\widehat{\mathbf{A}};\mathbf{b}(\rho),\rho\} - \frac{1}{\sqrt{n}}\sum_{i=1}^n S_i\{\mathbf{A};\mathbf{b}(\rho),\rho\}\right] &= O(1). \end{aligned} \tag{S.2}$$

(b) Under the Conditions (C7) and (C8), we have

$$\left\|\frac{1}{n}\sum_{i=1}^n \frac{\partial S_i\{\widehat{\mathbf{A}};\mathbf{b}(\rho),\rho\}}{\partial \mathbf{b}(\rho)^{\mathrm{T}}} - \frac{1}{n}\sum_{i=1}^n \frac{\partial S_i\{\mathbf{A};\mathbf{b}(\rho),\rho\}}{\partial \mathbf{b}(\rho)^{\mathrm{T}}}\right\| \quad = \quad o_p(n^{-1/4}), \tag{S.3}$$

and

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\mathrm{d}S_i\{\widehat{\mathbf{A}};\mathbf{b}(\rho),\rho\}}{\mathrm{d}\rho} \quad = \quad \frac{1}{n}\sum_{i=1}^{n}\frac{\mathrm{d}S_i\{\mathbf{A};\mathbf{b}(\rho),\rho\}}{\mathrm{d}\rho}+o_p\left(n^{-1/2}\right). \tag{S.4}$$

Lemma S2.7. *If* $\mathbf{y}$ *is a random vector with mean* $\boldsymbol{\mu}$ *and covariance matrix* $\boldsymbol{\Sigma}$ *and if* $\mathbf{A}$ *is a symmetric matrix of constants, then*

$$E\left(\mathbf{y}^{\mathrm{T}}\mathbf{A}\mathbf{y}\right) = \mathit{trace}(\mathbf{A}\boldsymbol{\Sigma}) + \boldsymbol{\mu}^{\mathrm{T}}\mathbf{A}\boldsymbol{\mu}.$$

If $\mathbf{y}$ *is* $\mathcal{N}(\boldsymbol{\mu},\boldsymbol{\Sigma})$*, then*

$$\mathit{var}\left(\mathbf{y}^{\mathrm{T}}\mathbf{A}\mathbf{y}\right) = 2\mathit{trace}\left\{(\mathbf{A}\boldsymbol{\Sigma})^2\right\} + 4\boldsymbol{\mu}^{\mathrm{T}}\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}\boldsymbol{\mu}.$$

The above Lemma S2.7 provides the expectation and variance of quadratic forms and the proof of Lemma S2.7 can be found in many textbooks, such as Rencher and Schaalje (2008).

Lemma S2.8. *(a) Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be positive semidefinite* $n\times n$ *matrices, then*

$$0 \leqslant \mathit{trace}(\mathbf{AB}) \leqslant \lambda_{\max}(\mathbf{A})\mathit{trace}(\mathbf{B}) \leqslant \mathit{trace}(\mathbf{A})\mathit{trace}(\mathbf{B}).$$

(b) Let $\mathbf{A}$ *and* $\mathbf{B}$ *be symmetric* $n\times n$ *matrices, then*

$$\lambda_{\max}\left\{(\mathbf{AB})^{\mathrm{T}}(\mathbf{AB})\right\} \quad \leqslant \quad \lambda_{\max}^2(\mathbf{A})\lambda_{\max}^2(\mathbf{B}).$$

(c) Let $\mathbf{A}$ *be an* $n \times n$ *matrix, then*

$$trace\left(\mathbf{A}^2\right) \leqslant trace\left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)$$

with equality if and only if $\mathbf{A}$ *is symmetric.*

Lemma S2.9. *For any* $n \times n$ *matrices* $\mathbf{M}_1 = (m_{1,ij}) \in \mathbb{R}^{n\times n}$ *and* $\mathbf{M}_2 = (m_{2,ij}) \in \mathbb{R}^{n\times n}$*, and n-dimensional vectors* $\mathbf{U}_1 \in \mathbb{R}^{n\times 1}$ *and* $\mathbf{U}_2 \in \mathbb{R}^{n\times 1}$*. Let* $\mathbf{X} = (X_1, \cdots, X_n)^{\mathrm{T}} \in \mathbb{R}^n$*, where* $X_i \in \mathbb{R}(1 \leqslant i \leqslant n)$ *are identically and independently distributed variables. Write* $Q_1 = \mathbf{X}^{\mathrm{T}}\mathbf{M}_1\mathbf{X} + \mathbf{U}_1^{\mathrm{T}}\mathbf{X}$*, and* $Q_2 = \mathbf{X}^{\mathrm{T}}\mathbf{M}_2\mathbf{X} + \mathbf{U}_2^{\mathrm{T}}\mathbf{X}$*. Assume the following conditions are satisfied: (1)* $E(X_i) = 0$ *and* $E(X_i^3) = 0$ *for* $i = 1, \ldots, n$*; (2)* $E(X_i^2) = 1$ *and* $E\left(X_i^4\right) < \infty$*, then*

$$\begin{aligned} cov\left(Q_1, Q_2\right) &= \left\{trace(\mathbf{M}_1\mathbf{M}_2^{\mathrm{T}}) + trace(\mathbf{M}_1\mathbf{M}_2)\right\} \\ &\quad + trace\left\{diag(\mathbf{M}_1)\, diag(\mathbf{M}_2)\right\}\left\{E(X_1^4) - 3\right\} + \mathbf{U}_1^{\mathrm{T}}\mathbf{U}_2. \end{aligned}$$

The above lemma is similar to Lemma 2 of Huang et al. (2019), which is used in the proof of Theorem 2. We provide the proof of Lemma S2.8 and Lemma S2.9 in Supplementary Material S3.8 and S3.9, respectively.

Lemma S2.10. *(a) Assume Conditions (C1), (C2), (C5) and (C8) hold. Then*

$$\left|\lambda_{\min}(\widehat{\mathbf{\Lambda}}) - \lambda_{\min}(\mathbf{\Lambda})\right| = o_p(1) \quad and \quad \left|\lambda_{\max}(\widehat{\mathbf{\Lambda}}) - \lambda_{\max}(\mathbf{\Lambda})\right| = o_p(1),$$

where $\widehat{\mathbf{\Lambda}} = n^{-1}\widehat{\mathbf{A}}_{K*}^{\mathrm{T}}\widehat{\mathbf{A}}_{K*}$ *and* $\mathbf{\Lambda} = n^{-1}\mathbf{A}_{K*}^{\mathrm{T}}\mathbf{A}_{K*}$*, with* K^* *being an arbitrary positive integer.*

(b) Under Conditions (C1) and (C2), the projection matrices defined in (3.20) *satisfy the following:*

(i)

$$\left\| \frac{1}{\sqrt{n}} \left(\widehat{\mathbf{A}}_0^{\mathrm{T}} - \mathbf{A}_0^{\mathrm{T}} \right) \mathbb{A}_k \right\| \quad = \quad O_p\left(\frac{K^{3/2-\alpha/2}}{n^{1/4}} \right).$$

(ii) Assume Conditions (C5) and (C8) hold. Then

$$\|\widehat{\mathbb{A}}_k^{proj} - \mathbb{A}_k^{proj}\| = O_p(1).$$

The above lemma quantifies the asymptotic orders of several key expressions that will be used in the proofs of Theorems 3 and 4.

S3 Proof of theory

S3.1 Verifications of several claims

Note that for $k = 1, \ldots, K$,

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} \widehat{\xi}_{ik}^2 &= \frac{1}{n}\sum_{i=1}^{n} \left\{ \int Z_i(t)\widehat{\phi}_k(t)dt \right\}^2 \\
&= \frac{1}{n}\sum_{i=1}^{n} \left(\int Z_i(t)\widehat{\phi}_k(t)dt \right) \left(\int Z_i(s)\widehat{\phi}_k(s)ds \right) \\
&= \frac{1}{n}\sum_{i=1}^{n} \iint Z_i(t)Z_i(s)\widehat{\phi}_k(t)\widehat{\phi}_k(s)dtds \\
&= \iint \left[\frac{1}{n}\sum_{i=1}^{n} Z_i(t)Z_i(s) \right] \widehat{\phi}_k(t)\widehat{\phi}_k(s)dtds \\
&= \iint \widehat{G}(t,s)\widehat{\phi}_k(t)\widehat{\phi}_k(s)dtds \\
&= \mathrm{cov}(\widehat{\xi}_k, \widehat{\xi}_k) = \widehat{\lambda}_k.
\end{aligned}$$

Let $\mathbf{Y}^* \equiv \mathbf{S}\mathbf{Y}$, $\widehat{g}(t) \equiv n^{-1}\sum_{i=1}^n Y_i^* Z_i(t)$ and $g(t) \equiv E[Y_i^* Z_i(t)]$. Equation (2.11) leads to

$$\begin{aligned}
\widehat{b}_k(\rho) &= \mathbf{e}_k^{\mathrm{T}} \widehat{\mathbf{\Lambda}}^{-1} \left(\frac{1}{n} \widehat{\mathbf{A}}^{\mathrm{T}} \mathbf{Y}^* \right) \\
&= \widehat{\lambda}_k^{-1} n^{-1} \sum_{i=1}^n \boldsymbol{\xi}_{ik} Y_i^* \\
&= \widehat{\lambda}_k^{-1} n^{-1} \sum_{i=1}^n \int Z_i(t) \widehat{\phi}_k(t) dt Y_i^* \\
&= \widehat{\lambda}_k^{-1} \left\langle n^{-1} \sum_{i=1}^n Y_i^* Z_i(t), \widehat{\phi}_k(t) \right\rangle \\
&= \widehat{\lambda}_k^{-1} \left\langle \widehat{g}(t), \widehat{\phi}_k(t) \right\rangle . \qquad \text{(S.5)}
\end{aligned}$$

Below, we show that

$$E\left\{ \frac{\partial Q(\mathbf{b}, \rho)}{\partial \rho} \right\} \neq 0,$$

where $Q(\mathbf{b}, \rho)$ is defined in (2.9). For ease of notation, let $\boldsymbol{f}(\rho) = \frac{1}{\sqrt{n}}(\mathbf{S}(\rho)\mathbf{Y} - \widehat{\mathbf{A}}\mathbf{b})$, then

$$Q(\mathbf{b}, \rho) = \boldsymbol{f}(\rho)^{\mathrm{T}} \boldsymbol{f}(\rho),$$

which implies that

$$\begin{aligned}
\frac{\partial Q(\mathbf{b}, \rho)}{\partial \rho} &= 2\boldsymbol{f}(\rho)^{\mathrm{T}} \frac{\partial \boldsymbol{f}(\rho)}{\partial \rho} \\
&= \frac{2}{n} \left\{ \mathbf{S}(\rho)\mathbf{Y} - \widehat{\mathbf{A}}\mathbf{b} \right\}^{\mathrm{T}} \frac{\partial \mathbf{S}(\rho)}{\partial \rho} \mathbf{Y} \\
&= \frac{2}{n} \left\{ \mathbf{Y}^{\mathrm{T}} \mathbf{S}(\rho)^{\mathrm{T}} \frac{\partial \mathbf{S}(\rho)}{\partial \rho} \mathbf{Y} - \mathbf{b}^{\mathrm{T}} \widehat{\mathbf{A}}^{\mathrm{T}} \frac{\partial \mathbf{S}(\rho)}{\partial \rho} \mathbf{Y} \right\} .
\end{aligned}$$

Recall that the mean and covariance of response $\mathbf{Y}$ are $\mathbf{S}^{-1}(\rho)\widehat{\mathbf{A}}\mathbf{b}$ and $\sigma^2 \mathbf{S}^{-1}(\rho)(\mathbf{S}^{-1}(\rho))^{\mathrm{T}}$, respectively. Then,

$$
\begin{aligned}
& E\left\{\frac{\partial Q(\mathbf{b},\rho)}{\partial \rho} \mid \widehat{\mathbf{A}}\right\} \\
= \;& \frac{2}{n}E\left\{\mathbf{Y}^{\mathrm{T}}\mathbf{S}(\rho)^{\mathrm{T}}\frac{\partial \mathbf{S}(\rho)}{\partial \rho}\mathbf{Y} \mid \widehat{\mathbf{A}}\right\} - \frac{2}{n}\mathbf{b}^{\mathrm{T}}\widehat{\mathbf{A}}^{\mathrm{T}}\frac{\partial \mathbf{S}(\rho)}{\partial \rho}E\left(\mathbf{Y} \mid \widehat{\mathbf{A}}\right) \\
= \;& \frac{2}{n}\mathrm{trace}\left\{\mathbf{S}(\rho)^{\mathrm{T}}\frac{\partial \mathbf{S}(\rho)}{\partial \rho}\mathrm{cov}(\mathbf{Y} \mid \widehat{\mathbf{A}})\right\} + \frac{2}{n}E\left(\mathbf{Y} \mid \widehat{\mathbf{A}}\right)^{\mathrm{T}}\left\{\mathbf{S}(\rho)^{\mathrm{T}}\frac{\partial \mathbf{S}(\rho)}{\partial \rho}\right\}E\left(\mathbf{Y} \mid \widehat{\mathbf{A}}\right) \\
& -\frac{2}{n}\mathbf{b}^{\mathrm{T}}\widehat{\mathbf{A}}^{\mathrm{T}}\frac{\partial \mathbf{S}(\rho)}{\partial \rho}E\left(\mathbf{Y} \mid \widehat{\mathbf{A}}\right) \\
= \;& \frac{2}{n}\mathrm{trace}\left\{\mathbf{S}(\rho)^{\mathrm{T}}\frac{\partial \mathbf{S}(\rho)}{\partial \rho}\sigma^2\mathbf{S}^{-1}(\rho)(\mathbf{S}^{-1}(\rho))^{\mathrm{T}}\right\} + \frac{2}{n}\mathbf{b}^{\mathrm{T}}\widehat{\mathbf{A}}^{\mathrm{T}}(\mathbf{S}^{-1}(\rho))^{\mathrm{T}}\left\{\mathbf{S}(\rho)^{\mathrm{T}}\frac{\partial \mathbf{S}(\rho)}{\partial \rho}\right\}\mathbf{S}^{-1}(\rho)\widehat{\mathbf{A}}\mathbf{b} \\
& -\frac{2}{n}\mathbf{b}^{\mathrm{T}}\widehat{\mathbf{A}}^{\mathrm{T}}\frac{\partial \mathbf{S}(\rho)}{\partial \rho}\mathbf{S}^{-1}(\rho)\widehat{\mathbf{A}}\mathbf{b} \\
= \;& \frac{2\sigma^2}{n}\mathrm{trace}\left\{\frac{\partial \mathbf{S}(\rho)}{\partial \rho}\mathbf{S}^{-1}(\rho)(\mathbf{S}^{-1}(\rho))^{\mathrm{T}}\mathbf{S}(\rho)^{\mathrm{T}}\right\} + \frac{2}{n}\mathbf{b}^{\mathrm{T}}\widehat{\mathbf{A}}^{\mathrm{T}}\frac{\partial \mathbf{S}(\rho)}{\partial \rho}\mathbf{S}^{-1}(\rho)\widehat{\mathbf{A}}\mathbf{b} \\
& -\frac{2}{n}\mathbf{b}^{\mathrm{T}}\widehat{\mathbf{A}}^{\mathrm{T}}\frac{\partial \mathbf{S}(\rho)}{\partial \rho}\mathbf{S}^{-1}(\rho)\widehat{\mathbf{A}}\mathbf{b} \\
= \;& \frac{2\sigma^2}{n}\mathrm{trace}\left\{\frac{\partial \mathbf{S}(\rho)}{\partial \rho}\mathbf{S}^{-1}(\rho)\right\} \\
= \;& -\frac{2\sigma^2}{n}\mathrm{trace}\left\{\mathbf{W}(\mathbf{I}-\rho\mathbf{W})^{-1}\right\},
\end{aligned}
$$

which is not zero in general. For example, when all eigenvalues of $\rho\mathbf{W}$ satisfy $|\lambda(\rho\mathbf{W})| < 1$, we have

$$
\begin{aligned}
& \mathrm{trace}\left\{\mathbf{W}(\mathbf{I}-\rho\mathbf{W})^{-1}\right\} \\
= \;& \mathrm{trace}\left[\mathbf{W}\left\{\mathbf{I}+(\rho\mathbf{W})+(\rho\mathbf{W})^2+(\rho\mathbf{W})^3+\cdots+(\rho\mathbf{W})^k+\cdots\right\}\right] \\
= \;& \mathrm{trace}\left(\mathbf{W}+\rho\mathbf{W}^2+\rho^2\mathbf{W}^3+\rho^3\mathbf{W}^4+\cdots\right) \\
= \;& \mathrm{trace}\left(\rho\mathbf{W}^2+\rho^2\mathbf{W}^3+\rho^3\mathbf{W}^4+\cdots\right) \geqslant 0
\end{aligned}
$$

in general, where the last euality is because the diagonal elements of $\mathbf{W}$ equal to zero, and the last inequality is because each element of $\mathbf{W}$ are nonnegative, and $\rho \neq 0, \mathbf{W} \neq 0$ in general. Hence, trace$\{\mathbf{W}(\mathbf{I}-\rho\mathbf{W})^{-1}\} \neq 0$ unless there is no network structure. Based on the above results, the expectation of the score function is not equal to zero.

S3.2 Proof of Lemma S2.2

(a) Let $[k/2]$ be the largest integer less than $k/2$. Then

$$\sum_{\ell:\ell\neq k}^{\infty} \ell^{-2(\alpha+\tau)}(\lambda_k-\lambda_\ell)^{-4} = \left(\sum_{\ell=1}^{[k/2]} + \sum_{\ell=[k/2]+1,\ell\neq k}^{\ell=2k} + \sum_{\ell=2k+1}^{\infty}\right)\ell^{-2(\alpha+\tau)}(\lambda_k-\lambda_\ell)^{-4}.$$

Under Condition (C2), we note that

$$\begin{aligned}
\sum_{\ell=1}^{[k/2]} \ell^{-2(\alpha+\tau)}(\lambda_k-\lambda_\ell)^{-4} &\leqslant \sum_{\ell=1}^{[k/2]} \ell^{-2(\alpha+\tau)}[\lambda_\ell(1-\lambda_k/\lambda_{[k/2]})]^{-4} \\
&\leqslant C_1 \sum_{\ell=1}^{[k/2]} \ell^{-2(\alpha+\tau)}\lambda_\ell^{-4} \\
&\leqslant C_2 \sum_{\ell=1}^{[k/2]} \ell^{2\alpha-2\tau} \\
&\leqslant C_3 \begin{cases} 1, & \text{if } \alpha+\frac{1}{2}<\tau, \\ \log k, & \text{if } \alpha+\frac{1}{2}=\tau, \\ k^{2\alpha-2\tau+1}, & \text{if } \alpha+\frac{1}{2}>\tau, \end{cases}
\end{aligned}$$

and

$$\begin{aligned}
\sum_{\ell=2k+1}^{\infty} \ell^{-2(\alpha+\tau)}(\lambda_k-\lambda_\ell)^{-4} &\leqslant \sum_{\ell=2k+1}^{\infty} \ell^{-2(\alpha+\tau)}[\lambda_k(1-\lambda_{2k+1}/\lambda_k)]^{-4} \\
&\leqslant C_4\lambda_k^{-4} \sum_{\ell=2k+1}^{\infty} \ell^{-2(\alpha+\tau)} \\
&\leqslant C_5 k^{2\alpha-2\tau+1}.
\end{aligned}$$

Again, using Condition (C2), for $[k/2] < \ell < k$, we observe that

$$\begin{aligned}
\lambda_\ell - \lambda_k &= \sum_{t=\ell}^{k-1} (\lambda_t - \lambda_{t+1}) \\
&\geqslant C^{-1} \sum_{t=\ell}^{k-1} t^{-\alpha-1} \\
&\geqslant k^{-\alpha-1} C^{-1} (k-\ell). \qquad \text{(S.6)}
\end{aligned}$$

Similarly, for $k < \ell \leqslant 2k$, let $C_6 = C^{-1} 2^{-\alpha-1}$, then

$$\begin{aligned}
\lambda_k - \lambda_\ell &= \sum_{t=k}^{\ell-1} (\lambda_t - \lambda_{t+1}) \\
&\geqslant C^{-1} \sum_{t=k}^{\ell-1} t^{-\alpha-1} \\
&\geqslant C^{-1} \sum_{t=k}^{\ell-1} (2k)^{-\alpha-1} \\
&= C_6 k^{-\alpha-1} |\ell - k|. \qquad \text{(S.7)}
\end{aligned}$$

This implies that, for $C_7 = \max(C_6^{-4}, C^4)$, $C_8 = C_7 2^{2\alpha+2\tau}$,

$$\begin{aligned}
\sum_{\ell=[k/2]+1, \ell\neq k}^{2k} \ell^{-2(\alpha+\tau)} (\lambda_k - \lambda_\ell)^{-4} &\leqslant \sum_{\ell=[k/2]+1, \ell\neq k}^{2k} \frac{C_7 \ell^{-2(\alpha+\tau)} k^{4(\alpha+1)}}{(k-\ell)^4} \\
&\leqslant \sum_{\ell=[k/2]+1, \ell\neq k}^{2k} \frac{C_7 (k/2)^{-2(\alpha+\tau)} k^{4(\alpha+1)}}{(k-\ell)^4} \\
&= \sum_{\ell=[k/2]+1, \ell\neq k}^{2k} \frac{C_8 k^{2\alpha-2\tau+4}}{(k-\ell)^4} \\
&\leqslant C_9 k^{2\alpha-2\tau+4},
\end{aligned}$$

where the last inequality is because $\sum_{k=1}^{\infty} k^{-4}$ is a convergence sequence. By combining the above results, for $C_{10} = \max(C_3, C_5, C_9)$, we have

$$\sum_{\ell:\ell\neq k}^{\infty} \ell^{-2(\alpha+\tau)}(\lambda_k - \lambda_\ell)^{-4} \leqslant C_{10}(k^{2\alpha-2\tau+4} + \log k).$$

(b) Similar to the treatment of (a), using $-(\tau + \frac{1}{2}\alpha) < -(\alpha+1)$ obtained by $\tau > \frac{1}{2}\alpha + 1$, we have

$$\begin{aligned}
\sum_{\ell=1}^{[k/2]} \ell^{-\frac{3}{2}\alpha-\tau}|\lambda_k - \lambda_\ell|^{-1} &\leqslant \sum_{\ell=1}^{[k/2]} \ell^{-\frac{3}{2}\alpha-\tau}[\lambda_\ell(1 - \lambda_k/\lambda_{[k/2]})]^{-1} \\
&\leqslant C_{11} \sum_{\ell=1}^{[k/2]} \ell^{-\frac{3}{2}\alpha-\tau}\lambda_\ell^{-1} \\
&\leqslant C_{12} \sum_{\ell=1}^{[k/2]} \ell^{-\frac{1}{2}\alpha-\tau} \to C_{13}.
\end{aligned}$$

Again using $-(\tau + \frac{1}{2}\alpha) < -(\alpha+1)$ and $\alpha > 1$, we have

$$\begin{aligned}
\sum_{\ell=2k+1}^{\infty} \ell^{-\frac{3}{2}\alpha-\tau}|\lambda_k - \lambda_\ell|^{-1} &\leqslant \sum_{\ell=2k+1}^{\infty} \ell^{-\frac{3}{2}\alpha-\tau}[\lambda_k(1 - \lambda_{2k+1}/\lambda_k)]^{-1} \\
&\leqslant C_{14}\lambda_k^{-1} \sum_{\ell=2k+1}^{\infty} \ell^{-\frac{3}{2}\alpha-\tau} \\
&\leqslant C_{15}k^{-\frac{1}{2}\alpha-\tau+1} \\
&\leqslant C_{15}k^{-\alpha},
\end{aligned}$$

where the last term converges to zero for large k. Using (S.6) and (S.7), we have $|\lambda_k - \lambda_\ell|^{-1} \leqslant C_{16}k^{\alpha+1}|\ell - k|^{-1}$, where $C_{16} = \max(C, C_6^{-1})$, which implies that

$$\sum_{\ell=[k/2]+1,\ell\neq k}^{2k} \ell^{-\frac{3}{2}\alpha-\tau}|\lambda_k - \lambda_\ell|^{-1} \leqslant \sum_{\ell=[k/2]+1,\ell\neq k}^{2k} \frac{C_{16}\ell^{-\frac{3}{2}\alpha-\tau}k^{\alpha+1}}{|k-\ell|}$$

$$
\begin{aligned}
&\leqslant k^{\alpha+1} \sum_{\ell=[k/2]+1,\ell\neq k}^{2k} \frac{C_{16}\ell^{-\frac{3}{2}\alpha-\tau}}{|k-\ell|} \\
&\leqslant C_{17}k^{-\frac{1}{2}\alpha-\tau+1}\log k,
\end{aligned}
$$

Also, the last term converges to zero for large k since for any $x \in [1,\infty]$ and $\gamma > 0$, $\ln x/x^{\gamma} \to 0$.

(c) Since $\tau > \frac{1}{2}\alpha + 1$, we have

$$
\begin{aligned}
\sum_{\ell=1}^{[k/2]} \ell^{-\alpha-\tau}|\lambda_k-\lambda_\ell|^{-1} &\leqslant \sum_{\ell=1}^{[k/2]} \ell^{-\alpha-\tau}[\lambda_\ell(1-\lambda_k/\lambda_{[k/2]})]^{-1} \\
&\leqslant C_{18}\sum_{\ell=1}^{[k/2]} \ell^{-\tau} \to C_{19},
\end{aligned}
$$

and

$$
\begin{aligned}
\sum_{\ell=2k+1}^{\infty} \ell^{-\alpha-\tau}|\lambda_k-\lambda_\ell|^{-1} &\leqslant \sum_{\ell=2k+1}^{\infty} \ell^{-\alpha-\tau}[\lambda_k(1-\lambda_{2k+1}/\lambda_k)]^{-1} \\
&\leqslant C_{20}\lambda_k^{-1}\sum_{\ell=2k+1}^{\infty} \ell^{-\alpha-\tau} \\
&\leqslant C_{21}k^{-\tau+1},
\end{aligned}
$$

where the last term converges to zero for large k.

Similarly, we have

$$
\begin{aligned}
\sum_{\ell=[k/2]+1,\ell\neq k}^{2k} \ell^{-\alpha-\tau}|\lambda_k-\lambda_\ell|^{-1} &\leqslant \sum_{\ell=[k/2]+1,\ell\neq k}^{2k} \frac{C_{16}\ell^{-\alpha-\tau}k^{\alpha+1}}{|k-\ell|} \\
&\leqslant k^{\alpha+1}\sum_{\ell=[k/2]+1,\ell\neq k}^{2k} \frac{C_{16}\ell^{-\alpha-\tau}}{|k-\ell|} \\
&\leqslant C_{22}k^{-\tau+1}\log k.
\end{aligned}
$$

Also, the last term converges to zero for large k since $\tau > 1$. This completes the proof of Lemma S2.2. □

S3.3 Proof of Lemma S2.3

Recall that

$$\begin{aligned}\widehat{g}(t) &= n^{-1}\sum_{i=1}^{n}\left(Y_i - \rho\sum_{j=1}^{n} w_{ij}Y_j\right) Z_i(t)\\ &= n^{-1}\sum_{i=1}^{n} Y_i^* Z_i(t),\end{aligned}$$

and

$$g(t) = E\{Y_i^* Z_i(t)\},$$

where Y_i^*'s are independent and identically distributed according to (2.6). Using Condition (C1), we note that

$$\begin{aligned}\|g\| &= \left\{\int g^2(t)dt\right\}^{1/2}\\ &= \left(\int [E\{Y_i^* Z_i(t)\}]^2 dt\right)^{1/2}\\ &\leqslant \left(\int E\{Y_i^* Z_i(t)\}^2 dt\right)^{1/2}\\ &\leqslant \left(\int \{E\left(Y_i^{*4}\right)\}^{1/2}[E\left\{Z_i^4(t)\right\}]^{1/2} dt\right)^{1/2}\\ &< \left[C^{1/2}\int [E\left\{Z_i^4(t)\right\}]^{1/2} dt\right]^{1/2}\\ &< C^{3/4},\end{aligned}$$

and

$$
\begin{aligned}
E\left\|\widehat{g}(t)-g(t)\right\|^{2} &= E\int\{\widehat{g}(t)-g(t)\}^{2}\,dt \\
&= \int E\{\widehat{g}(t)-g(t)\}^{2}\,dt \\
&= \int E\left[n^{-1}\sum_{i=1}^{n}Y_i^{*}Z_i(t)-E\{Y_i^{*}Z_i(t)\}\right]^{2}dt \\
&= \int E\left(n^{-1}\sum_{i=1}^{n}[Y_i^{*}Z_i(t)-E\{Y_i^{*}Z_i(t)\}]\right)^{2}dt \\
&= \int n^{-1}E\,[Y_i^{*}Z_i(t)-E\{Y_i^{*}Z_i(t)\}]^{2}\,dt \\
&\leqslant \int n^{-1}E\,\{Y_i^{*}Z_i(t)\}^{2}\,dt \\
&\leqslant n^{-1}\int\{E\left(Y_i^{*4}\right)\}^{1/2}[E\left\{Z_i^{4}(t)\right\}]^{1/2}dt \\
&< n^{-1}C^{1/2}\int[E\left\{Z_i^{4}(t)\right\}]^{1/2}dt \\
&< n^{-1}C^{1/2}C \\
&= O\left(n^{-1}\right).
\end{aligned}
$$

Therefore,

$$
E\left\|\widehat{g}(t)-g(t)\right\|^{2} = O(n^{-1}).
$$

This leads to $E\left\|\widehat{g}(t)-g(t)\right\| = O(n^{-1/2})$. □

S3.4 Proof of Theorem 1

We use the following facts in the proof

$$\sum_{\ell=1}^{m} \ell^p \asymp \begin{cases} m^{p+1} & p > -1 \\ \log m & p = -1 \\ C & p < -1 \end{cases}$$

$$\sum_{\ell=m+1}^{\infty} m^{-p} \asymp m^{-p+1} \quad p > 1.$$

For i.i.d. case,

$$\begin{aligned} & \left(\widehat{\beta}_\rho(t) - \beta(t)\right)^2 \\ = & \left(\sum_{k=1}^{K} \widehat{b}_k(\rho)\widehat{\phi}_k(t) - \beta(t)\right)^2 \\ = & \left(\sum_{k=1}^{K} \widehat{b}_k(\rho)\widehat{\phi}_k(t) - \sum_{k=1}^{K} b_k\phi_k(t) + \sum_{k=1}^{K} b_k\phi_k(t) - \beta(t)\right)^2 \\ \leqslant & \; 2\left(\sum_{k=1}^{K} \widehat{b}_k(\rho)\widehat{\phi}_k(t) - \sum_{k=1}^{K} b_k\phi_k(t)\right)^2 + 2\left(\sum_{k=1}^{K} b_k\phi_k(t) - \beta(t)\right)^2. \end{aligned} \tag{S.8}$$

Based on Condition (C3), we note that the second term on the right-hand side of (S.8) can be shown to be

$$\begin{aligned} \int \left\{\sum_{k=1}^{K} b_k\phi_k(t) - \beta(t)\right\}^2 dt &= \int \left\{\sum_{k=1}^{K} b_k\phi_k(t) - \sum_{k=1}^{\infty} b_k\phi_k(t)\right\}^2 dt \\ &= \int \left\{\sum_{k=K+1}^{\infty} b_k\phi_k(t)\right\}^2 dt \end{aligned}$$

$$
\begin{aligned}
&= \int \sum_{k=K+1}^{\infty} b_k^2 \phi_k^2(t) dt \\
&\leqslant \sum_{k=K+1}^{\infty} k^{-2\tau} \\
&= O(K^{-2\tau+1}) \\
&= O(n^{(1-2\tau)/(\alpha+2\tau)}).
\end{aligned}
$$

Next we will show the convergence rate of the first term on the right-hand side of (S.8). Recall that $\widehat{b}_k(\rho) = \widehat{\lambda}_k^{-1}\langle\widehat{\phi}_k, \widehat{g}\rangle$. Note also that $b_k = \lambda_k^{-1}\langle\phi_k, g\rangle$.

Let $\Delta = \|\widehat{G} - G\|$ and $\Omega_K = \{\Delta \leqslant \lambda_K\}$. On the event Ω_K, using Lemma S2.4, we can see that $\sup_{k\leqslant K} |\widehat{\lambda}_k - \lambda_k| \leqslant \Delta \leqslant \lambda_K/2 \leqslant \lambda_k/2$, which implies that $2^{-1}\lambda_k \leqslant \widehat{\lambda}_k \leqslant 2\lambda_k$. Moreover, note that Δ is nonnegative, Lemma S2.1(b) yields $\Delta = O_p(n^{-1/2})$. Condition (C2) yields $\lambda_K = O\{n^{-\alpha/(\alpha+2\tau)}\}$. Conditions (C3) leads to that $\mathrm{pr}(\Omega_K) \to 1$. Hence it suffices to work with bounds that are established under the event Ω_K. We have

$$
\begin{aligned}
& \int \left(\sum_{k=1}^{K} \widehat{b}_k(\rho)\widehat{\phi}_k(t) - \sum_{k=1}^{K} b_k\phi_k(t) \right)^2 dt \\
= & \int \left(\sum_{k=1}^{K} \widehat{\lambda}_k^{-1}(\langle\widehat{\phi}_k, \widehat{g}\rangle - \langle\phi_k, g\rangle)\widehat{\phi}_k(t) + \sum_{k=1}^{K}(\widehat{\lambda}_k^{-1} - \lambda_k^{-1})\langle\phi_k, g\rangle\widehat{\phi}_k(t) \right. \\
& \left. + \sum_{k=1}^{K} \lambda_k^{-1}\langle\phi_k, g\rangle\{\widehat{\phi}_k(t) - \phi_k(t)\} \right)^2 dt \\
\leqslant & \ 3\sum_{k=1}^{K} \widehat{\lambda}_k^{-2}(\langle\widehat{\phi}_k, \widehat{g}\rangle - \langle\phi_k, g\rangle)^2 + 3\sum_{k=1}^{K}(\widehat{\lambda}_k^{-1} - \lambda_k^{-1})^2\langle\phi_k, g\rangle^2 \\
& +3K\sum_{k=1}^{K} \lambda_k^{-2}\langle\phi_k, g\rangle^2\|\widehat{\phi}_k(t) - \phi_k(t)\|^2 \\
\leqslant & \ 36\sum_{k=1}^{K} \lambda_k^{-2}\left(\left\langle\phi_k, \widehat{g} - g\right\rangle^2 + \left\langle\widehat{\phi}_k - \phi_k, g\right\rangle^2 + \left\langle\widehat{\phi}_k - \phi_k, \widehat{g} - g\right\rangle^2\right) \\
& +3\sum_{k=1}^{K}(\widehat{\lambda}_k^{-1} - \lambda_k^{-1})^2\left\langle\phi_k, g\right\rangle^2 + 3K\sum_{k=1}^{K} \lambda_k^{-2}\langle\phi_k, g\rangle^2\|\widehat{\phi}_k(t) - \phi_k(t)\|^2
\end{aligned}
$$

$$
\begin{aligned}
\leqslant\; & 36\sum_{k=1}^{K}\lambda_k^{-2}\big\langle\phi_k,\widehat{g}-g\big\rangle^2+36\sum_{k=1}^{K}\lambda_k^{-2}\big\langle\widehat{\phi}_k-\phi_k,g\big\rangle^2\\
& +36\sum_{k=1}^{K}\left(\lambda_k^{-2}\|\widehat{g}-g\|^2+Kb_k^2\right)\|\widehat{\phi}_k(t)-\phi_k(t)\|^2+3\sum_{k=1}^{K}(\widehat{\lambda}_k^{-1}-\lambda_k^{-1})^2\big\langle\phi_k,g\big\rangle^2. \quad \text{(S.9)}
\end{aligned}
$$

We now treat each term in (S.9). To handle the first term in (S.9), we note that

$$
\begin{aligned}
g(t) &= E\left\{Y_i^* Z_i(t)\right\}\\
&= E\left[\left\{\int\beta(t)Z_i(t)dt\right\}\left\{\sum_{k'=1}^{\infty}\xi_{ik'}\phi_{k'}(t)\right\}\right]\\
&= \sum_{k'=1}^{\infty}E\left\{\int\beta(t)Z_i(t)dt\xi_{ik'}\right\}\phi_{k'}(t).
\end{aligned}
$$

Then by the orthonormality of $\{\phi_k(t)\}_{k=1,\ldots}$, it yields that

$$
\begin{aligned}
\int g(t)\phi_k(t)dt &= \int\sum_{k'=1}^{\infty}E\left\{\int\beta(s)Z_i(s)ds\xi_{ik'}\right\}\phi_{k'}(t)\phi_k(t)dt\\
&= E\left\{\int\beta(t)Z_i(t)dt\xi_{ik}\right\}.
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
\int\widehat{g}(t)\phi_k(t)dt &= \int\left\{n^{-1}\sum_{i=1}^{n}Y_i^*Z_i(t)\right\}\phi_k(t)dt\\
&= n^{-1}\sum_{i=1}^{n}\int\left\{\int\beta(t)Z_i(t)dt+\varepsilon_i\right\}\left\{\sum_{k'=1}^{\infty}\xi_{ik'}\phi_{k'}(t)\right\}\phi_k(t)dt\\
&= n^{-1}\sum_{i=1}^{n}\left\{\int\beta(t)Z_i(t)dt+\varepsilon_i\right\}\xi_{ik}.
\end{aligned}
$$

Noting that

$$
E\left(\langle\phi_k,\widehat{g}-g\rangle\right)
$$

$$
\begin{aligned}
&= E\left[n^{-1}\sum_{i=1}^{n}\left\{\int\beta(t)Z_i(t)dt+\varepsilon_i\right\}\xi_{ik}-E\left\{\int\beta(t)Z_i(t)dt\xi_{ik}\right\}\right] \\
&= 0.
\end{aligned}
$$

Now, using Condition (C1), we have

$$
\begin{aligned}
& E\left(\langle\phi_k,\widehat{g}-g\rangle^2\right) \\
= \; & \operatorname{var}\left[n^{-1}\sum_{i=1}^{n}\left\{\int\beta(t)Z_i(t)dt+\varepsilon_i\right\}\xi_{ik}-E\left\{\int\beta(t)Z_i(t)dt\xi_{ik}\right\}\right] \\
= \; & n^{-1}\left[\operatorname{var}\left\{\int\beta(t)Z_i(t)dt\xi_{ik}\right\}+\operatorname{var}\left(\varepsilon_i\xi_{ik}\right)\right] \\
\leqslant \; & n^{-1}\left(\left[E\left\{\int\beta(t)Z_i(t)dt\right\}^4\right]^{1/2}\left(E\xi_{ik}^4\right)^{1/2}+\left(E\varepsilon_i^4\right)^{1/2}\left(E\xi_{ik}^4\right)^{1/2}\right) \\
\leqslant \; & Cn^{-1}\lambda_k.
\end{aligned}
$$

This implies that

$$
\begin{aligned}
\sum_{k=1}^{K}\lambda_k^{-2}\langle\phi_k,\widehat{g}-g\rangle^2 &= O_p(n^{-1}\sum_{k=1}^{K}\lambda_k^{-1}) \\
&= O_p(n^{-1}\sum_{k=1}^{K}k^{\alpha}) \\
&= O_p(n^{-(2\tau-1)/(\alpha+2\tau)}).
\end{aligned}
$$

Next we deal with the second term in (S.9). Write $g(t)=\sum_{k=1}^{\infty}g_k\phi_k(t)$, where $g_k=\langle g,\phi_k\rangle=\lambda_k b_k$. Using (5.4) in Hall and Horowitz (2007), we have

$$
\widehat{\phi}_k(t)-\phi_k(t)=\sum_{\ell:\ell\neq k}^{\infty}(\widehat{\lambda}_k-\lambda_\ell)^{-1}\phi_\ell(t)\langle\widehat{\phi}_k,\langle\widehat{G}-G,\phi_\ell\rangle\rangle+\phi_k(t)\langle\widehat{\phi}_k-\phi_k,\phi_k\rangle.
$$

Hence, combing the orthonormality of $\{\phi_k(t)\}_{k=1,\dots}$, we deduce that

$$
\begin{aligned}
& \left\langle \widehat{\phi}_k - \phi_k, g \right\rangle \\
= & \int \left\{ \sum_{\ell:\ell\neq k}^{\infty} (\widehat{\lambda}_k - \lambda_\ell)^{-1} \phi_\ell(t) \langle \widehat{\phi}_k, \langle \widehat{G} - G, \phi_\ell \rangle \rangle + \phi_k(t) \langle \widehat{\phi}_k - \phi_k, \phi_k \rangle \right\} \left\{ \sum_{k'=1}^{\infty} g_{k'} \phi_{k'}(t) \right\} dt \\
= & \sum_{\ell:\ell\neq k}^{\infty} g_\ell (\widehat{\lambda}_k - \lambda_\ell)^{-1} \langle \widehat{\phi}_k, \langle \widehat{G} - G, \phi_\ell \rangle \rangle + g_k \langle \widehat{\phi}_k - \phi_k, \phi_k \rangle \\
= & R_{k1} + R_{k2} + R_{k3} + R_{k4},
\end{aligned}
\tag{S.10}
$$

where

$$
\begin{aligned}
R_{k1} &= \sum_{\ell:\ell\neq k}^{\infty} \lambda_\ell b_\ell \{ (\widehat{\lambda}_k - \lambda_\ell)^{-1} - (\lambda_k - \lambda_\ell)^{-1} \} \langle \phi_k, \langle \widehat{G} - G, \phi_\ell \rangle \rangle, \\
R_{k2} &= \sum_{\ell:\ell\neq k}^{\infty} \lambda_\ell b_\ell (\lambda_k - \lambda_\ell)^{-1} \langle \phi_k, \langle \widehat{G} - G, \phi_\ell \rangle \rangle, \\
R_{k3} &= \sum_{\ell:\ell\neq k}^{\infty} \lambda_\ell b_\ell (\widehat{\lambda}_k - \lambda_\ell)^{-1} \langle \widehat{\phi}_k - \phi_k, \langle \widehat{G} - G, \phi_\ell \rangle \rangle, \\
R_{k4} &= \lambda_k b_k \langle \widehat{\phi}_k - \phi_k, \phi_k \rangle.
\end{aligned}
$$

Next we will first focus on the convergence rate of R_{k1}. Define the set of realizations such that, for constant C' and sample size n,

$$
\mathcal{F}_K = \left\{ 2^{-1} (\widehat{\lambda}_k - \lambda_\ell)^{-2} \leqslant (\lambda_k - \lambda_\ell)^{-2} \leqslant C' n^{2(\alpha+1)/(\alpha+2\tau)},\ k, \ell = 1, \dots, K,\ k \neq \ell \right\}.
$$

We note that, by Lemma S2.1,

$$
\begin{aligned}
|\widehat{\lambda}_k - \lambda_\ell| &= |\widehat{\lambda}_k - \lambda_k + \lambda_k - \lambda_\ell| \\
&= |O_p(n^{-1/2}) + \lambda_k - \lambda_\ell| \\
&= |\lambda_k - \lambda_\ell| \{1 + o_p(1)\},
\end{aligned}
$$

hence $2^{-1}(\widehat{\lambda}_k - \lambda_\ell)^{-2} \leqslant (\lambda_k - \lambda_\ell)^{-2}$. Further,

$$\begin{aligned}(\lambda_k - \lambda_\ell)^2 &\geqslant \min\{(\lambda_{k-1} - \lambda_k)^2, (\lambda_k - \lambda_{k+1})^2\} \\ &\geqslant C^{-2} k^{-2(\alpha+1)} \\ &\geqslant C^{-2} K^{-2(\alpha+1)},\end{aligned}$$

hence $(\lambda_k - \lambda_\ell)^{-2} \leqslant C^2 K^{2(\alpha+1)} \leqslant C' n^{\frac{2(\alpha+1)}{\alpha+2\tau}}$ for sufficiently large n. Thus, we have $P(\mathcal{F}_K) \to 1$ as $n \to \infty$.

Hence it suffices to work with bounds that are established under the event $\mathcal{F}_K$. Using conditions (C2) and (C3),

$$\begin{aligned}R_{k1}^2 &= \left\{\sum_{\ell:\ell\neq k} \lambda_\ell b_\ell \frac{\widehat{\lambda}_k - \lambda_k}{(\widehat{\lambda}_k - \lambda_\ell)(\lambda_k - \lambda_\ell)} \langle \phi_k, \langle \widehat{G} - G, \phi_\ell \rangle\rangle\right\}^2 \\ &\leqslant \left\{\sum_{\ell:\ell\neq k} \lambda_\ell^2 b_\ell^2 \frac{(\widehat{\lambda}_k - \lambda_k)^2}{(\widehat{\lambda}_k - \lambda_\ell)^2(\lambda_k - \lambda_\ell)^2}\right\} \left\{\sum_{\ell:\ell\neq k} \langle \phi_k, \langle \widehat{G} - G, \phi_\ell \rangle\rangle^2\right\} \\ &\leqslant \left\{\sum_{\ell:\ell\neq k} 2C^4 \ell^{-2(\alpha+\tau)} \frac{(\widehat{\lambda}_k - \lambda_k)^2}{(\lambda_k - \lambda_\ell)^4}\right\} \left\{\sum_{\ell=1}^{\infty} \langle \phi_k, \langle \widehat{G} - G, \phi_\ell \rangle\rangle^2\right\} \\ &= 2C^4 (\widehat{\lambda}_k - \lambda_k)^2 \sum_{\ell:\ell\neq k} \ell^{-2(\alpha+\tau)} (\lambda_k - \lambda_\ell)^{-4} \|\langle \widehat{G} - G, \phi_k \rangle\|^2,\end{aligned}$$

where the last equation is because Parseval's identity asserts that for any x and orthonormal basis e_n in Hilbert space, $\sum_n |\langle x, e_n \rangle|^2 = \|x\|^2$. Based on Lemma S2.1 and Lemma S2.2, we have

$$\begin{aligned}\sum_{k=1}^{K} \lambda_k^{-2} R_{k1}^2 &\leqslant 2C^4 \sum_{k=1}^{K} \lambda_k^{-2} (\widehat{\lambda}_k - \lambda_k)^2 \|\langle \widehat{G} - G, \phi_k \rangle\|^2 \sum_{\ell:\ell\neq k} \ell^{-2(\alpha+\tau)} (\lambda_k - \lambda_\ell)^{-4} \\ &\leqslant 2C^6 \sum_{k=1}^{K} (\widehat{\lambda}_k - \lambda_k)^2 \|\langle \widehat{G} - G, \phi_k \rangle\|^2 k^{2\alpha} \sum_{\ell:\ell\neq k} \ell^{-2(\alpha+\tau)} (\lambda_k - \lambda_\ell)^{-4}\end{aligned}$$

$$
\begin{aligned}
&\leqslant C_1 \sum_{k=1}^{K} (\widehat{\lambda}_k - \lambda_k)^2 \|\langle \widehat{G} - G, \phi_k \rangle\|^2 \left(k^{2\alpha}\log n + k^{4\alpha - 2\tau + 4}\right) \\
&= O_p\left\{ n^{-1} \sum_{k=1}^{K} E(\|\langle \widehat{G} - G, \phi_k\rangle\|^2)(k^{2\alpha}\log n + k^{4\alpha-2\tau+4}) \right\} \\
&= O_p\left\{ n^{-2} \sum_{k=1}^{K} (k^{2\alpha}\log n + k^{4\alpha-2\tau+4}) \right\} \\
&= O_p\left\{ n^{-2}\left(K^{2\alpha+1}\log n + K^{4\alpha-2\tau+5}\right)\right\} \\
&= o_p\{n^{-(2\tau-1)/(\alpha+2\tau)}\}.
\end{aligned}
$$

In the following, we use Lemma S2.1 and the assumption $\tau > \alpha/2 + 1$ in (C3) repetitively to show the similar result for R_{k2}, R_{k3} and R_{k4} as for R_{k1}. For R_{k2}, observe that for all ℓ, k, because

$$
E(\xi_k \xi_\ell) = \iint G(s,t)\phi_k(s)\phi_\ell(t)dsdt = \begin{cases} \lambda_k, & \text{if } k = \ell, \\ 0, & \text{if } k \neq \ell, \end{cases}
$$

for $\ell \neq k$,

$$
\begin{aligned}
&\iint \left\{ \widehat{G}(s,t) - G(s,t) \right\} \phi_k(s)\phi_\ell(t)dsdt \\
&= \iint \left[\frac{1}{n}\sum_{i=1}^{n} Z_i(s)Z_i(t) - G(s,t)\right] \phi_k(s)\phi_\ell(t)dsdt \\
&= \frac{1}{n}\sum_{i=1}^{n} \xi_{ik}\xi_{i\ell}.
\end{aligned}
$$

Thus,

$$
R_{k2} = n^{-1} \sum_{i=1}^{n} \sum_{\ell:\ell\neq k}^{\infty} \lambda_\ell b_\ell (\lambda_k - \lambda_\ell)^{-1} \xi_{ik}\xi_{i\ell},
$$

and $ER_{k2}=0$, we can get

$$\begin{aligned}
nE(R_{k2}^2) &\leqslant 2E\left\{\xi_{ik}\sum_{\ell:\ell\neq k}^{\infty}\lambda_\ell b_\ell(\lambda_k-\lambda_\ell)^{-1}\xi_{i\ell}\right\}^2\\
&\leqslant 2(E\xi_k^4)^{1/2}\left[E\left\{\sum_{\ell:\ell\neq k}^{\infty}\lambda_\ell b_\ell(\lambda_k-\lambda_\ell)^{-1}\xi_\ell\right\}^4\right]^{1/2},
\end{aligned}$$

where $(E\xi_k^4)^{1/2}\leqslant C^{3/2}k^{-\alpha}$ by condition (C1). It is easy to see that

$$\begin{aligned}
&E\left\{\sum_{\ell:\ell\neq k}^{\infty}\lambda_\ell b_\ell(\lambda_k-\lambda_\ell)^{-1}\xi_\ell\right\}^4\\
\leqslant\ & \sum_{\ell_1:\ell_1\neq k}\cdots\sum_{\ell_4:\ell_4\neq k}\left|\lambda_{\ell_1}b_{\ell_1}(\lambda_k-\lambda_{\ell_1})^{-1}\right|\cdots\left|\lambda_{\ell_4}b_{\ell_4}(\lambda_k-\lambda_{\ell_4})^{-1}\right|E(|\xi_{\ell_1}\xi_{\ell_2}\xi_{\ell_3}\xi_{\ell_4}|)\\
\leqslant\ & C^8\sum_{\ell_1:\ell_1\neq k}\cdots\sum_{\ell_4:\ell_4\neq k}\left|\ell_1^{-(\alpha+\tau)}(\lambda_k-\lambda_{\ell_1})^{-1}\right|\cdots\left|\ell_4^{-(\alpha+\tau)}(\lambda_k-\lambda_{\ell_4})^{-1}\right|E\left(|\xi_{\ell_1}\xi_{\ell_2}\xi_{\ell_3}\xi_{\ell_4}|\right),
\end{aligned}$$

and repeated application of Hölder's inequality yields that

$$\begin{aligned}
E(|\xi_{\ell_1}\xi_{\ell_2}\xi_{\ell_3}\xi_{\ell_4}|) &\leqslant \left\{E\left|\xi_{\ell_1}\xi_{\ell_2}\xi_{\ell_3}\right|^{4/3}\right\}^{3/4}(E\xi_{\ell_4}^4)^{1/4}\\
&\leqslant ([E\{(\xi_{\ell_1}\xi_{\ell_2})^{4/3}\}^{3/2}]^{2/3}\{E(\xi_{\ell_3}^{4/3})^3\}^{1/3})^{3/4}(E\xi_{\ell_4}^4)^{1/4}\\
&\leqslant \{E(\xi_{\ell_1}\xi_{\ell_2})^2\}^{1/2}(E\xi_{\ell_3}^4)^{1/4}(E\xi_{\ell_4}^4)^{1/4}\\
&\leqslant \left(E\xi_{\ell_1}^4\right)^{1/4}\cdots\left(E\xi_{\ell_4}^4\right)^{1/4}\\
&\leqslant C^3\ell_1^{-\alpha/2}\cdots\ell_4^{-\alpha/2}.
\end{aligned}$$

Hence we have

$$\begin{aligned}
&(E\xi_k^4)^{1/2}\left[E\left\{\sum_{\ell:\ell\neq k}^{\infty}\lambda_\ell b_\ell(\lambda_k-\lambda_\ell)^{-1}\xi_\ell\right\}^4\right]^{1/2}\\
\leqslant\ & C^{3/2}k^{-\alpha}\left\{C^8\sum_{\ell_1:\ell_1\neq k}\cdots\sum_{\ell_4:\ell_4\neq k}\left|\ell_1^{-(\alpha+\tau)}(\lambda_k-\lambda_{\ell_1})^{-1}\right|\cdots\left|\ell_4^{-(\alpha+\tau)}(\lambda_k-\lambda_{\ell_4})^{-1}\right|E\left(|\xi_{\ell_1}\xi_{\ell_2}\xi_{\ell_3}\xi_{\ell_4}|\right)\right\}^{1/2}
\end{aligned}$$

$$
\begin{aligned}
&\leqslant C^{3/2}k^{-\alpha}\left\{C^{11}\sum_{\ell_1:\ell_1\neq k}\cdots\sum_{\ell_4:\ell_4\neq k}\left|\ell_1^{-(\alpha+\tau)}(\lambda_k-\lambda_{\ell_1})^{-1}\right|\cdots\left|\ell_4^{-(\alpha+\tau)}(\lambda_k-\lambda_{\ell_4})^{-1}\right|\ell_1^{-\alpha/2}\cdots\ell_4^{-\alpha/2}\right\}^{1/2}\\
&\leqslant C^7k^{-\alpha}\left\{\left(\sum_{\ell:\ell\neq k}\frac{\ell^{-\frac{3}{2}\alpha-\tau}}{|\lambda_k-\lambda_\ell|}\right)^4\right\}^{1/2}\\
&= C^7k^{-\alpha}\left(\sum_{\ell:\ell\neq k}\frac{\ell^{-\frac{3}{2}\alpha-\tau}}{|\lambda_k-\lambda_\ell|}\right)^2\\
&= O(k^{-\alpha}).
\end{aligned}
$$

The last equality follows from Lemma S2.2. We conclude that $E(R_{k2}^2)=O(n^{-1}k^{-\alpha})$, which implies that

$$
\begin{aligned}
\sum_{k=1}^{K}\lambda_k^{-2}R_{k2}^2 &= O_p(n^{-1}K^{\alpha+1})\\
&= O_p(n^{-(2\tau-1)/(\alpha+2\tau)}).
\end{aligned}
$$

It is seen that under $\mathcal{F}_K$,

$$
\begin{aligned}
R_{k3} &\leqslant \sum_{\ell:\ell\neq k}^{\infty}\lambda_\ell b_\ell|\widehat{\lambda}_k-\lambda_\ell|^{-1}\langle\widehat{\phi}_k-\phi_k,\langle\widehat{G}-G,\phi_\ell\rangle\rangle\\
&\leqslant \sqrt{2}\sum_{\ell:\ell\neq k}^{\infty}\lambda_\ell b_\ell|\lambda_k-\lambda_\ell|^{-1}\iint\left\{\widehat{G}(s,t)-G(s,t)\right\}\left\{\widehat{\phi}_k(s)-\phi_k(s)\right\}ds\phi_\ell(t)dt\\
&\leqslant \sqrt{2}C^2\sum_{\ell:\ell\neq k}^{\infty}\ell^{-(\alpha+\tau)}|\lambda_k-\lambda_\ell|^{-1}\\
&\qquad\times\int\left[\int\left\{\widehat{G}(s,t)-G(s,t)\right\}^2ds\right]^{1/2}\left[\int\left\{\widehat{\phi}_k(s)-\phi_k(s)\right\}^2ds\right]^{1/2}\phi_\ell(t)dt\\
&\leqslant \sqrt{2}C^2\sum_{\ell:\ell\neq k}^{\infty}\ell^{-(\alpha+\tau)}|\lambda_k-\lambda_\ell|^{-1}\left[\int\left\{\widehat{\phi}_k(s)-\phi_k(s)\right\}^2ds\right]^{1/2}\\
&\qquad\times\int\left[\int\left\{\widehat{G}(s,t)-G(s,t)\right\}^2ds\right]^{1/2}\phi_\ell(t)dt\\
&\leqslant \sqrt{2}C^2\sum_{\ell:\ell\neq k}^{\infty}\ell^{-(\alpha+\tau)}|\lambda_k-\lambda_\ell|^{-1}\|\widehat{\phi}_k-\phi_k\|
\end{aligned}
$$

$$
\begin{aligned}
& \times \left[\int\int \left\{ \widehat{G}(s,t) - G(s,t) \right\}^2 dsdt \right]^{1/2} \left\{ \int \phi_\ell^2(t) dt \right\}^{1/2} \\
= \;& \sqrt{2} C^2 \sum_{\ell:\ell\neq k}^{\infty} \ell^{-(\alpha+\tau)} |\lambda_k - \lambda_\ell|^{-1} \|\widehat{\phi}_k - \phi_k\| \|\widehat{G} - G\| \\
\leqslant \;& C_2 \|\widehat{\phi}_k - \phi_k\| \|\widehat{G} - G\|,
\end{aligned}
$$

Here the last equality is due to Lemma S2.2.

Then we have

$$
\begin{aligned}
\sum_{k=1}^{K} \lambda_k^{-2} R_{k3}^2 &\leqslant C^2 \sum_{k=1}^{K} k^{2\alpha} 2 C^4 C_1^2 \|\widehat{\phi}_k - \phi_k\|^2 \|\widehat{G} - G\|^2 \\
&= O_p\left(n^{-2} \sum_{k=1}^{K} k^{2\alpha+2} \right) \\
&= O_p\left(n^{\frac{-4\tau+3}{\alpha+2\tau}} \right) \\
&= o_p(n^{-(2\tau-1)/(\alpha+2\tau)}).
\end{aligned}
$$

Similarly, by (C3),

$$
\begin{aligned}
\sum_{k=1}^{K} \lambda_k^{-2} R_{k4}^2 &= \sum_{k=1}^{K} b_k^2 \langle \widehat{\phi}_k - \phi_k, \phi_k \rangle^2 \\
&\leqslant \sum_{k=1}^{K} b_k^2 \|\widehat{\phi}_k - \phi_k\|^2 \|\phi_k\|^2 \\
&\leqslant \sum_{k=1}^{K} C^2 k^{-2\tau} \|\widehat{\phi}_k - \phi_k\|^2 \\
&= O_p\left(n^{-1} \sum_{k=1}^{K} k^{-2(\tau-1)} \right) \\
&= O_p\left(n^{-1} \right) = o_p(n^{-(2\tau-1)/(\alpha+2\tau)}).
\end{aligned}
$$

The last equality obtained by $\sum_{k=1}^{K} k^{-2(\tau-1)}$ is a convergence sequence. Combining the results regarding $R_{k1}, R_{k2}, R_{k3}, R_{k4}$, we get

$$\sum_{k=1}^{K} \lambda_k^{-2} \left(\left\langle \widehat{\phi}_k - \phi_k, g \right\rangle^2 \right) = O_p \left(n^{-(2\tau-1)/(\alpha+2\tau)} \right).$$

We next consider the third term in (S.9). By combining Conditions (C2), (C3), Lemma S2.1 and invoking Lemma S2.3, we have

$$\begin{aligned} \sum_{k=1}^{K} \left(\lambda_k^{-2} \|\widehat{g} - g\|^2 + K b_k^2 \right) \|\widehat{\phi}_k - \phi_k\|^2 &\leqslant C^2 \sum_{k=1}^{K} (k^{2\alpha} n^{-1} + K k^{-2\tau}) \|\widehat{\phi}_k - \phi_k\|^2 \\ &= O_p(K^{2\alpha+3} n^{-2} + K n^{-1}) \\ &= o_p(n^{-(2\tau-1)/(\alpha+2\tau)}). \end{aligned}$$

For the last term of (S.9), given Condition (C2), we have $\delta_k^{-1} \leqslant C k^{\alpha+1}$, where $\delta_k = \min_{1 \leqslant \ell \leqslant k} (\lambda_\ell - \lambda_{\ell+1})$. Provided that event Ω_K holds, according to (5.7) in Hall and Horowitz (2007), we have

$$\left| \widehat{\lambda}_k - \lambda_k \right| \leqslant \left| \iint \left(\widehat{G}(s,t) - G(s,t) \right) \phi_k(s) \phi_k(t) ds dt \right| + \delta_k^{-1} \Delta \left\{ \Delta + \|(\widehat{G} - G)\phi_k\| \right\}.$$

Combing above results and Lemma S2.1, repeatedly using Conditions (C2) and (C3), we obtain,

$$\begin{aligned} \sum_{k=1}^{K} (\widehat{\lambda}_k^{-1} - \lambda_k^{-1})^2 \left\langle \phi_k, g \right\rangle^2 &\leqslant \sum_{k=1}^{K} \left(\frac{\widehat{\lambda}_k - \lambda_k}{\widehat{\lambda}_k \lambda_k} \right)^2 \left\langle \phi_k, g \right\rangle^2 \\ &\leqslant 4 \sum_{k=1}^{K} \lambda_k^{-2} b_k^2 (\widehat{\lambda}_k - \lambda_k)^2 \\ &\leqslant 8 \sum_{k=1}^{K} \lambda_k^{-2} b_k^2 \left\{ \iint \left(\widehat{G}(s,t) - G(s,t) \right) \phi_k(s) \phi_k(t) ds dt \right\}^2 \\ &\quad + 16 \Delta^2 \sum_{k=1}^{K} \delta_k^{-2} \lambda_k^{-2} b_k^2 \left\{ \Delta^2 + \|(\widehat{G} - G)\phi_k\|^2 \right\} \end{aligned}$$

$$
\begin{aligned}
&= O_p\left\{n^{-1}\sum_{k=1}^{K}\lambda_k^{-2}b_k^2\lambda_k^2 + n^{-2}\sum_{k=1}^{K}\delta_k^{-2}\lambda_k^{-2}b_k^2\right\} \\
&= O_p\left\{n^{-1}\sum_{k=1}^{K}k^{-2\tau} + n^{-2}\sum_{k=1}^{K}k^{4\alpha-2\tau+2}\right\} \\
&= O_p(n^{-1}K + n^{-2}K^{4\alpha-2\tau+3}) \\
&= o_p\left(n^{-(2\tau-1)/(\alpha+2\tau)}\right).
\end{aligned}
$$

This completes the proof of Theorem 1. □

S3.5 Proof of Lemma S2.4

Part (a) is the results from Bhatia et al. (1983).

(b) Under $k = o(n^{\frac{1}{2\alpha+2}})$, according to the result of (S.1) in Wong et al. (2019), we have

$$
\widehat{\lambda}_k - \lambda_k = \left\langle \left(\widehat{G} - G\right)\phi_k, \phi_k \right\rangle + O_p(k/n), \tag{S.11}
$$

$$
\widehat{\xi}_{ik} - \xi_{ik} = \sum_{\ell\neq k}(\lambda_k - \lambda_\ell)^{-1}\xi_{ik}\left\langle \left(\widehat{G} - G\right)\phi_\ell, \phi_k \right\rangle \times \{1 + o_p(1)\},
$$

and

$$
\widehat{\phi}_k(t) - \phi_k(t) = \left\{\sum_{\ell\neq k}(\lambda_k - \lambda_\ell)^{-1}\phi_\ell\left\langle \left(\widehat{G} - G\right)\phi_\ell, \phi_k \right\rangle\right\} \times \left\{1 + O_p(n^{-1/2}\delta_k'^{-1})\right\}, \tag{S.12}
$$

where $\delta_k' = \frac{1}{2}\min_{k'\neq k}|\lambda_{k'} - \lambda_k|$, which is no less than $\frac{1}{2}C^{-1}k^{-\alpha-1}$ under condition (C2). Let

$\Lambda_{nk} = \sum_{\ell \neq k}(\lambda_k - \lambda_\ell)^{-1}\phi_\ell \langle (\widehat{G} - G)\phi_\ell, \phi_k \rangle$. Note that for $k \neq \ell$,

$$\iint \left\{ \widehat{G}(s,t) - G(s,t) \right\} \phi_k(s)\phi_\ell(t) ds dt = \frac{1}{n}\sum_{i=1}^{n} \xi_{i\ell}\xi_{ik} - E(\xi_{i\ell}\xi_{ik}) = \frac{1}{n}\sum_{i=1}^{n} \xi_{i\ell}\xi_{ik},$$

and

$$E\left\{ \frac{1}{n}\sum_{i=1}^{n} \xi_{i\ell}\xi_{ik} \right\}^2 = \frac{1}{n}E(\xi_{i\ell}^2\xi_{ik}^2) \leqslant \frac{1}{n}\{E(\xi_{i\ell}^4)E(\xi_{ik}^4)\}^{1/2} \leqslant C\frac{1}{n}\lambda_\ell\lambda_k,$$

where the last inequality is due to Condition (C1). Thus,

$$E\left\{ \langle (\widehat{G} - G)\phi_\ell, \phi_k \rangle \right\}^2 = O(n^{-1}\lambda_\ell\lambda_k).$$

We have, by Condition (C4),

$$\Lambda_{nk} = O_p\left\{ n^{-1/2}\sum_{\ell \neq k}(\lambda_k - \lambda_\ell)^{-1}\phi_\ell\lambda_\ell^{1/2}\lambda_k^{1/2} \right\} = O_p\left\{ n^{-1/2}\sum_{\ell \neq k}(\lambda_k - \lambda_\ell)^{-1}\lambda_\ell^{1/2}\lambda_k^{1/2} \right\}.$$

Write

$$\begin{aligned} & \sum_{\ell \neq k}(\lambda_k - \lambda_\ell)^{-1}\lambda_\ell^{1/2}\lambda_k^{1/2} \\ = & \left(\sum_{\ell=1}^{[k/2]} + \sum_{\ell=[k/2]+1}^{k-1} + \sum_{\ell=k+1}^{[3k/2]} + \sum_{\ell=[3k/2]+1}^{2k} + \sum_{\ell=2k+1}^{\infty} \right)(\lambda_k - \lambda_\ell)^{-1}\lambda_\ell^{1/2}\lambda_k^{1/2}. \end{aligned} \tag{S.13}$$

To examine the first summation in (S.13), by Condition (C2), we have

$$\begin{aligned} \lambda_{k/2} - \lambda_k &= \lambda_{k/2} - \lambda_{k/2+1} + \cdots + \lambda_{k-1} - \lambda_k \\ &\geqslant C^{-1}(k/2)^{-(\alpha+1)} + \cdots + C^{-1}(k-1)^{-(\alpha+1)} \\ &\geqslant C^{-1}(k/2)k^{-(\alpha+1)} \end{aligned}$$

$$= \quad C^{-1}\frac{1}{2}k^{-\alpha}.$$

and

$$C^{-1}(k/2)^{\alpha} \leqslant 1/\lambda_{k/2} \leqslant C(k/2)^{\alpha}.$$

This leads to

$$1 - \lambda_k/\lambda_{k/2} \geqslant C^{-1}\frac{1}{2}k^{-\alpha}C^{-1}(k/2)^{\alpha} = 2^{-\alpha-1}C^{-2}.$$

Then, under Condition (C2),

$$\begin{aligned}
\sum_{\ell=1}^{[k/2]}(\lambda_k - \lambda_\ell)^{-1}\lambda_\ell^{1/2}\lambda_k^{1/2} &= \sum_{\ell=1}^{[k/2]}\frac{\lambda_\ell^{1/2}\lambda_k^{1/2}}{\lambda_\ell(1-\lambda_k/\lambda_\ell)} \\
&\leqslant \sum_{\ell=1}^{[k/2]}\frac{\lambda_\ell^{1/2}\lambda_k^{1/2}}{\lambda_\ell(1-\lambda_k/\lambda_{k/2})} \\
&\leqslant 2^{1+\alpha}C^2\sum_{\ell=1}^{[k/2]}\lambda_\ell^{-1/2}\lambda_k^{1/2} \\
&\leqslant 2^{1+\alpha}C^3k^{-\alpha/2}\sum_{\ell=1}^{[k/2]}\ell^{\alpha/2} \\
&\leqslant C_2k,
\end{aligned}$$

for some constant C_2.

For the second summation in (S.13), under Condition (C2), $\lambda_\ell - \lambda_k \geqslant C^{-1}k^{-\alpha-1}(k-\ell)$ for $\ell < k$, and

$$\sum_{\ell=[k/2]+1}^{k-1}|\lambda_k - \lambda_\ell|^{-1}\lambda_\ell^{1/2}\lambda_k^{1/2} \leqslant \sum_{\ell=[k/2]+1}^{k-1}Ck^{\alpha+1}\frac{C^{1/2}k^{-\alpha/2}C^{1/2}\ell^{-\alpha/2}}{k-\ell}$$

$$
\begin{aligned}
&= C^2 \sum_{\ell=[k/2]+1}^{k-1} k^{\alpha+1} \frac{k^{-\alpha/2}\ell^{-\alpha/2}}{k-\ell} \\
&\leqslant C^2 \sum_{\ell=[k/2]+1}^{k-1} k^{\alpha+1} \frac{k^{-\alpha/2}(k/2)^{-\alpha/2}}{k-\ell} \\
&= C^2 2^{\alpha/2} \sum_{\ell=[k/2]+1}^{k-1} k/(k-\ell) \\
&\leqslant C_3 k \log k,
\end{aligned}
$$

for some constant C_3.

We now examine the third term in (S.13). From Condition (C2), $\lambda_k - \lambda_\ell \geqslant C^{-1}\ell^{-\alpha-1}(\ell - k) \geqslant C^{-1}(2k)^{-\alpha-1}(\ell-k)$ for $3k/2 \geqslant \ell > k$. Thus, let $C_4 = C^2 2^{\alpha+1}$, incorporating Condition (C2), we have

$$
\begin{aligned}
\sum_{\ell=k+1}^{[3k/2]} (\lambda_k - \lambda_\ell)^{-1} \lambda_\ell^{1/2} \lambda_k^{1/2} &\leqslant C \sum_{\ell=k+1}^{[3k/2]} (2k)^{\alpha+1} \frac{\lambda_\ell^{1/2}\lambda_k^{1/2}}{\ell-k} \\
&\leqslant C_4 \sum_{\ell=k+1}^{[3k/2]} k^{\alpha+1} \frac{k^{-\alpha/2}\ell^{-\alpha/2}}{\ell-k} \\
&\leqslant C_4 k \sum_{\ell=k+1}^{[3k/2]} 1/(\ell-k) \\
&\leqslant C_5 k \log k,
\end{aligned}
$$

for some constant C_5.

Consider the fourth summation in (S.13). By condition (C2), for $\ell > k$,

$$
\begin{aligned}
\lambda_k - \lambda_\ell &= \sum_{t=k}^{\ell-1} (\lambda_t - \lambda_{t+1}) \\
&\geqslant C^{-1} \sum_{t=k}^{\ell-1} t^{-\alpha-1} \\
&\geqslant C^{-1} \int_k^\ell x^{-\alpha-1} dx
\end{aligned}
$$

$$= \frac{1}{C\alpha}(k^{-\alpha} - \ell^{-\alpha}).$$

Hence, by $\alpha > 1$,

$$\begin{aligned}
\sum_{\ell=[3k/2]+1}^{2k} (\lambda_k - \lambda_\ell)^{-1}\lambda_\ell^{1/2}\lambda_k^{1/2} &\leqslant C^2\alpha \sum_{\ell=[3k/2]+1}^{2k} \frac{k^{-\alpha/2}\ell^{-\alpha/2}}{k^{-\alpha} - \ell^{-\alpha}} \\
&= C^2\alpha k \frac{1}{k} \sum_{\ell=[3k/2]+1}^{2k} \frac{(\ell/k)^{\alpha/2}}{(\ell/k)^{\alpha} - 1} \\
&\leqslant C^2\alpha k \int_{3/2}^{2} \frac{x^{\alpha/2}}{x^\alpha - 1} dx \\
&\leqslant C^2\alpha k \int_{3/2}^{2} \frac{x^{\alpha}}{x^\alpha - 1} dx \\
&= C^2\alpha k \int_{3/2}^{2} \frac{x^{\alpha} - 1}{x^\alpha - 1} dx + C^2\alpha k \int_{3/2}^{2} \frac{1}{x^\alpha - 1} dx \\
&\leqslant C^2\alpha k \frac{1}{2} + C^2\alpha k \int_{3/2}^{2} \frac{1}{x - 1} dx \\
&\leqslant C_6 k,
\end{aligned}$$

for some constant C_6.

For the last term in summation in (S.13). Using Condition (C2) and $\ell > 2k + 1$, we have

$$\begin{aligned}
\lambda_k - \lambda_\ell &\geqslant \lambda_k - \lambda_{k+1} + \cdots + \lambda_{\ell-1} - \lambda_\ell \\
&\geqslant C^{-1}k^{-(\alpha+1)} + \cdots + C^{-1}(2k)^{-(\alpha+1)} + \cdots + C^{-1}(\ell-1)^{-(\alpha+1)} \\
&\geqslant C^{-1}k \times (2k)^{-(\alpha+1)} \\
&= C^{-1}2^{-(\alpha+1)}k^{-\alpha}.
\end{aligned}$$

This implies that

$$\lambda_k/\lambda_\ell \geqslant \frac{\lambda_\ell + C^{-1}2^{-(\alpha+1)}k^{-\alpha}}{\lambda_\ell}$$

$$
\begin{aligned}
&\geqslant 1+C^{-1}2^{-(\alpha+1)}k^{-\alpha}C^{-1}\ell^{\alpha} \\
&\geqslant 1+C^{-2}2^{-(\alpha+1)}k^{-\alpha}(2k)^{\alpha} \\
&\geqslant 1+C^{-2}2^{-1} \\
&= 1+(2C^2)^{-1},
\end{aligned}
$$

which leads to $\frac{1}{2C^2+1}\lambda_k \leqslant \lambda_k - \lambda_\ell \leqslant \lambda_k$ when $\ell > 2k+1$. When $\alpha > 2$, we have

$$
\begin{aligned}
\sum_{\ell=2k+1}^{\infty}(\lambda_k-\lambda_\ell)^{-1}\lambda_\ell^{1/2}\lambda_k^{1/2} &\leqslant (2C^2+1)\sum_{\ell=2k+1}^{\infty}\frac{\lambda_\ell^{1/2}\lambda_k^{1/2}}{\lambda_k} \\
&= (2C^2+1)\sum_{\ell=2k+1}^{\infty}\lambda_\ell^{1/2}\lambda_k^{-1/2} \\
&\leqslant (2C^2+1)Ck^{\alpha/2}\sum_{\ell=2k+1}^{\infty}\ell^{-\alpha/2} \\
&\leqslant C_7 k,
\end{aligned}
$$

for some constant C_7. Thus, according to (S.13), we conclude that, if $\alpha > 2$,

$$
\sum_{\ell\neq k}(\lambda_k-\lambda_\ell)^{-1}\lambda_\ell^{1/2}\lambda_k^{1/2} = O(k\log k). \tag{S.14}
$$

Now, it follows that

$$
\begin{aligned}
\Lambda_{nk} &= O_p\left\{n^{-1/2}\sum_{\ell\neq k}(\lambda_k-\lambda_\ell)^{-1}\lambda_\ell^{1/2}\lambda_k^{1/2}\right\} \\
&= O_p\left\{n^{-1/2}O(k\log k)\right\} \\
&= O_p(n^{-1/2}k\log k).
\end{aligned}
$$

Combing (S.12) and $\delta_k'^{-1} \leqslant 2Ck^{\alpha+1}$, we have

$$\begin{aligned}
\widehat{\phi}_k(t) - \phi_k(t) &= \sum_{\ell \neq k} (\lambda_k - \lambda_\ell)^{-1} \phi_\ell \langle (\widehat{G} - G)\phi_\ell, \phi_k \rangle + O_p(\Lambda_{nk} \times n^{-1/2} \delta_k'^{-1}) \\
&= \sum_{\ell \neq k} (\lambda_k - \lambda_\ell)^{-1} \phi_\ell \langle (\widehat{G} - G)\phi_\ell, \phi_k \rangle + O_p(n^{-1} k^{\alpha+2} \log k).
\end{aligned}$$

Noting that

$$\begin{aligned}
& \iint \left\{ \widehat{G}(s,t) - G(s,t) \right\} \phi_k(s)\phi_k(t) ds dt \\
= \; & \iint \left[\frac{1}{n} \sum_{i=1}^{n} Z_i(s) Z_i(t) - G(s,t) \right] \phi_k(s)\phi_k(t) ds dt \\
= \; & \frac{1}{n} \sum_{i=1}^{n} \xi_{ik}^2 - \lambda_k, \qquad \text{(S.15)}
\end{aligned}$$

incorporating (S.11), we get

$$\widehat{\lambda}_k - \lambda_k = \frac{1}{n} \sum_{i=1}^{n} (\xi_{ik}^2 - \lambda_k) + O_p(k/n). \qquad \text{(S.16)}$$

(c) Using (S.15),

$$\begin{aligned}
& E\left[\iint \left\{ \widehat{G}(s,t) - G(s,t) \right\} \phi_k(s)\phi_k(t) ds dt \right] \\
= \; & E\left\{ n^{-1} \sum_{i=1}^{n} (\xi_{ik}^2 - \lambda_k) \right\} \\
= \; & 0
\end{aligned}$$

and

$$E\left[\iint \left\{ \widehat{G}(s,t) - G(s,t) \right\} \phi_k(s)\phi_k(t) ds dt \right]^2$$

$$
\begin{aligned}
&= E\left\{n^{-1}\sum_{i=1}^{n}(\xi_{ik}^2-\lambda_k)\right\}^2 \\
&= n^{-1}E(\xi_{ik}^2-\lambda_k)^2 \\
&\leqslant Cn^{-1}\lambda_k^2,
\end{aligned}
$$

where the last inequality is established by condition $E(\xi_k^2-\lambda_k)^2 < C\lambda_k^2$. Hence, by (S.16), for some constant C_8,

$$
\begin{aligned}
E(\widehat{\lambda}_k-\lambda_k)^2 &\leqslant 2E\left\{n^{-1}\sum_{i=1}^{n}(\xi_{ik}^2-\lambda_k)\right\}^2 + C_8k^2/n^2 \\
&= O(n^{-1}\lambda_k^2)+O(k^2/n^2) \\
&= O(n^{-1}\lambda_k^2),
\end{aligned}
$$

where the last equality is because $k = o(n^{\frac{1}{2\alpha+2}})$. Finally, the last equation

$$
\widehat{\xi}_{ik}-\xi_{ik} = O_p(k^{1-\alpha/2}n^{-1/2})
$$

is a result in the proof of Proposition 1 in Wong et al. (2019). This completes the proof. □

S3.6 Proof of Lemma S2.5

By Theorem 5.6.9. of Horn and Johnson (2012), it is obvious that

$$
\left|\lambda_{\max}\left\{\frac{1}{\sqrt{n}}\left(\widehat{\mathbf{A}}^{\mathrm{T}}-\mathbf{A}^{\mathrm{T}}\right)\left(\widehat{\mathbf{A}}-\mathbf{A}\right)\right\}\right| \leqslant \left\|\frac{1}{\sqrt{n}}\left(\widehat{\mathbf{A}}^{\mathrm{T}}-\mathbf{A}^{\mathrm{T}}\right)\left(\widehat{\mathbf{A}}-\mathbf{A}\right)\right\|_1,
$$

where $\|\cdot\|_1$ is the maximum column sum matrix norm. By Lemma S2.4 (c), we have

$$
\left\|\frac{1}{\sqrt{n}}\left(\widehat{\mathbf{A}}^{\mathrm{T}}-\mathbf{A}^{\mathrm{T}}\right)\left(\widehat{\mathbf{A}}-\mathbf{A}\right)\right\|_1 \leqslant \frac{1}{\sqrt{n}}\max_{1\leqslant k_2\leqslant K}\sum_{k_1=1}^{K}\sum_{i=1}^{n}\left|(\widehat{\xi}_{ik_1}-\xi_{ik_1})(\widehat{\xi}_{ik_2}-\xi_{ik_2})\right|
$$

$$
\begin{aligned}
&\leqslant C \max_{1\leqslant k_2\leqslant K}\sum_{k_1=1}^{K} {k_1'}^{1-\alpha/2}{k_2'}^{1-\alpha/2}n^{-1/2} \\
&= C\left(\sum_{k_1=1}^{K} {k_1'}^{1-\alpha/2}n^{-1/2}\right)\left(\max_{1\leqslant k_2\leqslant K} {k_2'}^{1-\alpha/2}\right),
\end{aligned}
$$

for some constant C. Note that, for any $1 \leqslant k \leqslant K$,

$$
\sum_{k=1}^{K} k^{1-\alpha/2}n^{-1/2} \asymp \gamma(n),
$$

where

$$
\gamma(n) = \begin{cases} K^{2-\alpha/2}n^{-1/2} & 2<\alpha<4, \\ \log K n^{-1/2} & \alpha = 4, \\ n^{-1/2} & \alpha > 4. \end{cases}
$$

Based on Condition (C5), we know $\gamma(n) \to 0$. Furthermore, $\max_{1\leqslant k_2\leqslant K} {k_2'}^{1-\alpha/2} = O(1)$ by $\alpha > 2$. Hence, it follows that $\lambda_{\max}\left\{\frac{1}{\sqrt{n}}\left(\widehat{\mathbf{A}}^{\mathrm{T}} - \mathbf{A}^{\mathrm{T}}\right)\left(\widehat{\mathbf{A}} - \mathbf{A}\right)\right\} = o_p(1)$. $\square$

S3.7 Proof of Lemma S2.6

The detailed arguments of (S.2), (S.3) and (S.4) are shown below, respectively.

Proof of (S.2). Denote $f(\mathbf{A}) = n^{-1/2}\sum_{i=1}^{n} S_i\{\mathbf{A}; \mathbf{b}(\rho), \rho\}$. According to (2.15), taking derivative of $f(\mathbf{A})$ with respect to $\mathbf{A}$, we have

$$
\begin{aligned}
&\frac{\partial f(\mathbf{A})}{\partial \mathbf{A}} \\
= &\frac{1}{\sqrt{n}}\Bigg(-\mathbf{S}(\rho)\mathbf{D}(\rho)\Bigg[\left\{\frac{\partial \mathbf{D}(\rho)}{\partial \rho}\mathbf{M}(\rho) + \mathbf{D}(\rho)\frac{\partial \mathbf{M}(\rho)}{\partial \rho}\right\}\mathbf{Y} \\
&-\left\{\frac{\partial \mathbf{D}(\rho)}{\partial \rho}\mathbf{S}(\rho)^{\mathrm{T}} + \mathbf{D}(\rho)\left(-\mathbf{W}^{\mathrm{T}}\right)\right\}\mathbf{A}\mathbf{b}(\rho) \ - \mathbf{D}(\rho)\mathbf{S}(\rho)^{\mathrm{T}}\mathbf{A}\frac{\partial \mathbf{b}(\rho)}{\partial \rho}\Bigg]\mathbf{b}(\rho)^{\mathrm{T}}
\end{aligned}
$$

$$
\begin{aligned}
& - \left\{ \mathbf{S}(\rho) \frac{\partial \mathbf{D}(\rho)}{\partial \rho} - \mathbf{W}\mathbf{D}(\rho) \right\} \left\{ \mathbf{D}(\rho)\mathbf{M}(\rho)\mathbf{Y} - \mathbf{D}(\rho)\mathbf{S}(\rho)^{\mathrm{T}}\mathbf{A}\mathbf{b}(\rho) \right\} \mathbf{b}(\rho)^{\mathrm{T}} \\
& - \mathbf{S}(\rho)\mathbf{D}(\rho) \left\{ \mathbf{D}(\rho)\mathbf{M}(\rho)\mathbf{Y} - \mathbf{D}(\rho)\mathbf{S}(\rho)^{\mathrm{T}}\mathbf{A}\mathbf{b}(\rho) \right\} \frac{\partial \mathbf{b}(\rho)^{\mathrm{T}}}{\partial \rho} \Bigg) .
\end{aligned}
$$

By Taylor's expansion,

$$
\begin{aligned}
f(\widehat{\mathbf{A}}) &= f(\mathbf{A}) + \mathrm{vec}^{\mathrm{T}} \left\{ \frac{\partial f(\mathbf{A})}{\partial \mathbf{A}} \right\} \mathrm{vec} \left(\widehat{\mathbf{A}} - \mathbf{A} \right) \\
& \quad + 2^{-1} \mathrm{vec}^{\mathrm{T}} \left\{ \frac{\partial^2 f(\mathbf{A})}{\partial \mathbf{A}^2} \right\} \mathrm{vec} \left\{ \left(\widehat{\mathbf{A}} - \mathbf{A} \right) \otimes \left(\widehat{\mathbf{A}} - \mathbf{A} \right) \right\}, \qquad \text{(S.17)}
\end{aligned}
$$

where $\partial^2 f(\mathbf{A})/\partial \mathbf{A}^2$ does not depend on $\mathbf{A}$, and

$$
\mathrm{vec}^{\mathrm{T}} \left\{ \frac{\partial f(\mathbf{A})}{\partial \mathbf{A}} \right\} \mathrm{vec} \left(\widehat{\mathbf{A}} - \mathbf{A} \right) = \mathscr{A}_{n1} + \mathscr{A}_{n2} + \mathscr{A}_{n3} \qquad \text{(S.18)}
$$

where

$$
\begin{aligned}
\mathscr{A}_{n1} &= \mathrm{vec}^{\mathrm{T}} \Bigg(-\frac{1}{\sqrt{n}} \mathbf{S}(\rho)\mathbf{D}(\rho) \Bigg[\left\{ \frac{\partial \mathbf{D}(\rho)}{\partial \rho} \mathbf{M}(\rho) + \mathbf{D}(\rho) \frac{\partial \mathbf{M}(\rho)}{\partial \rho} \right\} \mathbf{Y} \\
& \quad - \left\{ \frac{\partial \mathbf{D}(\rho)}{\partial \rho} \mathbf{S}(\rho)^{\mathrm{T}} + \mathbf{D}(\rho) \left(-\mathbf{W}^{\mathrm{T}} \right) \right\} \mathbf{A}\mathbf{b}(\rho) \ - \mathbf{D}(\rho)\mathbf{S}(\rho)^{\mathrm{T}}\mathbf{A} \frac{\partial \mathbf{b}(\rho)}{\partial \rho} \Bigg] \mathbf{b}(\rho)^{\mathrm{T}} \Bigg) \\
& \quad \times \mathrm{vec} \left(\widehat{\mathbf{A}} - \mathbf{A} \right),
\end{aligned}
$$

$$
\begin{aligned}
\mathscr{A}_{n2} &= \mathrm{vec}^{\mathrm{T}} \Bigg[-\frac{1}{\sqrt{n}} \left\{ \mathbf{S}(\rho) \frac{\partial \mathbf{D}(\rho)}{\partial \rho} - \mathbf{W}\mathbf{D}(\rho) \right\} \\
& \quad \times \left\{ \mathbf{D}(\rho)\mathbf{M}(\rho)\mathbf{Y} - \mathbf{D}(\rho)\mathbf{S}(\rho)^{\mathrm{T}}\mathbf{A}\mathbf{b}(\rho) \right\} \mathbf{b}(\rho)^{\mathrm{T}} \Bigg] \mathrm{vec} \left(\widehat{\mathbf{A}} - \mathbf{A} \right),
\end{aligned}
$$

and

$$
\mathscr{A}_{n3} = \mathrm{vec}^{\mathrm{T}} \left[-\frac{1}{\sqrt{n}} \mathbf{S}(\rho)\mathbf{D}(\rho) \left\{ \mathbf{D}(\rho)\mathbf{M}(\rho)\mathbf{Y} - \mathbf{D}(\rho)\mathbf{S}(\rho)^{\mathrm{T}}\mathbf{A}\mathbf{b}(\rho) \right\} \frac{\partial \mathbf{b}(\rho)^{\mathrm{T}}}{\partial \rho} \right] \mathrm{vec} \left(\widehat{\mathbf{A}} - \mathbf{A} \right).
$$

We now consider the term $\mathscr{A}_{n1}$. Let $\widehat{\boldsymbol{\delta}}_k = \left(\widehat{\xi}_{1k} - \xi_{1k}, \ldots, \widehat{\xi}_{nk} - \xi_{nk}\right)^{\mathrm{T}}$ and $(\mathbf{B})_{\bullet,j}$ denote the jth column of $\mathbf{B}$ for any $n \times m$ matrix $\mathbf{B}$. For succinct presentation, we write

$$
\begin{aligned}
\mathscr{B}_{n1} &= \sum_{k=1}^{K} \frac{-1}{\sqrt{n}} \{Y_1 b_k(\rho), \ldots, Y_n b_k(\rho)\} \left\{\mathbf{M}(\rho)\frac{\partial \mathbf{D}(\rho)}{\partial \rho} + \frac{\partial \mathbf{M}(\rho)}{\partial \rho}\mathbf{D}(\rho)\right\} \mathbf{D}(\rho)\mathbf{S}(\rho)^{\mathrm{T}}, \\
\mathscr{B}_{n2} &= \sum_{k=1}^{K} \frac{1}{\sqrt{n}} \left\{\sum_{k'=1}^{K} \xi_{1k'} b_{k'}(\rho) b_k(\rho), \ldots, \sum_{k'=1}^{K} \xi_{nk'} b_{k'}(\rho) b_k(\rho)\right\} \\
&\quad \times \left\{\mathbf{S}(\rho)\frac{\partial \mathbf{D}(\rho)}{\partial \rho} - \mathbf{W}\mathbf{D}(\rho)\right\} \mathbf{D}(\rho)\mathbf{S}(\rho)^{\mathrm{T}},
\end{aligned}
$$

and

$$
\mathscr{B}_{n3} = \sum_{k=1}^{K} \frac{1}{\sqrt{n}} \left\{\sum_{k'=1}^{K} \xi_{1k'} \frac{\partial b_{k'}(\rho)}{\partial \rho} b_k(\rho), \ldots, \sum_{k'=1}^{K} \xi_{nk'} \frac{\partial b_{k'}(\rho)}{\partial \rho} b_k(\rho)\right\} \mathbf{S}(\rho)\mathbf{D}(\rho)^2\mathbf{S}(\rho)^{\mathrm{T}}.
$$

Then we have

$$
\begin{aligned}
\mathscr{A}_{n1} &= \sum_{k=1}^{K} \left[-\frac{1}{\sqrt{n}}\mathbf{S}(\rho)\mathbf{D}(\rho)\left\{\frac{\partial \mathbf{D}(\rho)}{\partial \rho}\mathbf{M}(\rho) + \mathbf{D}(\rho)\frac{\partial \mathbf{M}(\rho)}{\partial \rho}\right\} \mathbf{Y}\mathbf{b}(\rho)^{\mathrm{T}}\right]_{\bullet,k}^{\mathrm{T}} \widehat{\boldsymbol{\delta}}_k \\
&\quad + \sum_{k=1}^{K} \left[\frac{1}{\sqrt{n}}\mathbf{S}(\rho)\mathbf{D}(\rho)\left\{\frac{\partial \mathbf{D}(\rho)}{\partial \rho}\mathbf{S}(\rho)^{\mathrm{T}} + \mathbf{D}(\rho)\left(-\mathbf{W}^{\mathrm{T}}\right)\right\} \mathbf{A}\mathbf{b}(\rho)\mathbf{b}(\rho)^{\mathrm{T}}\right]_{\bullet,k}^{\mathrm{T}} \widehat{\boldsymbol{\delta}}_k \\
&\quad + \sum_{k=1}^{K} \frac{1}{\sqrt{n}} \left\{\mathbf{S}(\rho)\mathbf{D}(\rho)^2\mathbf{S}(\rho)^{\mathrm{T}}\mathbf{A}\frac{\partial \mathbf{b}(\rho)}{\partial \rho}\mathbf{b}(\rho)^{\mathrm{T}}\right\}_{\bullet,k}^{\mathrm{T}} \widehat{\boldsymbol{\delta}}_k \\
&= \left(\mathscr{B}_{n1} + \mathscr{B}_{n2} + \mathscr{B}_{n3}\right) \widehat{\boldsymbol{\delta}}_k. \qquad \text{(S.19)}
\end{aligned}
$$

Using the conditions $\xi_{ik} = O_p(1)$ for any $k = 1, \ldots, K, i = 1, \ldots, n$, and $\sup_{\rho \in [-1,1]} |b_k(\rho)| \leqslant Ck^{-\tau}, \sup_{\rho \in [-1,1]} |\partial b_k(\rho)/\partial \rho| \leqslant Ck^{-\tau}$ for each $k \geqslant 1$ in Condition (C7), we have $\sum_{k'=1}^{K} \xi_{ik'} b_{k'}(\rho) = O_p(1)$ and $\sum_{k'=1}^{K} \xi_{ik'} \{\partial b_{k'}(\rho)/\partial \rho\} = O_p(1)$, which indicate $\sum_{k'=1}^{K} \xi_{ik'} b_{k'}(\rho) b_k(\rho) = O_p(k^{-\tau})$ and $\sum_{k'=1}^{K} \xi_{ik'} \{\partial b_{k'}(\rho)/\partial \rho\} b_k(\rho) = O_p(k^{-\tau})$, for $i = 1, \ldots, n$. Thus, combing Lemma S2.4 (c), Condition (C3),(C7) and (C8), it follows that

$$
\begin{aligned}
& \mathscr{A}_{n1} \\
\leqslant\ & \sum_{k=1}^{K} O_p\left(k^{-\tau}\right) O_p\left(k^{1-\alpha/2}\right)\left[\lambda_{\max}\left\{\frac{\partial \mathbf{D}(\rho)}{\partial \rho}\right\}\lambda_{\max}\left\{\mathbf{M}(\rho)\right\}+\lambda_{\max}\left\{\mathbf{D}(\rho)\right\}\lambda_{\max}\left\{\frac{\partial \mathbf{M}(\rho)}{\partial \rho}\right\}\right] \\
& \quad \times \lambda_{\max}\left\{\mathbf{D}(\rho)\right\}\lambda_{\max}^{1/2}\left\{\mathbf{S}(\rho)\mathbf{S}(\rho)^{\mathrm{T}}\right\} \\
& + \sum_{k=1}^{K} O_p\left(k^{-\tau}\right) O_p\left(k^{1-\alpha/2}\right)\left[\lambda_{\max}\left\{\frac{\partial \mathbf{D}(\rho)}{\partial \rho}\right\}\lambda_{\max}^{1/2}\left\{\mathbf{S}(\rho)\mathbf{S}(\rho)^{\mathrm{T}}\right\}+\lambda_{\max}\left\{\mathbf{D}(\rho)\right\}\lambda_{\max}^{1/2}\left(\mathbf{W}\mathbf{W}^{\mathrm{T}}\right)\right] \\
& \quad \times \lambda_{\max}\left\{\mathbf{D}(\rho)\right\}\lambda_{\max}^{1/2}\left\{\mathbf{S}(\rho)\mathbf{S}(\rho)^{\mathrm{T}}\right\} \\
& + \sum_{k=1}^{K} O_p\left(k^{-\tau}\right) O_p\left(k^{1-\alpha/2}\right)\lambda_{\max}\left\{\mathbf{D}(\rho)^2\right\}\lambda_{\max}\left\{\mathbf{S}(\rho)\mathbf{S}(\rho)^{\mathrm{T}}\right\} \\
=\ & \sum_{k=1}^{K} O_p(k^{1-\tau-\alpha/2}) \\
=\ & O_p(1). \qquad \text{(S.20)}
\end{aligned}
$$

where the last equality is because condition $\tau > \alpha/2 + 1$ in (C7) and the assumption $\alpha > 1$ in (C2).

On the other hand, by Lemma S2.4 and its proof concerning $\langle(\widehat{G}-G)\phi_\ell, \phi_k\rangle$, we get

$$
\begin{aligned}
\widehat{\xi}_{ik}-\xi_{ik} &= \sum_{\ell\neq k}(\lambda_k-\lambda_\ell)^{-1}\xi_{ik}\left\langle\left(\widehat{G}-G\right)\phi_\ell,\phi_k\right\rangle\times\{1+o_p(1)\} \\
&= n^{-1}\sum_{i_1=1}^{n}\xi_{i_1k}\xi_{i_1\ell}\sum_{\ell\neq k}(\lambda_k-\lambda_\ell)^{-1}\xi_{ik}\times\{1+o_p(1)\} \\
&= n^{-1}\left(\sum_{i_1\neq i}^{n}\xi_{i_1k}\xi_{i_1\ell}\xi_{ik}+\xi_{ik}^2\xi_{i\ell}\right)\sum_{\ell\neq k}(\lambda_k-\lambda_\ell)^{-1}\times\{1+o_p(1)\} \\
&= (\delta_{i1k}+\delta_{i2k})\times\{1+o_p(1)\},
\end{aligned}
$$

where $\delta_{i1k} = n^{-1}\left(\sum_{i_1\neq i}^{n}\xi_{i_1k}\xi_{i_1\ell}\xi_{ik}\right)\sum_{\ell\neq k}(\lambda_k-\lambda_\ell)^{-1}$, $\delta_{i2k} = n^{-1}\xi_{ik}^2\xi_{i\ell}\sum_{\ell\neq k}(\lambda_k-\lambda_\ell)^{-1}$. Note that $E\delta_{i1k}=0$ and by the proof of Proposition 1 in Wong et al. (2019), $E\delta_{i1k}^2 \leqslant Ck^{2-\alpha}/n$, for $i = 1, \ldots, n$. With

$$\begin{aligned} E\left|\xi_{ik}^2 \xi_{i\ell}\right| &\leqslant (E\xi_{ik}^4)^{1/2}(E\xi_{i\ell}^2)^{1/2} \\ &\leqslant C^{1/2}\lambda_k \lambda_\ell^{1/2} \end{aligned}$$

implied by Condition (C1), and

$$\sum_{\ell \neq k}(\lambda_k - \lambda_\ell)^{-1}\lambda_\ell^{1/2}\lambda_k^{1/2} = O(k\log k),$$

implied by (S.14), we can get

$$\begin{aligned} E\left|\delta_{i2k}\right| &\leqslant n^{-1}C^{1/2}\sum_{\ell \neq k}(\lambda_k - \lambda_\ell)^{-1}\lambda_\ell^{1/2}\lambda_k \\ &= O(n^{-1}k^{1-\alpha/2}\log k). \end{aligned}$$

Let $\boldsymbol{\delta}_{1k} = (\delta_{11k}, \ldots, \delta_{n1k})^{\mathrm{T}}$ and $\boldsymbol{\delta}_{2k} = (\delta_{12k}, \ldots, \delta_{n2k})^{\mathrm{T}}$. Based on above results, $E\boldsymbol{\delta}_{1k} = 0$, $\|\boldsymbol{\delta}_{2k}\| = O_p(n^{-1/2}k^{1-\alpha/2}\log k)$, and Condition (C3),(C7) and (C8), (S.19) can be written as

$$\begin{aligned} \mathscr{A}_{n1} &= \left(\mathscr{B}_{n1} + \mathscr{B}_{n2} + \mathscr{B}_{n3}\right)\left(\boldsymbol{\delta}_{1k} + \boldsymbol{\delta}_{2k}\right)\{1 + o_p(1)\} \\ &= \left(\mathscr{B}_{n1} + \mathscr{B}_{n2} + \mathscr{B}_{n3}\right)\boldsymbol{\delta}_{1k} + o_p(1), \end{aligned}$$

where the second equality is established as follows. By Conditions (C3),(C5),(C7),(C8) and similar to the proof of (S.20),

$$\begin{aligned} &\left(\mathscr{B}_{n1} + \mathscr{B}_{n2} + \mathscr{B}_{n3}\right)\boldsymbol{\delta}_{2k} \\ \leqslant\ & \sum_{k=1}^{K} O_p\left(k^{-\tau}\right) O_p\left(n^{-1/2}k^{1-\alpha/2}\log k\right)\left[\lambda_{\max}\left\{\frac{\partial \mathbf{D}(\rho)}{\partial \rho}\right\}\lambda_{\max}\left\{\mathbf{M}(\rho)\right\}\right. \end{aligned}$$

$$
\begin{aligned}
& + \lambda_{\max}\left\{\mathbf{D}(\rho)\right\}\lambda_{\max}\left\{\frac{\partial \mathbf{M}(\rho)}{\partial \rho}\right\}\Bigg]\lambda_{\max}\left\{\mathbf{D}(\rho)\right\}\lambda_{\max}^{1/2}\left\{\mathbf{S}(\rho)\mathbf{S}(\rho)^{\mathrm{T}}\right\} \\
& + \sum_{k=1}^{K} O_p\left(k^{-\tau}\right) O_p\left(n^{-1/2}k^{1-\alpha/2}\log k\right)\left[\lambda_{\max}\left\{\frac{\partial \mathbf{D}(\rho)}{\partial \rho}\right\}\lambda_{\max}^{1/2}\left\{\mathbf{S}(\rho)\mathbf{S}(\rho)^{\mathrm{T}}\right\}\right. \\
& \left. + \lambda_{\max}\left\{\mathbf{D}(\rho)\right\}\lambda_{\max}^{1/2}\left(\mathbf{W}\mathbf{W}^{\mathrm{T}}\right)\right] \times \lambda_{\max}\left\{\mathbf{D}(\rho)\right\}\lambda_{\max}^{1/2}\left\{\mathbf{S}(\rho)\mathbf{S}(\rho)^{\mathrm{T}}\right\} \\
& + \sum_{k=1}^{K} O_p\left(k^{-\tau}\right) O_p\left(n^{-1/2}k^{1-\alpha/2}\log k\right)\lambda_{\max}\left\{\mathbf{D}(\rho)^2\right\}\lambda_{\max}\left\{\mathbf{S}(\rho)\mathbf{S}(\rho)^{\mathrm{T}}\right\} \\
= \;& o_p(1).
\end{aligned}
$$

For $\mathscr{B}_{n2}\boldsymbol{\delta}_{1k}$,

$$
\begin{aligned}
& E\left(\mathscr{B}_{n2}\boldsymbol{\delta}_{1k}\right) \\
= \;& E\left[\sum_{k=1}^{K}\frac{1}{\sqrt{n}}\left\{\sum_{k'=1}^{K}\xi_{1k'}b_{k'}(\rho)b_k(\rho),\ldots,\sum_{k'=1}^{K}\xi_{nk'}b_{k'}(\rho)b_k(\rho)\right\}\right. \\
& \left. \times\left\{\mathbf{S}(\rho)\frac{\partial \mathbf{D}(\rho)}{\partial \rho} - \mathbf{W}\mathbf{D}(\rho)\right\}\mathbf{D}(\rho)\mathbf{S}(\rho)^{\mathrm{T}}\boldsymbol{\delta}_{1k}\right] \\
= \;& \sum_{k=1}^{K}\frac{1}{\sqrt{n}}\sum_{p=1}^{n}\sum_{q=1}^{n}\sum_{k'=1}^{K} E\left\{\xi_{pk'}b_{k'}(\rho)b_k(\rho)Q_{pq}\left(n^{-1}\sum_{i_1\neq q}^{n}\xi_{i_1k}\xi_{i_1\ell}\right)\xi_{qk}\sum_{\ell\neq k}(\lambda_k-\lambda_\ell)^{-1}\right\} \\
= \;& \sum_{k=1}^{K}\frac{1}{\sqrt{n}}\sum_{p=1}^{n}\sum_{q=1}^{n}\sum_{k'=1}^{K} b_{k'}(\rho)b_k(\rho)Q_{pq}\sum_{\ell\neq k}(\lambda_k-\lambda_\ell)^{-1}E\left\{\xi_{pk'}\left(n^{-1}\sum_{i_1\neq q}^{n}\xi_{i_1k}\xi_{i_1\ell}\right)\xi_{qk}\right\} \\
= \;& 0,
\end{aligned}
$$

where Q_{pq} denotes the (p,q)th element of the matrix $[\mathbf{S}(\rho)\{\partial\mathbf{D}(\rho)/\partial\rho\} - \mathbf{W}\mathbf{D}(\rho)]\mathbf{D}(\rho)\mathbf{S}(\rho)^{\mathrm{T}}$ and is non-random. The last equality holds because for $p=q$, by $k\neq\ell$ and the uncorrelation between ξ_{i_1k} and $\xi_{i_1\ell}$,

$$
\begin{aligned}
E\left\{\xi_{pk'}\left(n^{-1}\sum_{i_1\neq q}^{n}\xi_{i_1k}\xi_{i_1\ell}\right)\xi_{qk}\right\} &= E\left\{\xi_{qk'}\xi_{qk}\left(n^{-1}\sum_{i_1\neq q}^{n}\xi_{i_1k}\xi_{i_1\ell}\right)\right\} \\
&= E\left(\xi_{qk'}\xi_{qk}\right)\left\{n^{-1}\sum_{i_1\neq q}^{n}E\left(\xi_{i_1k}\xi_{i_1\ell}\right)\right\} \\
&= E\left(\xi_{qk'}\xi_{qk}\right)0 = 0,
\end{aligned}
$$

and, for $p \neq q$,

$$
\begin{aligned}
& E\left\{\xi_{pk'}\left(n^{-1}\sum_{i_1\neq q}^{n}\xi_{i_1k}\xi_{i_1\ell}\right)\xi_{qk}\right\} \\
= \; & E\left\{\xi_{pk'}\left(n^{-1}\sum_{i_1\neq q}^{n}\xi_{i_1k}\xi_{i_1\ell}\right)\right\}E\xi_{qk} \\
= \; & 0.
\end{aligned}
$$

Similar derivation can be used to show that $E\left(\mathscr{B}_{n3}\boldsymbol{\delta}_{1k}\right) = 0$. For the term $\mathscr{B}_{n1}\boldsymbol{\delta}_{1k}$, by $E(\mathbf{Y}|\mathbf{A}) = \mathbf{S}^{-1}\mathbf{A}\mathbf{b}(\rho_0)$, we get

$$
\begin{aligned}
& E\left(\mathscr{B}_{n1}\boldsymbol{\delta}_{1k}\right) \\
= \; & E\left[\sum_{k=1}^{K}\frac{-1}{\sqrt{n}}\{Y_1b_k(\rho),\ldots,Y_nb_k(\rho)\}\left\{\mathbf{M}(\rho)\frac{\partial\mathbf{D}(\rho)}{\partial\rho}+\frac{\partial\mathbf{M}(\rho)}{\partial\rho}\mathbf{D}(\rho)\right\}\mathbf{D}(\rho)\mathbf{S}(\rho)^{\mathrm{T}}\boldsymbol{\delta}_{1k}\right] \\
= \; & \sum_{k=1}^{K}\frac{-1}{\sqrt{n}}E\left\{\sum_{j=1}^{n}\sum_{i=1}^{n}Y_ib_k(\rho)Q_{ij}\left(n^{-1}\sum_{i_1\neq j}^{n}\xi_{i_1k}\xi_{i_1\ell}\right)\xi_{jk}\sum_{\ell\neq k}(\lambda_k-\lambda_\ell)^{-1}\right\} \\
= \; & \sum_{k=1}^{K}\frac{-1}{\sqrt{n}}\sum_{j=1}^{n}\sum_{i=1}^{n}b_k(\rho)Q_{ij}\sum_{\ell\neq k}(\lambda_k-\lambda_\ell)^{-1}E\left\{Y_i\left(n^{-1}\sum_{i_1\neq j}^{n}\xi_{i_1k}\xi_{i_1\ell}\right)\xi_{jk}\right\} \\
= \; & \sum_{k=1}^{K}\frac{-1}{\sqrt{n}}\sum_{j=1}^{n}\sum_{i=1}^{n}b_k(\rho)Q_{ij}\sum_{\ell\neq k}(\lambda_k-\lambda_\ell)^{-1}E\left[E\left\{Y_in^{-1}\sum_{i_1\neq j}^{n}\xi_{i_1k}\xi_{i_1\ell}\xi_{jk}\middle|\mathbf{A}\right\}\right] \\
= \; & \sum_{k=1}^{K}\frac{-1}{\sqrt{n}}\sum_{j=1}^{n}\sum_{i=1}^{n}\sum_{k'=1}^{K}\sum_{q=1}^{n}b_k(\rho)Q_{ij}s_{iq}^{-1}b_{k'}(\rho_0)\sum_{\ell\neq k}(\lambda_k-\lambda_\ell)^{-1}E\left(n^{-1}\sum_{i_1\neq j}^{n}\xi_{i_1k}\xi_{i_1\ell}\xi_{jk}\xi_{qk'}\right) \\
= \; & 0,
\end{aligned}
$$

where s_{iq}^{-1} is the (i,q)th element of $\mathbf{S}^{-1}$ and Q_{ij} is the (i,j)th element of $[\mathbf{M}(\rho)\{\partial\mathbf{D}(\rho)/\partial\rho\}+\{\partial\mathbf{M}(\rho)/\partial\rho\}\mathbf{D}(\rho)]\mathbf{D}(\rho)\mathbf{S}(\rho)^{\mathrm{T}}$, both are non-random. Combining the above results, we get $E\mathscr{A}_{n1} = o(1)$. Almost identical derivation also leads to $\mathscr{A}_{n2} = O_p(1)$, $E\mathscr{A}_{n2} = o(1)$, and

$\mathscr{A}_{n3} = O_p(1)$, $E\mathscr{A}_{n3} = o(1)$ and we skip the details. Hence,

$$E\left[\mathrm{vec}^{\mathrm{T}}\left\{\frac{\partial f(\mathbf{A})}{\partial \mathbf{A}}\right\}\mathrm{vec}\left(\widehat{\mathbf{A}} - \mathbf{A}\right)\right] = o(1). \tag{S.21}$$

By (S.18),

$$\begin{aligned}
\mathrm{var}\left[\mathrm{vec}^{\mathrm{T}}\left\{\frac{\partial f(\mathbf{A})}{\partial \mathbf{A}}\right\}\mathrm{vec}\left(\widehat{\mathbf{A}} - \mathbf{A}\right)\right] &\leqslant E\left[\mathrm{vec}^{\mathrm{T}}\left\{\frac{\partial f(\mathbf{A})}{\partial \mathbf{A}}\right\}\mathrm{vec}\left(\widehat{\mathbf{A}} - \mathbf{A}\right)\right]^2 \\
&\leqslant 3E\mathscr{A}_{n1}^2 + 3E\mathscr{A}_{n2}^2 + 3E\mathscr{A}_{n3}^2 \\
&= O(1). \qquad\qquad \text{(S.22)}
\end{aligned}$$

We now consider the last term of (S.17). Using the fact, for any matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ of sizes $m \times n$, $n \times p$, and $p \times q$, $\mathrm{vec}(\mathbf{ABC}) = (\mathbf{C}^{\mathrm{T}} \otimes \mathbf{A})\mathrm{vec}(\mathbf{B})$, we have

$$\begin{aligned}
&\frac{\partial^2 f(\mathbf{A})}{\partial \mathbf{A}^2} \\
= \;& \partial \mathrm{vec}\left\{\frac{\partial f(\mathbf{A})}{\partial \mathbf{A}}\right\}\Big/\partial \mathrm{vec}(\mathbf{A})^{\mathrm{T}} \\
= \;& \partial \mathrm{vec}\left(\frac{1}{\sqrt{n}}\left[\mathbf{S}(\rho)\mathbf{D}(\rho)\left\{\frac{\partial \mathbf{D}(\rho)}{\partial \rho}\mathbf{S}(\rho)^{\mathrm{T}} - \mathbf{D}(\rho)\mathbf{W}^{\mathrm{T}}\right\}\mathbf{A}\mathbf{b}(\rho)\mathbf{b}(\rho)^{\mathrm{T}}\right.\right. \\
&+\mathbf{S}(\rho)\mathbf{D}(\rho)^2\mathbf{S}(\rho)^{\mathrm{T}}\mathbf{A}\frac{\partial \mathbf{b}(\rho)}{\partial \rho}\mathbf{b}(\rho)^{\mathrm{T}} + \left\{\mathbf{S}(\rho)\frac{\partial \mathbf{D}(\rho)}{\partial \rho} - \mathbf{W}\mathbf{D}(\rho)\right\}\mathbf{D}(\rho)\mathbf{S}(\rho)^{\mathrm{T}}\mathbf{A}\mathbf{b}(\rho)\mathbf{b}(\rho)^{\mathrm{T}} \\
&\left.\left.+\mathbf{S}(\rho)\mathbf{D}(\rho)^2\mathbf{S}(\rho)^{\mathrm{T}}\mathbf{A}\mathbf{b}(\rho)\frac{\partial \mathbf{b}(\rho)^{\mathrm{T}}}{\partial \rho}\right]\right)\Big/\partial \mathrm{vec}(\mathbf{A})^{\mathrm{T}} \\
= \;& \frac{1}{\sqrt{n}}\partial\left\{\left(\left\{\mathbf{b}(\rho)\mathbf{b}(\rho)^{\mathrm{T}}\right\} \otimes \left[\mathbf{S}(\rho)\mathbf{D}(\rho)\left\{\frac{\partial \mathbf{D}(\rho)}{\partial \rho}\mathbf{S}(\rho)^{\mathrm{T}} - \mathbf{D}(\rho)\mathbf{W}^{\mathrm{T}}\right\}\right]\right.\right. \\
&+\left\{\mathbf{b}(\rho)\frac{\partial \mathbf{b}(\rho)^{\mathrm{T}}}{\partial \rho}\right\} \otimes \left\{\mathbf{S}(\rho)\mathbf{D}(\rho)^2\mathbf{S}(\rho)^{\mathrm{T}}\right\} \\
&+\left\{\mathbf{b}(\rho)\mathbf{b}(\rho)^{\mathrm{T}}\right\} \otimes \left\{\mathbf{S}(\rho)\frac{\partial \mathbf{D}(\rho)}{\partial \rho} - \mathbf{W}\mathbf{D}(\rho)\right\}\mathbf{D}(\rho)\mathbf{S}(\rho)^{\mathrm{T}} \\
&\left.\left.+\left\{\frac{\partial \mathbf{b}(\rho)}{\partial \rho}\mathbf{b}(\rho)^{\mathrm{T}}\right\} \otimes \left\{\mathbf{S}(\rho)\mathbf{D}(\rho)^2\mathbf{S}(\rho)^{\mathrm{T}}\right\}\right)\mathrm{vec}(\mathbf{A})\right\}\Big/\partial \mathrm{vec}(\mathbf{A})^{\mathrm{T}} \\
= \;& \frac{1}{\sqrt{n}}\left(\frac{\partial\left\{\mathbf{b}(\rho)\mathbf{b}(\rho)^{\mathrm{T}}\right\}}{\partial \rho} \otimes \left\{\mathbf{S}(\rho)\mathbf{D}(\rho)^2\mathbf{S}(\rho)^{\mathrm{T}}\right\} + \left\{\mathbf{b}(\rho)\mathbf{b}(\rho)^{\mathrm{T}}\right\}\right.
\end{aligned}$$

$$
\begin{aligned}
&\quad \otimes\left\{2\mathbf{S}(\rho)\mathbf{D}(\rho)\frac{\partial\mathbf{D}(\rho)}{\partial\rho}\mathbf{S}(\rho)^{\mathrm{T}}+2\rho\mathbf{W}\mathbf{D}(\rho)^2\mathbf{W}^{\mathrm{T}}-\mathbf{D}(\rho)^2\mathbf{W}^{\mathrm{T}}-\mathbf{W}\mathbf{D}(\rho)^2\right\}\Bigg)\\
\equiv&\quad \frac{1}{\sqrt{n}}\mathbf{Q}(\rho),
\end{aligned}
$$

which yields that

$$
\begin{aligned}
&2^{-1}\mathrm{vec}^{\mathrm{T}}\left\{\frac{\partial^2 f(\mathbf{A})}{\partial\mathbf{A}^2}\right\}\mathrm{vec}\left\{\left(\widehat{\mathbf{A}}-\mathbf{A}\right)\otimes\left(\widehat{\mathbf{A}}-\mathbf{A}\right)\right\}\\
=&\ 2^{-1}\frac{1}{\sqrt{n}}\left[\mathrm{vec}\left\{\mathbf{Q}(\rho)\right\}\right]^{\mathrm{T}}\mathrm{vec}\left\{\left(\widehat{\mathbf{A}}-\mathbf{A}\right)\otimes\left(\widehat{\mathbf{A}}-\mathbf{A}\right)\right\}. \qquad \text{(S.23)}
\end{aligned}
$$

Recall that $(w_{ij})_{i=1,\dots,n;j=1,\dots,n}$ is row-normalized adjacency matrix with zero diagonal and $\mathbf{D}(\rho)=\{\mathrm{diag}\mathbf{M}(\rho)\}^{-1}$. We use (k_1,k_2,i,j) to denote the index of vec$\{\mathbf{Q}(\rho)\}$, and use (i',k_1',k_2',j') to denote the corresponding index of vec$\{(\widehat{\mathbf{A}}-\mathbf{A})\otimes(\widehat{\mathbf{A}}-\mathbf{A})\}$. Obviously, there is a one-to-one mapping between (k_1,k_2,i,j) and (i',k_1',k_2',j'). Under Condition (C7) and (C8), Lemma S2.4(c), Remark 2 and $\alpha>2$ imply that, for some constant C,

$$
\begin{aligned}
&\left|\frac{1}{\sqrt{n}}\left[\mathrm{vec}\left\{\mathbf{Q}(\rho)\right\}\right]^{\mathrm{T}}\mathrm{vec}\left\{\left(\widehat{\mathbf{A}}-\mathbf{A}\right)\otimes\left(\widehat{\mathbf{A}}-\mathbf{A}\right)\right\}\right|\\
=&\ \Bigg|\frac{1}{\sqrt{n}}\sum_{i\neq j}^{n}\sum_{k_1=1}^{K}\sum_{k_2=1}^{K}\left\{b_{k_1}(\rho)\frac{\partial b_{k_2}(\rho)}{\partial\rho}+\frac{\partial b_{k_1}(\rho)}{\partial\rho}b_{k_2}(\rho)\right\}\\
&\quad\times\left\{-\rho w_{ij}D_{jj}(\rho)^2-\rho w_{ji}D_{ii}(\rho)^2+\rho^2\sum_{\ell\neq i,j}^{n}w_{i\ell}w_{j\ell}D_{\ell\ell}(\rho)^2\right\}\left\{\left(\widehat{\mathbf{A}}-\mathbf{A}\right)\otimes\left(\widehat{\mathbf{A}}-\mathbf{A}\right)\right\}_{i'k_1',k_2'j'}\\
&\quad+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{k_1=1}^{K}\sum_{k_2=1}^{K}\left\{b_{k_1}(\rho)\frac{\partial b_{k_2}(\rho)}{\partial\rho}+\frac{\partial b_{k_1}(\rho)}{\partial\rho}b_{k_2}(\rho)\right\}\\
&\quad\times\left\{D_{ii}(\rho)^2+\rho^2\sum_{\ell\neq i}^{n}w_{i\ell}^2D_{\ell\ell}(\rho)^2\right\}\left\{\left(\widehat{\mathbf{A}}-\mathbf{A}\right)\otimes\left(\widehat{\mathbf{A}}-\mathbf{A}\right)\right\}_{i'k_1',k_2'j'}\\
&\quad+\frac{1}{\sqrt{n}}\sum_{i\neq j}^{n}\sum_{k_1=1}^{K}\sum_{k_2=1}^{K}b_{k_1}(\rho)b_{k_2}(\rho)\Bigg\{-2\rho w_{ij}D_{jj}(\rho)\frac{\partial D_{jj}(\rho)}{\partial\rho}-2\rho w_{ji}D_{ii}(\rho)\frac{\partial D_{ii}(\rho)}{\partial\rho}\\
&\quad\ +2\rho^2\sum_{\ell\neq i,j}^{n}w_{i\ell}w_{j\ell}D_{\ell\ell}(\rho)\frac{\partial D_{\ell\ell}(\rho)}{\partial\rho}\Bigg\}\left\{\left(\widehat{\mathbf{A}}-\mathbf{A}\right)\otimes\left(\widehat{\mathbf{A}}-\mathbf{A}\right)\right\}_{i'k_1',k_2'j'}
\end{aligned}
$$

$$
\begin{aligned}
& +\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{k_1=1}^{K}\sum_{k_2=1}^{K} b_{k_1}(\rho)b_{k_2}(\rho)\left\{2D_{ii}(\rho)\frac{\partial D_{ii}(\rho)}{\partial\rho}+2\rho^2\sum_{\ell\neq i}^{n} w_{i\ell}^2 D_{\ell\ell}(\rho)\frac{\partial D_{\ell\ell}(\rho)}{\partial\rho}\right\} \\
& \times\left\{\left(\widehat{\mathbf{A}}-\mathbf{A}\right)\otimes\left(\widehat{\mathbf{A}}-\mathbf{A}\right)\right\}_{i'k_1',k_2'j'} \\
& +2\rho\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k_1=1}^{K}\sum_{k_2=1}^{K} b_{k_1}(\rho)b_{k_2}(\rho)\sum_{\ell\neq i,j}^{n} w_{i\ell}w_{j\ell}D_{\ell\ell}(\rho)^2\left\{\left(\widehat{\mathbf{A}}-\mathbf{A}\right)\otimes\left(\widehat{\mathbf{A}}-\mathbf{A}\right)\right\}_{i'k_1',k_2'j'} \\
& -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k_1=1}^{K}\sum_{k_2=1}^{K} b_{k_1}(\rho)b_{k_2}(\rho)\left\{w_{ij}D_{jj}(\rho)^2+w_{ji}D_{ii}(\rho)^2\right\} \\
& \times\left\{\left(\widehat{\mathbf{A}}-\mathbf{A}\right)\otimes\left(\widehat{\mathbf{A}}-\mathbf{A}\right)\right\}_{i'k_1',k_2'j'}\Bigg| \\
\leqslant\ & \sum_{i\neq j}^{n}\sum_{k_1=1}^{K}\sum_{k_2=1}^{K}\left\{|b_{k_1}(\rho)|\left|\frac{\partial b_{k_2}(\rho)}{\partial\rho}\right|+\left|\frac{\partial b_{k_1}(\rho)}{\partial\rho}\right||b_{k_2}(\rho)|\right\} \\
& \times\left\{|\rho|w_{ij}D_{jj}(\rho)^2+|\rho|w_{ji}D_{ii}(\rho)^2+\rho^2\sum_{\ell\neq i,j}^{n} w_{i\ell}w_{j\ell}D_{\ell\ell}(\rho)^2\right\}O_p\left({k_1'}^{1-\alpha/2}{k_2'}^{1-\alpha/2}n^{-3/2}\right) \\
& +\sum_{i=1}^{n}\sum_{k_1=1}^{K}\sum_{k_2=1}^{K}\left\{|b_{k_1}(\rho)|\left|\frac{\partial b_{k_2}(\rho)}{\partial\rho}\right|+\left|\frac{\partial b_{k_1}(\rho)}{\partial\rho}\right||b_{k_2}(\rho)|\right\} \\
& \times\left\{D_{ii}(\rho)^2+\rho^2\sum_{\ell\neq i}^{n} w_{i\ell}^2 D_{\ell\ell}(\rho)^2\right\}O_p\left({k_1'}^{1-\alpha/2}{k_2'}^{1-\alpha/2}n^{-3/2}\right) \\
& +\sum_{i\neq j}^{n}\sum_{k_1=1}^{K}\sum_{k_2=1}^{K}|b_{k_1}(\rho)|\,|b_{k_2}(\rho)|\left\{2|\rho|w_{ij}D_{jj}(\rho)\left|\frac{\partial D_{jj}(\rho)}{\partial\rho}\right|+2|\rho|w_{ji}D_{ii}(\rho)\left|\frac{\partial D_{ii}(\rho)}{\partial\rho}\right|\right. \\
& \left.+\,2\rho^2\sum_{\ell\neq i,j}^{n} w_{i\ell}w_{j\ell}D_{\ell\ell}(\rho)\left|\frac{\partial D_{\ell\ell}(\rho)}{\partial\rho}\right|\right\}O_p\left({k_1'}^{1-\alpha/2}{k_2'}^{1-\alpha/2}n^{-3/2}\right) \\
& +\sum_{i=1}^{n}\sum_{k_1=1}^{K}\sum_{k_2=1}^{K}|b_{k_1}(\rho)|\,|b_{k_2}(\rho)|\left\{2D_{ii}(\rho)\left|\frac{\partial D_{ii}(\rho)}{\partial\rho}\right|+2\rho^2\sum_{\ell\neq i}^{n} w_{i\ell}^2 D_{\ell\ell}(\rho)\left|\frac{\partial D_{\ell\ell}(\rho)}{\partial\rho}\right|\right\} \\
& \times O_p\left({k_1'}^{1-\alpha/2}{k_2'}^{1-\alpha/2}n^{-3/2}\right) \\
& +2|\rho|\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k_1=1}^{K}\sum_{k_2=1}^{K}|b_{k_1}(\rho)|\,|b_{k_2}(\rho)|\sum_{\ell\neq i,j}^{n} w_{i\ell}w_{j\ell}D_{\ell\ell}(\rho)^2 O_p\left({k_1'}^{1-\alpha/2}{k_2'}^{1-\alpha/2}n^{-3/2}\right) \\
& +\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k_1=1}^{K}\sum_{k_2=1}^{K}|b_{k_1}(\rho)|\,|b_{k_2}(\rho)|\left\{w_{ij}D_{jj}(\rho)^2+w_{ji}D_{ii}(\rho)^2\right\}O_p\left({k_1'}^{1-\alpha/2}{k_2'}^{1-\alpha/2}n^{-3/2}\right) \\
\leqslant\ & \sum_{i\neq j}^{n}\sum_{k_1=1}^{K}\sum_{k_2=1}^{K}\left\{|b_{k_1}(\rho)|\left|\frac{\partial b_{k_2}(\rho)}{\partial\rho}\right|+\left|\frac{\partial b_{k_1}(\rho)}{\partial\rho}\right||b_{k_2}(\rho)|\right\}
\end{aligned}
$$

$$
\begin{aligned}
&\times \left\{ |\rho| w_{ij} D_{jj}(\rho)^2 + |\rho| w_{ji} D_{ii}(\rho)^2 + \rho^2 \sum_{\ell \neq i,j}^{n} w_{i\ell} w_{j\ell} D_{\ell\ell}(\rho)^2 \right\} O_p\left(n^{-3/2}\right) \\
&+ \sum_{i=1}^{n} \sum_{k_1=1}^{K} \sum_{k_2=1}^{K} \left\{ |b_{k_1}(\rho)| \left| \frac{\partial b_{k_2}(\rho)}{\partial \rho} \right| + \left| \frac{\partial b_{k_1}(\rho)}{\partial \rho} \right| |b_{k_2}(\rho)| \right\} \\
&\times \left\{ D_{ii}(\rho)^2 + \rho^2 \sum_{\ell \neq i}^{n} w_{i\ell}^2 D_{\ell\ell}(\rho)^2 \right\} O_p\left(n^{-3/2}\right) \\
&+ \sum_{i \neq j}^{n} \sum_{k_1=1}^{K} \sum_{k_2=1}^{K} |b_{k_1}(\rho)| \, |b_{k_2}(\rho)| \left\{ 2|\rho| w_{ij} D_{jj}(\rho) \left| \frac{\partial D_{jj}(\rho)}{\partial \rho} \right| + 2|\rho| w_{ji} D_{ii}(\rho) \left| \frac{\partial D_{ii}(\rho)}{\partial \rho} \right| \right. \\
&\quad \left. + 2\rho^2 \sum_{\ell \neq i,j}^{n} w_{i\ell} w_{j\ell} D_{\ell\ell}(\rho) \left| \frac{\partial D_{\ell\ell}(\rho)}{\partial \rho} \right| \right\} O_p\left(n^{-3/2}\right) \\
&+ \sum_{i=1}^{n} \sum_{k_1=1}^{K} \sum_{k_2=1}^{K} |b_{k_1}(\rho)| \, |b_{k_2}(\rho)| \left\{ 2 D_{ii}(\rho) \left| \frac{\partial D_{ii}(\rho)}{\partial \rho} \right| + 2\rho^2 \sum_{\ell \neq i}^{n} w_{i\ell}^2 D_{\ell\ell}(\rho) \left| \frac{\partial D_{\ell\ell}(\rho)}{\partial \rho} \right| \right\} O_p\left(n^{-3/2}\right) \\
&+ 2|\rho| \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k_1=1}^{K} \sum_{k_2=1}^{K} |b_{k_1}(\rho)| \, |b_{k_2}(\rho)| \sum_{\ell \neq i,j}^{n} w_{i\ell} w_{j\ell} D_{\ell\ell}(\rho)^2 O_p\left(n^{-3/2}\right) \\
&+ \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k_1=1}^{K} \sum_{k_2=1}^{K} |b_{k_1}(\rho)| \, |b_{k_2}(\rho)| \left\{ w_{ij} D_{jj}(\rho)^2 + w_{ji} D_{ii}(\rho)^2 \right\} O_p\left(n^{-3/2}\right) \\
\leqslant\ & 2C^2 \left[\sum_{i \neq j}^{n} \left(\sum_{k_1=1}^{K} \sum_{k_2=1}^{K} k_1^{-\tau} k_2^{-\tau} \right) \left(|\rho| w_{ij} + |\rho| w_{ji} + \rho^2 \sum_{\ell \neq i,j}^{n} w_{i\ell} w_{j\ell} \right) O_p\left(n^{-3/2}\right) \right. \\
&\left. + \sum_{i=1}^{n} \left(\sum_{k_1=1}^{K} \sum_{k_2=1}^{K} k_1^{-\tau} k_2^{-\tau} \right) \left(1 + \rho^2 \sum_{\ell \neq i}^{n} w_{i\ell}^2 \right) O_p\left(n^{-3/2}\right) \right] \\
&+ C^2 \left[\sum_{i \neq j}^{n} \left(\sum_{k_1=1}^{K} \sum_{k_2=1}^{K} k_1^{-\tau} k_2^{-\tau} \right) \left(w_{ij} + w_{ji} + |\rho| \sum_{\ell \neq i,j}^{n} w_{i\ell} w_{j\ell} \right) O_p\left(n^{-3/2}\right) \right. \\
&+ \sum_{i=1}^{n} \left(\sum_{k_1=1}^{K} \sum_{k_2=1}^{K} k_1^{-\tau} k_2^{-\tau} \right) \left\{ \mathrm{I}(\rho \neq 0)/|\rho| + |\rho| \sum_{\ell \neq i}^{n} w_{i\ell}^2 \right\} O_p\left(n^{-3/2}\right) \\
&+ 2|\rho| \sum_{i=1}^{n} \sum_{j=1}^{n} \left(\sum_{k_1=1}^{K} \sum_{k_2=1}^{K} k_1^{-\tau} k_2^{-\tau} \right) \sum_{\ell \neq i,j}^{n} w_{i\ell} w_{j\ell} O_p\left(n^{-3/2}\right) \\
&\left. + \sum_{i=1}^{n} \sum_{j=1}^{n} \left(\sum_{k_1=1}^{K} \sum_{k_2=1}^{K} k_1^{-\tau} k_2^{-\tau} \right) (w_{ij} + w_{ji}) \, O_p\left(n^{-3/2}\right) \right] \\
=\ & o_p(1),
\end{aligned}
$$

where the last equality is because

$$\sum_{\ell=1}^{n} w_{i\ell} w_{j\ell} \leqslant (\sum_{\ell=1}^{n} w_{i\ell}) \|\mathbf{w}_j\|_\infty = \|\mathbf{w}_j\|_\infty = o(n^{-1/2}) \tag{S.24}$$

due to Condition (C9), and $\sum_{\ell=1}^{n} w_{i\ell}^2 \leqslant \sum_{\ell=1}^{n} w_{i\ell} = 1$. Thus, by (S.23), we can obtain

$$2^{-1}\mathrm{vec}^{\mathrm{T}}\left\{\frac{\partial^2 f(\mathbf{A})}{\partial \mathbf{A}^2}\right\}\mathrm{vec}\left\{\left(\widehat{\mathbf{A}}-\mathbf{A}\right)\otimes\left(\widehat{\mathbf{A}}-\mathbf{A}\right)\right\} \quad = \quad o_p(1). \tag{S.25}$$

Hence, it follows from combing (S.17), (S.21), (S.22) and (S.25) that,

$$\begin{aligned}
& E\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n} S_i\{\widehat{\mathbf{A}};\mathbf{b}(\rho),\rho\} - \frac{1}{\sqrt{n}}\sum_{i=1}^{n} S_i\{\mathbf{A};\mathbf{b}(\rho),\rho\}\right]^2 \\
= \; & E\left[\mathrm{vec}^{\mathrm{T}}\left\{\frac{\partial f(\mathbf{A})}{\partial \mathbf{A}}\right\}\mathrm{vec}\left(\widehat{\mathbf{A}}-\mathbf{A}\right) + 2^{-1}\mathrm{vec}^{\mathrm{T}}\left\{\frac{\partial^2 f(\mathbf{A})}{\partial \mathbf{A}^2}\right\}\mathrm{vec}\left\{\left(\widehat{\mathbf{A}}-\mathbf{A}\right)\otimes\left(\widehat{\mathbf{A}}-\mathbf{A}\right)\right\}\right]^2 \\
= \; & E\left[\mathrm{vec}^{\mathrm{T}}\left\{\frac{\partial f(\mathbf{A})}{\partial \mathbf{A}}\right\}\mathrm{vec}\left(\widehat{\mathbf{A}}-\mathbf{A}\right) + o_p(1)\right]^2 \\
= \; & O(1).
\end{aligned}$$

□

Proof of (S.3). Note that

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\partial S_i\{\widehat{\mathbf{A}};\mathbf{b}(\rho),\rho\}}{\partial \mathbf{b}(\rho)^{\mathrm{T}}} - \frac{1}{n}\sum_{i=1}^{n}\frac{\partial S_i\{\mathbf{A};\mathbf{b}(\rho),\rho\}}{\partial \mathbf{b}(\rho)^{\mathrm{T}}} \quad = \quad \mathbf{a}_{n1}+\mathbf{a}_{n2}+\mathbf{a}_{n3}, \tag{S.26}$$

where

$$\begin{aligned}
\mathbf{a}_{n1} \quad = \quad & \frac{2}{n}\left\{-\mathbf{Y}^{\mathrm{T}}\mathbf{M}(\rho)\frac{\partial \mathbf{D}(\rho)}{\partial \rho}\mathbf{D}(\rho)\mathbf{S}(\rho)^{\mathrm{T}}\left(\widehat{\mathbf{A}}-\mathbf{A}\right)\right. \\
& +\mathbf{b}(\rho)^{\mathrm{T}}\left(\widehat{\mathbf{A}}^{\mathrm{T}}-\mathbf{A}^{\mathrm{T}}\right)\mathbf{S}(\rho)\frac{\partial \mathbf{D}(\rho)}{\partial \rho}\mathbf{D}(\rho)\mathbf{S}(\rho)^{\mathrm{T}}\mathbf{A} \\
& +\mathbf{b}(\rho)^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{S}(\rho)\frac{\partial \mathbf{D}(\rho)}{\partial \rho}\mathbf{D}(\rho)\mathbf{S}(\rho)^{\mathrm{T}}\left(\widehat{\mathbf{A}}-\mathbf{A}\right)
\end{aligned}$$

$$+\mathbf{b}(\rho)^{\mathrm{T}}\left(\widehat{\mathbf{A}}^{\mathrm{T}}-\mathbf{A}^{\mathrm{T}}\right)\mathbf{S}(\rho)\frac{\partial\mathbf{D}(\rho)}{\partial\rho}\mathbf{D}(\rho)\mathbf{S}(\rho)^{\mathrm{T}}\left(\widehat{\mathbf{A}}-\mathbf{A}\right)\Big\},$$

$$\begin{aligned}
\mathbf{a}_{n2} \;=\;& \frac{1}{n}\Big\{-\mathbf{Y}^{\mathrm{T}}\frac{\partial\mathbf{M}(\rho)}{\partial\rho}\mathbf{D}(\rho)^2\mathbf{S}(\rho)^{\mathrm{T}}\left(\widehat{\mathbf{A}}-\mathbf{A}\right)\\
&-\mathbf{b}(\rho)^{\mathrm{T}}\left(\widehat{\mathbf{A}}^{\mathrm{T}}-\mathbf{A}^{\mathrm{T}}\right)\mathbf{W}\mathbf{D}(\rho)^2\mathbf{S}(\rho)^{\mathrm{T}}\mathbf{A}\\
&-\mathbf{b}(\rho)^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{D}(\rho)^2\mathbf{S}(\rho)^{\mathrm{T}}\left(\widehat{\mathbf{A}}-\mathbf{A}\right)\\
&-\mathbf{b}(\rho)^{\mathrm{T}}\left(\widehat{\mathbf{A}}^{\mathrm{T}}-\mathbf{A}^{\mathrm{T}}\right)\mathbf{W}\mathbf{D}(\rho)^2\mathbf{S}(\rho)^{\mathrm{T}}\left(\widehat{\mathbf{A}}-\mathbf{A}\right)\\
&+\frac{\partial\mathbf{b}(\rho)^{\mathrm{T}}}{\partial\rho}\left(\widehat{\mathbf{A}}^{\mathrm{T}}-\mathbf{A}^{\mathrm{T}}\right)\mathbf{S}(\rho)\mathbf{D}(\rho)^2\mathbf{S}(\rho)^{\mathrm{T}}\mathbf{A}\\
&+\frac{\partial\mathbf{b}(\rho)^{\mathrm{T}}}{\partial\rho}\mathbf{A}^{\mathrm{T}}\mathbf{S}(\rho)\mathbf{D}(\rho)^2\mathbf{S}(\rho)^{\mathrm{T}}\left(\widehat{\mathbf{A}}-\mathbf{A}\right)\\
&+\frac{\partial\mathbf{b}(\rho)^{\mathrm{T}}}{\partial\rho}\left(\widehat{\mathbf{A}}^{\mathrm{T}}-\mathbf{A}^{\mathrm{T}}\right)\mathbf{S}(\rho)\mathbf{D}(\rho)^2\mathbf{S}(\rho)^{\mathrm{T}}\left(\widehat{\mathbf{A}}-\mathbf{A}\right)\Big\},
\end{aligned}$$

and

$$\begin{aligned}
\mathbf{a}_{n3} \;=\;& \frac{1}{n}\Big\{\mathbf{Y}^{\mathrm{T}}\mathbf{M}(\rho)\mathbf{D}(\rho)^2\mathbf{W}^{\mathrm{T}}\left(\widehat{\mathbf{A}}-\mathbf{A}\right)\\
&-\mathbf{b}(\rho)^{\mathrm{T}}\left(\widehat{\mathbf{A}}^{\mathrm{T}}-\mathbf{A}^{\mathrm{T}}\right)\mathbf{S}(\rho)\mathbf{D}(\rho)^2\mathbf{W}^{\mathrm{T}}\mathbf{A}\\
&-\mathbf{b}(\rho)^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{S}(\rho)\mathbf{D}(\rho)^2\mathbf{W}^{\mathrm{T}}\left(\widehat{\mathbf{A}}-\mathbf{A}\right)\\
&-\mathbf{b}(\rho)^{\mathrm{T}}\left(\widehat{\mathbf{A}}^{\mathrm{T}}-\mathbf{A}^{\mathrm{T}}\right)\mathbf{S}(\rho)\mathbf{D}(\rho)^2\mathbf{W}^{\mathrm{T}}\left(\widehat{\mathbf{A}}-\mathbf{A}\right)\Big\}.
\end{aligned}$$

We now treat the first term on the right-hand side of (S.26). We have

$$\begin{aligned}
&\|\mathbf{a}_{n1}\|^2\\
\leqslant\;& \frac{16}{n^2}\mathbf{Y}^{\mathrm{T}}\mathbf{M}(\rho)\frac{\partial\mathbf{D}(\rho)}{\partial\rho}\mathbf{D}(\rho)\mathbf{S}(\rho)^{\mathrm{T}}\left(\widehat{\mathbf{A}}-\mathbf{A}\right)\left(\widehat{\mathbf{A}}^{\mathrm{T}}-\mathbf{A}^{\mathrm{T}}\right)\mathbf{S}(\rho)\mathbf{D}(\rho)\frac{\partial\mathbf{D}(\rho)}{\partial\rho}\mathbf{M}(\rho)\mathbf{Y}\\
&+\frac{16}{n^2}\mathbf{b}(\rho)^{\mathrm{T}}\left(\widehat{\mathbf{A}}^{\mathrm{T}}-\mathbf{A}^{\mathrm{T}}\right)\mathbf{S}(\rho)\frac{\partial\mathbf{D}(\rho)}{\partial\rho}\mathbf{D}(\rho)\mathbf{S}(\rho)^{\mathrm{T}}\mathbf{A}\mathbf{A}^{\mathrm{T}}\mathbf{S}(\rho)\mathbf{D}(\rho)\frac{\partial\mathbf{D}(\rho)}{\partial\rho}\mathbf{S}(\rho)^{\mathrm{T}}\left(\widehat{\mathbf{A}}-\mathbf{A}\right)\mathbf{b}(\rho)
\end{aligned}$$

$$
\begin{aligned}
& +\frac{16}{n^2}\mathbf{b}(\rho)^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{S}(\rho)\frac{\partial\mathbf{D}(\rho)}{\partial\rho}\mathbf{D}(\rho)\mathbf{S}(\rho)^{\mathrm{T}}\left(\widehat{\mathbf{A}}-\mathbf{A}\right)\left(\widehat{\mathbf{A}}^{\mathrm{T}}-\mathbf{A}^{\mathrm{T}}\right)\mathbf{S}(\rho)\mathbf{D}(\rho)\frac{\partial\mathbf{D}(\rho)}{\partial\rho}\mathbf{S}(\rho)^{\mathrm{T}}\mathbf{A}\mathbf{b}(\rho)\\
& +\frac{16}{n^2}\mathbf{b}(\rho)^{\mathrm{T}}\left(\widehat{\mathbf{A}}^{\mathrm{T}}-\mathbf{A}^{\mathrm{T}}\right)\mathbf{S}(\rho)\frac{\partial\mathbf{D}(\rho)}{\partial\rho}\mathbf{D}(\rho)\mathbf{S}(\rho)^{\mathrm{T}}\left(\widehat{\mathbf{A}}-\mathbf{A}\right)\\
& \quad\times\left(\widehat{\mathbf{A}}^{\mathrm{T}}-\mathbf{A}^{\mathrm{T}}\right)\mathbf{S}(\rho)\mathbf{D}(\rho)\frac{\partial\mathbf{D}(\rho)}{\partial\rho}\mathbf{S}(\rho)^{\mathrm{T}}\left(\widehat{\mathbf{A}}-\mathbf{A}\right)\mathbf{b}(\rho)\\
\leqslant\ & 16\lambda_{\max}\left\{\frac{1}{n}\left(\widehat{\mathbf{A}}-\mathbf{A}\right)\left(\widehat{\mathbf{A}}^{\mathrm{T}}-\mathbf{A}^{\mathrm{T}}\right)\right\}\\
& \quad\times\lambda_{\max}\left\{\mathbf{M}(\rho)\right\}\lambda_{\max}\left\{\mathbf{D}(\rho)^2\right\}\lambda_{\max}\left[\left\{\frac{\partial\mathbf{D}(\rho)}{\partial\rho}\right\}^2\right]\lambda_{\max}\left\{\mathbf{M}(\rho)^2\right\}\frac{1}{n}\|\mathbf{Y}\|^2\\
& +16\lambda_{\max}\left(\frac{1}{n}\mathbf{A}\mathbf{A}^{\mathrm{T}}\right)\lambda_{\max}\left\{\mathbf{M}(\rho)\right\}\lambda_{\max}\left\{\mathbf{D}(\rho)^2\right\}\lambda_{\max}\left[\left\{\frac{\partial\mathbf{D}(\rho)}{\partial\rho}\right\}^2\right]\\
& \quad\times\lambda_{\max}\left\{\mathbf{S}(\rho)\mathbf{S}(\rho)^{\mathrm{T}}\right\}\lambda_{\max}\left\{\frac{1}{n}\left(\widehat{\mathbf{A}}^{\mathrm{T}}-\mathbf{A}^{\mathrm{T}}\right)\left(\widehat{\mathbf{A}}-\mathbf{A}\right)\right\}\|\mathbf{b}(\rho)\|^2\\
& +16\lambda_{\max}\left\{\frac{1}{n}\left(\widehat{\mathbf{A}}-\mathbf{A}\right)\left(\widehat{\mathbf{A}}^{\mathrm{T}}-\mathbf{A}^{\mathrm{T}}\right)\right\}\lambda_{\max}\left\{\mathbf{M}(\rho)\right\}\lambda_{\max}\left\{\mathbf{D}(\rho)^2\right\}\lambda_{\max}\left[\left\{\frac{\partial\mathbf{D}(\rho)}{\partial\rho}\right\}^2\right]\\
& \quad\times\lambda_{\max}\left\{\mathbf{S}(\rho)\mathbf{S}(\rho)^{\mathrm{T}}\right\}\lambda_{\max}\left(\frac{1}{n}\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)\|\mathbf{b}(\rho)\|^2\\
& +16\lambda_{\max}\left\{\frac{1}{n}\left(\widehat{\mathbf{A}}-\mathbf{A}\right)\left(\widehat{\mathbf{A}}^{\mathrm{T}}-\mathbf{A}^{\mathrm{T}}\right)\right\}\lambda_{\max}\left\{\mathbf{M}(\rho)\right\}\lambda_{\max}\left\{\mathbf{D}(\rho)^2\right\}\lambda_{\max}\left[\left\{\frac{\partial\mathbf{D}(\rho)}{\partial\rho}\right\}^2\right]\\
& \quad\times\lambda_{\max}\left\{\mathbf{S}(\rho)\mathbf{S}(\rho)^{\mathrm{T}}\right\}\lambda_{\max}\left\{\frac{1}{n}\left(\widehat{\mathbf{A}}^{\mathrm{T}}-\mathbf{A}^{\mathrm{T}}\right)\left(\widehat{\mathbf{A}}-\mathbf{A}\right)\right\}\|\mathbf{b}(\rho)\|^2\\
=\ & o_p(n^{-1/2}),
\end{aligned}
$$

where the last equality is due to Conditions (C7)-(C8), Remark 2, and Lemma S2.5. Similar arguments can be used to show that $\|\mathbf{a}_{n2}\|^2 = o_p(n^{-1/2})$ and $\|\mathbf{a}_{n3}\|^2 = o_p(n^{-1/2})$. Thus,

$$
\begin{aligned}
\left\|\frac{1}{n}\sum_{i=1}^{n}\frac{\partial S_i\{\widehat{\mathbf{A}};\mathbf{b}(\rho),\rho\}}{\partial\mathbf{b}(\rho)^{\mathrm{T}}}-\frac{1}{n}\sum_{i=1}^{n}\frac{\partial S_i\{\mathbf{A};\mathbf{b}(\rho),\rho\}}{\partial\mathbf{b}(\rho)^{\mathrm{T}}}\right\|^2 &\leqslant 3\sum_{j=1}^{3}\|\mathbf{a}_{nj}\|^2\\
&= o_p(n^{-1/2}).
\end{aligned}
$$

$\square$

Proof of (S.4). Observe that

$$\frac{1}{n}\sum_{i=1}^{n}\frac{dS_i\{\widehat{\mathbf{A}};\mathbf{b}(\rho),\rho\}}{d\rho}-\frac{1}{n}\sum_{i=1}^{n}\frac{dS_i\{\mathbf{A};\mathbf{b}(\rho),\rho\}}{d\rho} \quad = \quad \delta_{n1}+\delta_{n2}+\delta_{n3}+\delta_{n4}+\delta_{n5}+\delta_{n6},$$

where

$$\begin{aligned}\delta_{n1} \quad = \quad & \frac{1}{n}\left\{\mathbf{b}(\rho)^{\mathrm{T}}(\mathbf{A}^{\mathrm{T}}-\widehat{\mathbf{A}}^{\mathrm{T}})\mathbf{S}(\rho)\frac{\partial\mathbf{D}(\rho)}{\partial\rho}+\mathbf{b}(\rho)^{\mathrm{T}}(\widehat{\mathbf{A}}^{\mathrm{T}}-\mathbf{A}^{\mathrm{T}})\mathbf{W}\mathbf{D}(\rho)\right.\\ & \left.+\frac{\partial\mathbf{b}(\rho)^{\mathrm{T}}}{\partial\rho}(\mathbf{A}^{\mathrm{T}}-\widehat{\mathbf{A}}^{\mathrm{T}})\mathbf{S}(\rho)\mathbf{D}(\rho)\right\}\left\{\frac{\partial\mathbf{D}(\rho)}{\partial\rho}\mathbf{M}(\rho)+\mathbf{D}(\rho)\frac{\partial\mathbf{M}(\rho)}{\partial\rho}\right\}\mathbf{Y},\end{aligned}$$

$$\begin{aligned}&\delta_{n2}\\ = \quad & \frac{1}{n}\left\{\mathbf{Y}^{\mathrm{T}}\mathbf{M}(\rho)\frac{\partial\mathbf{D}(\rho)}{\partial\rho}+\mathbf{Y}^{\mathrm{T}}\frac{\partial\mathbf{M}(\rho)}{\partial\rho}\mathbf{D}(\rho)\right\}\left\{\frac{\partial\mathbf{D}(\rho)}{\partial\rho}\mathbf{S}(\rho)^{\mathrm{T}}+\mathbf{D}(\rho)\left(-\mathbf{W}^{\mathrm{T}}\right)\right\}(\mathbf{A}-\widehat{\mathbf{A}})\mathbf{b}(\rho)\\ & +\frac{1}{n}\mathbf{b}(\rho)^{\mathrm{T}}\left(\widehat{\mathbf{A}}^{\mathrm{T}}-\mathbf{A}^{\mathrm{T}}\right)\mathbf{S}(\rho)\frac{\partial\mathbf{D}(\rho)}{\partial\rho}\left\{\frac{\partial\mathbf{D}(\rho)}{\partial\rho}\mathbf{S}(\rho)^{\mathrm{T}}+\mathbf{D}(\rho)\left(-\mathbf{W}^{\mathrm{T}}\right)\right\}\left(\widehat{\mathbf{A}}-\mathbf{A}\right)\mathbf{b}(\rho)\\ & +\frac{1}{n}\mathbf{b}(\rho)^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{S}(\rho)\frac{\partial\mathbf{D}(\rho)}{\partial\rho}\left\{\frac{\partial\mathbf{D}(\rho)}{\partial\rho}\mathbf{S}(\rho)^{\mathrm{T}}+\mathbf{D}(\rho)\left(-\mathbf{W}^{\mathrm{T}}\right)\right\}\left(\widehat{\mathbf{A}}-\mathbf{A}\right)\mathbf{b}(\rho)\\ & +\frac{1}{n}\mathbf{b}(\rho)^{\mathrm{T}}\left(\widehat{\mathbf{A}}^{\mathrm{T}}-\mathbf{A}^{\mathrm{T}}\right)\mathbf{S}(\rho)\frac{\partial\mathbf{D}(\rho)}{\partial\rho}\left\{\frac{\partial\mathbf{D}(\rho)}{\partial\rho}\mathbf{S}(\rho)^{\mathrm{T}}+\mathbf{D}(\rho)\left(-\mathbf{W}^{\mathrm{T}}\right)\right\}\mathbf{A}\mathbf{b}(\rho)\\ & +\frac{1}{n}\mathbf{b}(\rho)^{\mathrm{T}}\left(\widehat{\mathbf{A}}^{\mathrm{T}}-\mathbf{A}^{\mathrm{T}}\right)(-\mathbf{W})\,\mathbf{D}(\rho)\left\{\frac{\partial\mathbf{D}(\rho)}{\partial\rho}\mathbf{S}(\rho)^{\mathrm{T}}+\mathbf{D}(\rho)\left(-\mathbf{W}^{\mathrm{T}}\right)\right\}\left(\widehat{\mathbf{A}}-\mathbf{A}\right)\mathbf{b}(\rho)\\ & +\frac{1}{n}\mathbf{b}(\rho)^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}(-\mathbf{W})\,\mathbf{D}(\rho)\left\{\frac{\partial\mathbf{D}(\rho)}{\partial\rho}\mathbf{S}(\rho)^{\mathrm{T}}+\mathbf{D}(\rho)\left(-\mathbf{W}^{\mathrm{T}}\right)\right\}\left(\widehat{\mathbf{A}}-\mathbf{A}\right)\mathbf{b}(\rho)\\ & +\frac{1}{n}\mathbf{b}(\rho)^{\mathrm{T}}\left(\widehat{\mathbf{A}}^{\mathrm{T}}-\mathbf{A}^{\mathrm{T}}\right)(-\mathbf{W})\,\mathbf{D}(\rho)\left\{\frac{\partial\mathbf{D}(\rho)}{\partial\rho}\mathbf{S}(\rho)^{\mathrm{T}}+\mathbf{D}(\rho)\left(-\mathbf{W}^{\mathrm{T}}\right)\right\}\mathbf{A}\mathbf{b}(\rho)\\ & +\frac{1}{n}\frac{\partial\mathbf{b}(\rho)^{\mathrm{T}}}{\partial\rho}\left(\widehat{\mathbf{A}}^{\mathrm{T}}-\mathbf{A}^{\mathrm{T}}\right)\mathbf{S}(\rho)\mathbf{D}(\rho)\left\{\frac{\partial\mathbf{D}(\rho)}{\partial\rho}\mathbf{S}(\rho)^{\mathrm{T}}+\mathbf{D}(\rho)\left(-\mathbf{W}^{\mathrm{T}}\right)\right\}\left(\widehat{\mathbf{A}}-\mathbf{A}\right)\mathbf{b}(\rho)\\ & +\frac{1}{n}\frac{\partial\mathbf{b}(\rho)^{\mathrm{T}}}{\partial\rho}\mathbf{A}^{\mathrm{T}}\mathbf{S}(\rho)\mathbf{D}(\rho)\left\{\frac{\partial\mathbf{D}(\rho)}{\partial\rho}\mathbf{S}(\rho)^{\mathrm{T}}+\mathbf{D}(\rho)\left(-\mathbf{W}^{\mathrm{T}}\right)\right\}\left(\widehat{\mathbf{A}}-\mathbf{A}\right)\mathbf{b}(\rho)\\ & +\frac{1}{n}\frac{\partial\mathbf{b}(\rho)^{\mathrm{T}}}{\partial\rho}\left(\widehat{\mathbf{A}}^{\mathrm{T}}-\mathbf{A}^{\mathrm{T}}\right)\mathbf{S}(\rho)\mathbf{D}(\rho)\left\{\frac{\partial\mathbf{D}(\rho)}{\partial\rho}\mathbf{S}(\rho)^{\mathrm{T}}+\mathbf{D}(\rho)\left(-\mathbf{W}^{\mathrm{T}}\right)\right\}\mathbf{A}\mathbf{b}(\rho),\end{aligned}$$

$$
\begin{aligned}
\delta_{n3} &= \frac{1}{n}\left\{\mathbf{Y}^{\mathrm{T}}\mathbf{M}(\rho)\frac{\partial\mathbf{D}(\rho)}{\partial\rho}+\mathbf{Y}^{\mathrm{T}}\frac{\partial\mathbf{M}(\rho)}{\partial\rho}\mathbf{D}(\rho)\right\}\mathbf{D}(\rho)\mathbf{S}(\rho)^{\mathrm{T}}(\mathbf{A}-\widehat{\mathbf{A}})\frac{\partial\mathbf{b}(\rho)}{\partial\rho}\\
&+\frac{1}{n}\left\{\mathbf{b}(\rho)^{\mathrm{T}}\left(\widehat{\mathbf{A}}^{\mathrm{T}}-\mathbf{A}^{\mathrm{T}}\right)\mathbf{S}(\rho)\frac{\partial\mathbf{D}(\rho)}{\partial\rho}\mathbf{D}(\rho)\mathbf{S}(\rho)^{\mathrm{T}}\left(\widehat{\mathbf{A}}-\mathbf{A}\right)\frac{\partial\mathbf{b}(\rho)}{\partial\rho}\right\}\\
&+\frac{1}{n}\left\{\mathbf{b}(\rho)^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{S}(\rho)\frac{\partial\mathbf{D}(\rho)}{\partial\rho}\mathbf{D}(\rho)\mathbf{S}(\rho)^{\mathrm{T}}\left(\widehat{\mathbf{A}}-\mathbf{A}\right)\frac{\partial\mathbf{b}(\rho)}{\partial\rho}\right\}\\
&+\frac{1}{n}\left\{\mathbf{b}(\rho)^{\mathrm{T}}\left(\widehat{\mathbf{A}}^{\mathrm{T}}-\mathbf{A}^{\mathrm{T}}\right)\mathbf{S}(\rho)\frac{\partial\mathbf{D}(\rho)}{\partial\rho}\mathbf{D}(\rho)\mathbf{S}(\rho)^{\mathrm{T}}\mathbf{A}\frac{\partial\mathbf{b}(\rho)}{\partial\rho}\right\}\\
&+\frac{1}{n}\left\{\mathbf{b}(\rho)^{\mathrm{T}}\left(\widehat{\mathbf{A}}^{\mathrm{T}}-\mathbf{A}^{\mathrm{T}}\right)(-\mathbf{W})\,\mathbf{D}(\rho)^2\mathbf{S}(\rho)^{\mathrm{T}}\left(\widehat{\mathbf{A}}-\mathbf{A}\right)\frac{\partial\mathbf{b}(\rho)}{\partial\rho}\right\}\\
&+\frac{1}{n}\left\{\mathbf{b}(\rho)^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}(-\mathbf{W})\,\mathbf{D}(\rho)^2\mathbf{S}(\rho)^{\mathrm{T}}\left(\widehat{\mathbf{A}}-\mathbf{A}\right)\frac{\partial\mathbf{b}(\rho)}{\partial\rho}\right\}\\
&+\frac{1}{n}\left\{\mathbf{b}(\rho)^{\mathrm{T}}\left(\widehat{\mathbf{A}}^{\mathrm{T}}-\mathbf{A}^{\mathrm{T}}\right)(-\mathbf{W})\,\mathbf{D}(\rho)^2\mathbf{S}(\rho)^{\mathrm{T}}\mathbf{A}\frac{\partial\mathbf{b}(\rho)}{\partial\rho}\right\}\\
&+\frac{1}{n}\left\{\frac{\partial\mathbf{b}(\rho)^{\mathrm{T}}}{\partial\rho}\left(\widehat{\mathbf{A}}^{\mathrm{T}}-\mathbf{A}^{\mathrm{T}}\right)\mathbf{S}(\rho)\mathbf{D}(\rho)^2\mathbf{S}(\rho)^{\mathrm{T}}\left(\widehat{\mathbf{A}}-\mathbf{A}\right)\frac{\partial\mathbf{b}(\rho)}{\partial\rho}\right\}\\
&+\frac{1}{n}\left\{\frac{\partial\mathbf{b}(\rho)^{\mathrm{T}}}{\partial\rho}\mathbf{A}^{\mathrm{T}}\mathbf{S}(\rho)\mathbf{D}(\rho)^2\mathbf{S}(\rho)^{\mathrm{T}}\left(\widehat{\mathbf{A}}-\mathbf{A}\right)\frac{\partial\mathbf{b}(\rho)}{\partial\rho}\right\}\\
&+\frac{1}{n}\left\{\frac{\partial\mathbf{b}(\rho)^{\mathrm{T}}}{\partial\rho}\left(\widehat{\mathbf{A}}^{\mathrm{T}}-\mathbf{A}^{\mathrm{T}}\right)\mathbf{S}(\rho)\mathbf{D}(\rho)^2\mathbf{S}(\rho)^{\mathrm{T}}\mathbf{A}\frac{\partial\mathbf{b}(\rho)}{\partial\rho}\right\},
\end{aligned}
$$

$$
\begin{aligned}
\delta_{n4} &= \frac{1}{n}\mathbf{b}(\rho)^{\mathrm{T}}(\mathbf{A}^{\mathrm{T}}-\widehat{\mathbf{A}}^{\mathrm{T}})\mathbf{S}(\rho)\mathbf{D}(\rho)\\
&\quad\times\left\{\frac{\partial^2\mathbf{D}(\rho)}{\partial\rho^2}\mathbf{M}(\rho)+2\frac{\partial\mathbf{D}(\rho)}{\partial\rho}\frac{\partial\mathbf{M}(\rho)}{\partial\rho}+2\mathbf{D}(\rho)\left(\mathbf{W}^{\mathrm{T}}\mathbf{W}\right)\right\}\mathbf{Y},
\end{aligned}
$$

$$
\begin{aligned}
\delta_{n5} &= \frac{1}{n}\mathbf{Y}^{\mathrm{T}}\mathbf{M}(\rho)\mathbf{D}(\rho)\left\{\frac{\partial^2\mathbf{D}(\rho)}{\partial\rho^2}\mathbf{S}(\rho)^{\mathrm{T}}+2\frac{\partial\mathbf{D}(\rho)}{\partial\rho}\left(-\mathbf{W}^{\mathrm{T}}\right)\right\}(\mathbf{A}-\widehat{\mathbf{A}})\mathbf{b}(\rho)\\
&+\frac{1}{n}\mathbf{b}(\rho)^{\mathrm{T}}\left(\widehat{\mathbf{A}}^{\mathrm{T}}-\mathbf{A}^{\mathrm{T}}\right)\mathbf{S}(\rho)\mathbf{D}(\rho)\left\{\frac{\partial^2\mathbf{D}(\rho)}{\partial\rho^2}\mathbf{S}(\rho)^{\mathrm{T}}+2\frac{\partial\mathbf{D}(\rho)}{\partial\rho}\left(-\mathbf{W}^{\mathrm{T}}\right)\right\}\left(\widehat{\mathbf{A}}-\mathbf{A}\right)\mathbf{b}(\rho)\\
&+\frac{1}{n}\mathbf{b}(\rho)^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{S}(\rho)\mathbf{D}(\rho)\left\{\frac{\partial^2\mathbf{D}(\rho)}{\partial\rho^2}\mathbf{S}(\rho)^{\mathrm{T}}+2\frac{\partial\mathbf{D}(\rho)}{\partial\rho}\left(-\mathbf{W}^{\mathrm{T}}\right)\right\}\left(\widehat{\mathbf{A}}-\mathbf{A}\right)\mathbf{b}(\rho)\\
&+\frac{1}{n}\mathbf{b}(\rho)^{\mathrm{T}}\left(\widehat{\mathbf{A}}^{\mathrm{T}}-\mathbf{A}^{\mathrm{T}}\right)\mathbf{S}(\rho)\mathbf{D}(\rho)\left\{\frac{\partial^2\mathbf{D}(\rho)}{\partial\rho^2}\mathbf{S}(\rho)^{\mathrm{T}}+2\frac{\partial\mathbf{D}(\rho)}{\partial\rho}\left(-\mathbf{W}^{\mathrm{T}}\right)\right\}\mathbf{A}\mathbf{b}(\rho),
\end{aligned}
$$

and

$$
\begin{aligned}
\delta_{n6} \quad = \quad & \frac{2}{n}\mathbf{Y}^{\mathrm{T}}\mathbf{M}(\rho)\mathbf{D}(\rho)\left\{\frac{\partial\mathbf{D}(\rho)}{\partial\rho}\mathbf{S}(\rho)^{\mathrm{T}}+\mathbf{D}(\rho)\left(-\mathbf{W}^{\mathrm{T}}\right)\right\}(\mathbf{A}-\widehat{\mathbf{A}})\frac{\partial\mathbf{b}(\rho)}{\partial\rho} \\
& +\frac{2}{n}\mathbf{b}(\rho)^{\mathrm{T}}\left(\widehat{\mathbf{A}}^{\mathrm{T}}-\mathbf{A}^{\mathrm{T}}\right)\mathbf{S}(\rho)\mathbf{D}(\rho)\left\{\frac{\partial\mathbf{D}(\rho)}{\partial\rho}\mathbf{S}(\rho)^{\mathrm{T}}+\mathbf{D}(\rho)\left(-\mathbf{W}^{\mathrm{T}}\right)\right\}\left(\widehat{\mathbf{A}}-\mathbf{A}\right)\frac{\partial\mathbf{b}(\rho)}{\partial\rho} \\
& +\frac{2}{n}\mathbf{b}(\rho)^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{S}(\rho)\mathbf{D}(\rho)\left\{\frac{\partial\mathbf{D}(\rho)}{\partial\rho}\mathbf{S}(\rho)^{\mathrm{T}}+\mathbf{D}(\rho)\left(-\mathbf{W}^{\mathrm{T}}\right)\right\}\left(\widehat{\mathbf{A}}-\mathbf{A}\right)\frac{\partial\mathbf{b}(\rho)}{\partial\rho} \\
& +\frac{2}{n}\mathbf{b}(\rho)^{\mathrm{T}}\left(\widehat{\mathbf{A}}^{\mathrm{T}}-\mathbf{A}^{\mathrm{T}}\right)\mathbf{S}(\rho)\mathbf{D}(\rho)\left\{\frac{\partial\mathbf{D}(\rho)}{\partial\rho}\mathbf{S}(\rho)^{\mathrm{T}}+\mathbf{D}(\rho)\left(-\mathbf{W}^{\mathrm{T}}\right)\right\}\mathbf{A}\frac{\partial\mathbf{b}(\rho)}{\partial\rho}.
\end{aligned}
$$

We have

$$
\begin{aligned}
|\delta_{n1}| \quad \leqslant \quad & \left\{\left\|\frac{\partial\mathbf{D}(\rho)}{\partial\rho}\mathbf{S}(\rho)^{\mathrm{T}}\frac{1}{\sqrt{n}}(\mathbf{A}-\widehat{\mathbf{A}})\mathbf{b}(\rho)\right\|+\left\|\mathbf{D}(\rho)\left(-\mathbf{W}^{\mathrm{T}}\right)\frac{1}{\sqrt{n}}(\mathbf{A}-\widehat{\mathbf{A}})\mathbf{b}(\rho)\right\|\right. \\
& \left.+\left\|\mathbf{D}(\rho)\mathbf{S}(\rho)^{\mathrm{T}}\frac{1}{\sqrt{n}}(\mathbf{A}-\widehat{\mathbf{A}})\frac{\partial\mathbf{b}(\rho)}{\partial\rho}\right\|\right\}\left\|\left\{\frac{\partial\mathbf{D}(\rho)}{\partial\rho}\mathbf{M}(\rho)+\mathbf{D}(\rho)\frac{\partial\mathbf{M}(\rho)}{\partial\rho}\right\}\frac{1}{\sqrt{n}}\mathbf{Y}\right\| \\
\leqslant \quad & \left[\lambda_{\max}\left\{\frac{\partial\mathbf{D}(\rho)}{\partial\rho}\right\}\lambda_{\max}^{1/2}\left\{\mathbf{S}(\rho)\mathbf{S}(\rho)^{\mathrm{T}}\right\}\lambda_{\max}^{1/2}\left\{\frac{1}{n}(\mathbf{A}^{\mathrm{T}}-\widehat{\mathbf{A}}^{\mathrm{T}})(\mathbf{A}-\widehat{\mathbf{A}})\right\}\|\mathbf{b}(\rho)\|\right. \\
& \quad+\lambda_{\max}\left\{\mathbf{D}(\rho)\right\}\lambda_{\max}^{1/2}\left(\mathbf{W}^{\mathrm{T}}\mathbf{W}\right)\lambda_{\max}^{1/2}\left\{\frac{1}{n}(\mathbf{A}^{\mathrm{T}}-\widehat{\mathbf{A}}^{\mathrm{T}})(\mathbf{A}-\widehat{\mathbf{A}})\right\}\|\mathbf{b}(\rho)\| \\
& \quad\left.+\lambda_{\max}\left\{\mathbf{D}(\rho)\right\}\lambda_{\max}^{1/2}\left\{\mathbf{M}(\rho)\right\}\lambda_{\max}^{1/2}\left\{\frac{1}{n}(\mathbf{A}^{\mathrm{T}}-\widehat{\mathbf{A}}^{\mathrm{T}})(\mathbf{A}-\widehat{\mathbf{A}})\right\}\left\|\frac{\partial\mathbf{b}(\rho)}{\partial\rho}\right\|\right] \\
& \times\left[\lambda_{\max}\left\{\frac{\partial\mathbf{D}(\rho)}{\partial\rho}\right\}\lambda_{\max}\left\{\mathbf{M}(\rho)\right\}+\lambda_{\max}\left\{\mathbf{D}(\rho)\right\}\lambda_{\max}\left\{\frac{\partial\mathbf{M}(\rho)}{\partial\rho}\right\}\right]\frac{1}{\sqrt{n}}\|\mathbf{Y}\| \\
= \quad & o_p\left(n^{-1/2}\right),
\end{aligned}
$$

where the last equality is due to Conditions (C7)-(C8), Remark 2, and Lemma S2.5. Similarly,

$$
\begin{aligned}
& |\delta_{n2}| \\
\leqslant \quad & \left[\lambda_{\max}^{2}\left\{\frac{\partial\mathbf{D}(\rho)}{\partial\rho}\right\}\lambda_{\max}\left\{\mathbf{M}(\rho)\right\}\lambda_{\max}^{1/2}\left\{\mathbf{S}(\rho)\mathbf{S}(\rho)^{\mathrm{T}}\right\}\right. \\
& \quad+\lambda_{\max}\left\{\frac{\partial\mathbf{D}(\rho)}{\partial\rho}\right\}\lambda_{\max}\left\{\mathbf{M}(\rho)\right\}\lambda_{\max}\left\{\mathbf{D}(\rho)\right\}\lambda_{\max}^{1/2}\left(\mathbf{W}\mathbf{W}^{\mathrm{T}}\right)
\end{aligned}
$$

$$
\begin{aligned}
& \quad + \lambda_{\max}\left\{\mathbf{D}(\rho)\right\} \lambda_{\max}\left\{\frac{\partial \mathbf{M}(\rho)}{\partial \rho}\right\} \lambda_{\max}\left\{\frac{\partial \mathbf{D}(\rho)}{\partial \rho}\right\} \lambda_{\max}^{1/2}\left\{\mathbf{S}(\rho)\mathbf{S}(\rho)^{\mathrm{T}}\right\} \\
& \quad + \lambda_{\max}\left\{\mathbf{D}(\rho)^2\right\} \lambda_{\max}\left\{\frac{\partial \mathbf{M}(\rho)}{\partial \rho}\right\} \lambda_{\max}^{1/2}\left(\mathbf{W}\mathbf{W}^{\mathrm{T}}\right)\Bigg] \\
& \times \lambda_{\max}^{1/2}\left\{\frac{1}{n}(\mathbf{A}^{\mathrm{T}} - \widehat{\mathbf{A}}^{\mathrm{T}})(\mathbf{A} - \widehat{\mathbf{A}})\right\} \frac{1}{\sqrt{n}} \|\mathbf{Y}\| \|\mathbf{b}(\rho)\| \\
+ & \left[\lambda_{\max}\left\{\frac{\partial \mathbf{D}(\rho)}{\partial \rho}\right\} \lambda_{\max}^{1/2}\left\{\mathbf{S}(\rho)\mathbf{S}(\rho)^{\mathrm{T}}\right\} + \lambda_{\max}\left\{\mathbf{D}(\rho)\right\} \lambda_{\max}^{1/2}\left(\mathbf{W}\mathbf{W}^{\mathrm{T}}\right)\right] \\
& \times \Bigg[\lambda_{\max}\left\{\frac{\partial \mathbf{D}(\rho)}{\partial \rho}\right\} \lambda_{\max}^{1/2}\left\{\mathbf{S}(\rho)\mathbf{S}(\rho)^{\mathrm{T}}\right\} \lambda_{\max}\left\{\frac{1}{n}\left(\widehat{\mathbf{A}}^{\mathrm{T}} - \mathbf{A}^{\mathrm{T}}\right)\left(\widehat{\mathbf{A}} - \mathbf{A}\right)\right\} \|\mathbf{b}(\rho)\|^2 \\
& \quad + 2\lambda_{\max}\left\{\frac{\partial \mathbf{D}(\rho)}{\partial \rho}\right\} \lambda_{\max}^{1/2}\left\{\mathbf{S}(\rho)\mathbf{S}(\rho)^{\mathrm{T}}\right\} \lambda_{\max}^{1/2}\left(\frac{1}{n}\mathbf{A}^{\mathrm{T}}\mathbf{A}\right) \\
& \quad\quad \times \lambda_{\max}^{1/2}\left\{\frac{1}{n}\left(\widehat{\mathbf{A}}^{\mathrm{T}} - \mathbf{A}^{\mathrm{T}}\right)\left(\widehat{\mathbf{A}} - \mathbf{A}\right)\right\} \|\mathbf{b}(\rho)\|^2 \\
& \quad + \lambda_{\max}^{1/2}\left(\mathbf{W}\mathbf{W}^{\mathrm{T}}\right) \lambda_{\max}\left\{\mathbf{D}(\rho)\right\} \lambda_{\max}\left\{\frac{1}{n}\left(\widehat{\mathbf{A}}^{\mathrm{T}} - \mathbf{A}^{\mathrm{T}}\right)\left(\widehat{\mathbf{A}} - \mathbf{A}\right)\right\} \|\mathbf{b}(\rho)\|^2 \\
& \quad + 2\lambda_{\max}^{1/2}\left(\mathbf{W}\mathbf{W}^{\mathrm{T}}\right) \lambda_{\max}\left\{\mathbf{D}(\rho)\right\} \lambda_{\max}^{1/2}\left(\frac{1}{n}\mathbf{A}^{\mathrm{T}}\mathbf{A}\right) \\
& \quad\quad \times \lambda_{\max}^{1/2}\left\{\frac{1}{n}\left(\widehat{\mathbf{A}}^{\mathrm{T}} - \mathbf{A}^{\mathrm{T}}\right)\left(\widehat{\mathbf{A}} - \mathbf{A}\right)\right\} \|\mathbf{b}(\rho)\|^2 \\
& \quad + \lambda_{\max}\left\{\mathbf{D}(\rho)\right\} \lambda_{\max}^{1/2}\left\{\mathbf{S}(\rho)\mathbf{S}(\rho)\right\} \lambda_{\max}\left\{\frac{1}{n}\left(\widehat{\mathbf{A}}^{\mathrm{T}} - \mathbf{A}^{\mathrm{T}}\right)\left(\widehat{\mathbf{A}} - \mathbf{A}\right)\right\} \|\mathbf{b}(\rho)\| \left\|\frac{\partial \mathbf{b}(\rho)}{\partial \rho}\right\| \\
& \quad + 2\lambda_{\max}\left\{\mathbf{D}(\rho)\right\} \lambda_{\max}^{1/2}\left\{\mathbf{S}(\rho)\mathbf{S}(\rho)\right\} \lambda_{\max}^{1/2}\left(\frac{1}{n}\mathbf{A}^{\mathrm{T}}\mathbf{A}\right) \\
& \quad\quad \times \lambda_{\max}^{1/2}\left\{\frac{1}{n}\left(\widehat{\mathbf{A}}^{\mathrm{T}} - \mathbf{A}^{\mathrm{T}}\right)\left(\widehat{\mathbf{A}} - \mathbf{A}\right)\right\} \|\mathbf{b}(\rho)\| \left\|\frac{\partial \mathbf{b}(\rho)}{\partial \rho}\right\| \Bigg] \\
= \quad & o_p\left(n^{-1/2}\right),
\end{aligned}
$$

where the last equality is also due to Conditions (C7)-(C8), Remark 2, and Lemma S2.5. The same rate can also be similarly seen to hold for δ_{n3}, δ_{n4}, δ_{n6} and δ_{n6} and the details are omitted. This implies that (S.4) holds. □

S3.8 Proof of Lemma S2.8

For part (a), let $\mathbf{A} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^{\mathrm{T}}$ be the eigendecomposition of $\mathbf{A}$, then

$$
\begin{aligned}
\mathrm{trace}(\mathbf{AB}) &= \mathrm{trace}(\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^{\mathrm{T}}\mathbf{B}) \\
&= \mathrm{trace}(\boldsymbol{\Lambda}\mathbf{U}^{\mathrm{T}}\mathbf{B}\mathbf{U}) \\
&\leqslant \lambda_{\max}(\boldsymbol{\Lambda})\mathrm{trace}(\mathbf{U}^{\mathrm{T}}\mathbf{B}\mathbf{U}) \\
&= \lambda_{\max}(\mathbf{A})\mathrm{trace}(\mathbf{B}) \\
&\leqslant \mathrm{trace}(\mathbf{A})\mathrm{trace}(\mathbf{B}).
\end{aligned}
$$

The proof of the inequality in the other direction is similar, so we omit it. For part (b), we have

$$
\begin{aligned}
\lambda_{\max}\left\{(\mathbf{AB})^{\mathrm{T}}(\mathbf{AB})\right\} &= \max_{\|\mathbf{x}\|=1} \mathbf{x}^{\mathrm{T}}(\mathbf{AB})^{\mathrm{T}}(\mathbf{AB})\mathbf{x} \\
&= \max_{\|\mathbf{x}\|=1} \mathbf{x}^{\mathrm{T}}\mathbf{BAABx} \\
&\leqslant \lambda_{\max}\left(\mathbf{A}^2\right) \max_{\|\mathbf{x}\|=1} \mathbf{x}^{\mathrm{T}}\mathbf{BBx} \\
&\leqslant \lambda_{\max}\left(\mathbf{A}^2\right)\lambda_{\max}\left(\mathbf{B}^2\right) = \lambda_{\max}^2(\mathbf{A})\lambda_{\max}^2(\mathbf{B}).
\end{aligned}
$$

The proof part (c) can be found in Abadir and Magnus (2005). $\square$

S3.9 Proof of Lemma S2.9

Note that X_i's are independent, from which we have $E(X_iX_j) = 0$, $E(X_iX_jX_k) = 0$ and $E(X_i^2X_j) = 0$ for any $i \neq j \neq k \neq i$. Together with the conditions in Lemma S2.9, simple calculation implies that

$$\operatorname{cov}(Q_1, Q_2) = \operatorname{cov}(\mathbf{X}^{\mathrm{T}}\mathbf{M}_1\mathbf{X}, \mathbf{X}^{\mathrm{T}}\mathbf{M}_2\mathbf{X}) + \operatorname{cov}(\mathbf{U}_1^{\mathrm{T}}\mathbf{X}, \mathbf{U}_2^{\mathrm{T}}\mathbf{X}),$$

where $\operatorname{cov}(\mathbf{U}_1^{\mathrm{T}}\mathbf{X}, \mathbf{U}_2^{\mathrm{T}}\mathbf{X}) = \mathbf{U}_1^{\mathrm{T}}\mathbf{U}_2$. Since $E(\mathbf{X}^{\mathrm{T}}\mathbf{M}_1\mathbf{X}) = \operatorname{trace}(\mathbf{M}_1)$ by Lemma S2.7, we have

$$\begin{aligned}
& \operatorname{cov}\left\{\left(\mathbf{X}^{\mathrm{T}}\mathbf{M}_1\mathbf{X}\right)\left(\mathbf{X}^{\mathrm{T}}\mathbf{M}_2\mathbf{X}\right)\right\} \\
= \; & E\left\{\left(\mathbf{X}^{\mathrm{T}}\mathbf{M}_1\mathbf{X}\right)\left(\mathbf{X}^{\mathrm{T}}\mathbf{M}_2\mathbf{X}\right)\right\} - E(\mathbf{X}^{\mathrm{T}}\mathbf{M}_1\mathbf{X})E(\mathbf{X}^{\mathrm{T}}\mathbf{M}_2\mathbf{X}) \\
= \; & \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{\ell=1}^{n} m_{1,ij}m_{2,k\ell}E\left(X_iX_jX_kX_\ell\right) - \operatorname{trace}(\mathbf{M}_1)\operatorname{trace}(\mathbf{M}_2) \\
= \; & \left(\sum_{i\neq j}^{n} m_{1,ii}m_{2,jj} + \sum_{i\neq j}^{n} m_{1,ij}m_{2,ij} + \sum_{i\neq j}^{n} m_{1,ij}m_{2,ji}\right) E\left(X_i^2X_j^2\right) \\
& + \sum_{i=1}^{n} m_{1,ii}m_{2,ii}E(X_i^4) - \operatorname{trace}(\mathbf{M}_1)\operatorname{trace}(\mathbf{M}_2) \\
= \; & \left\{\operatorname{trace}(\mathbf{M}_1)\operatorname{trace}(\mathbf{M}_2) + \operatorname{trace}(\mathbf{M}_1\mathbf{M}_2^{\mathrm{T}}) + \operatorname{trace}(\mathbf{M}_1\mathbf{M}_2)\right\} E(X_1^2X_2^2) \\
& + \operatorname{trace}\left\{\operatorname{diag}(\mathbf{M}_1)\operatorname{diag}(\mathbf{M}_2)\right\}\left\{E(X_1^4) - 3E(X_1^2X_2^2)\right\} - \operatorname{trace}(\mathbf{M}_1)\operatorname{trace}(\mathbf{M}_2).
\end{aligned}$$

Hence,

$$\begin{aligned}
\operatorname{cov}(Q_1, Q_2) = \; & \left\{\operatorname{trace}(\mathbf{M}_1)\operatorname{trace}(\mathbf{M}_2) + \operatorname{trace}(\mathbf{M}_1\mathbf{M}_2^{\mathrm{T}}) + \operatorname{trace}(\mathbf{M}_1\mathbf{M}_2)\right\} E(X_1^2X_2^2) \\
& + \operatorname{trace}\left\{\operatorname{diag}(\mathbf{M}_1)\operatorname{diag}(\mathbf{M}_2)\right\}\left\{E(X_1^4) - 3E(X_1^2X_2^2)\right\} \\
& - \operatorname{trace}(\mathbf{M}_1)\operatorname{trace}(\mathbf{M}_2) + \mathbf{U}_1\mathbf{U}_2^{\mathrm{T}}.
\end{aligned}$$

Again, since X_i's are independent, we have $E(X_i^2X_j^2) = E(X_i^2)E(X_j^2) = 1$ for $i \neq j$. This completes the proof. $\square$

S3.10 Proof of Theorem 2

We establish Theorem 2 in three steps. First, we expand $\widehat{b}_k(\rho_0) - b_k(\rho_0)$ as independent sums. Second, we verify that $\widehat{\rho}$ is $n^{1/4}$-consistent. Third, using the conclusions in the first two steps, we derive the asymptotic normality of $n^{1/4}(\widehat{\rho} - \rho_0)$.

Expand $\widehat{b}_k(\rho_0) - b_k(\rho_0)$ as independent sums

To employ the central limit theorem in the third step, we need to write $\widehat{b}_k(\rho_0) - b_k(\rho_0)$ as independent sums. So we first expand this term.

Recall that $\widehat{b}_k(\rho) = \widehat{\lambda}_k^{-1}\langle\widehat{\phi}_k, \widehat{g}\rangle$ and $b_k = \lambda_k^{-1}\langle\phi_k, g\rangle$. By Taylor's expansion, we have

$$
\begin{aligned}
& \widehat{b}_k(\rho_0) - b_k(\rho_0) \\
= \;& \widehat{\lambda}_k^{-1}\langle\widehat{\phi}_k, \widehat{g}\rangle - \lambda_k^{-1}\langle\phi_k, g\rangle \\
= \;& \left(\widehat{\lambda}_k^{-1} - \lambda_k^{-1}\right)\langle\widehat{\phi}_k - \phi_k, \widehat{g}\rangle + \left(\widehat{\lambda}_k^{-1} - \lambda_k^{-1}\right)\langle\phi_k, \widehat{g}\rangle + \lambda_k^{-1}\langle\widehat{\phi}_k - \phi_k, \widehat{g}\rangle + \lambda_k^{-1}\langle\phi_k, \widehat{g} - g\rangle \\
= \;& -\lambda_k^{-2}\left(\widehat{\lambda}_k - \lambda_k\right)\langle\widehat{\phi}_k - \phi_k, \widehat{g}\rangle + O_p\left\{\left(\widehat{\lambda}_k - \lambda_k\right)^2\right\}\langle\widehat{\phi}_k - \phi_k, \widehat{g}\rangle \\
& -\lambda_k^{-2}\left(\widehat{\lambda}_k - \lambda_k\right)\langle\phi_k, \widehat{g}\rangle + O_p\left\{\left(\widehat{\lambda}_k - \lambda_k\right)^2\right\}\langle\phi_k, \widehat{g}\rangle + \lambda_k^{-1}\langle\widehat{\phi}_k - \phi_k, \widehat{g}\rangle + \lambda_k^{-1}\langle\phi_k, \widehat{g} - g\rangle \\
= \;& \lambda_k^{-1}\langle\phi_k, \widehat{g} - g\rangle + \lambda_k^{-1}\langle\widehat{\phi}_k - \phi_k, g\rangle - \lambda_k^{-2}\left(\widehat{\lambda}_k - \lambda_k\right)\langle\phi_k, g\rangle \\
& +\lambda_k^{-1}\langle\widehat{\phi}_k - \phi_k, \widehat{g} - g\rangle - \lambda_k^{-2}\left(\widehat{\lambda}_k - \lambda_k\right)\langle\phi_k, \widehat{g} - g\rangle \\
& -\lambda_k^{-2}\left(\widehat{\lambda}_k - \lambda_k\right)\langle\widehat{\phi}_k - \phi_k, g\rangle - \lambda_k^{-2}\left(\widehat{\lambda}_k - \lambda_k\right)\langle\widehat{\phi}_k - \phi_k, \widehat{g} - g\rangle \\
& +O_p\left\{\left(\widehat{\lambda}_k - \lambda_k\right)^2\right\}\langle\widehat{\phi}_k - \phi_k, g\rangle + O_p\left\{\left(\widehat{\lambda}_k - \lambda_k\right)^2\right\}\langle\widehat{\phi}_k - \phi_k, \widehat{g} - g\rangle \\
& +O_p\left\{\left(\widehat{\lambda}_k - \lambda_k\right)^2\right\}\langle\phi_k, g\rangle + O_p\left\{\left(\widehat{\lambda}_k - \lambda_k\right)^2\right\}\langle\phi_k, \widehat{g} - g\rangle \\
= \;& \lambda_k^{-1}\langle\phi_k, \widehat{g} - g\rangle + \lambda_k^{-1}\langle\widehat{\phi}_k - \phi_k, g\rangle - \lambda_k^{-2}\left(\widehat{\lambda}_k - \lambda_k\right)\langle\phi_k, g\rangle \\
& +A_{k1} - A_{k2} - A_{k3} - A_{k4} + A_{k5} + A_{k6} + A_{k7} + A_{k8},
\end{aligned}
$$

where $A_{k1} = \lambda_k^{-1}\langle\widehat{\phi}_k - \phi_k, \widehat{g} - g\rangle$, $A_{k2} = \lambda_k^{-2}(\widehat{\lambda}_k - \lambda_k)\langle\phi_k, \widehat{g} - g\rangle$, $A_{k3} = \lambda_k^{-2}(\widehat{\lambda}_k - \lambda_k)\langle\widehat{\phi}_k - \phi_k, g\rangle$, $A_{k4} = \lambda_k^{-2}(\widehat{\lambda}_k - \lambda_k)\langle\widehat{\phi}_k - \phi_k, \widehat{g} - g\rangle$, $A_{k5} = O_p\{(\widehat{\lambda}_k - \lambda_k)^2\}\langle\widehat{\phi}_k - \phi_k, g\rangle$, $A_{k6} = O_p\{(\widehat{\lambda}_k - \lambda_k)^2\}\langle\widehat{\phi}_k - \phi_k, \widehat{g} - g\rangle$, $A_{k7} = O_p\{(\widehat{\lambda}_k - \lambda_k)^2\}\langle\phi_k, g\rangle$, $A_{k8} = O_p\{(\widehat{\lambda}_k - \lambda_k)^2\}\langle\phi_k, \widehat{g} - g\rangle$. We next use Cauchy-Schwarz inequality, Condition (C2), Lemma S2.4(c), Lemma S2.1(c) and Lemma S2.3 repetitively to show the order of $A_{k1} - A_{k8}$.

$$
\begin{aligned}
A_{k1} &= \lambda_k^{-1}\langle\widehat{\phi}_k - \phi_k, \widehat{g} - g\rangle \\
&\leqslant \lambda_k^{-1}\|\widehat{\phi}_k - \phi_k\|\|\widehat{g} - g\| \\
&= O_p(n^{-1}k^{\alpha+1}),
\end{aligned}
$$

$$
\begin{aligned}
A_{k2} &= \lambda_k^{-2}\left(\widehat{\lambda}_k - \lambda_k\right)\langle\phi_k, \widehat{g} - g\rangle \\
&\leqslant \lambda_k^{-2}|\widehat{\lambda}_k - \lambda_k|\|\phi_k\|\|\widehat{g} - g\| \\
&= O_p(n^{-1}k^{\alpha})\|\phi_k\| \\
&= O_p(n^{-1}k^{\alpha}),
\end{aligned}
$$

where the last equality is due to $\int \phi_k^2(t)dt = 1$. By Condition (C1),

$$
\begin{aligned}
A_{k3} &= \lambda_k^{-2}\left(\widehat{\lambda}_k - \lambda_k\right)\langle\widehat{\phi}_k - \phi_k, g\rangle \\
&\leqslant \lambda_k^{-2}|\widehat{\lambda}_k - \lambda_k|\|\widehat{\phi}_k - \phi_k\|\|g\| \\
&= O_p(n^{-1}k^{\alpha+1})\|g\| \\
&= O_p(n^{-1}k^{\alpha+1}),
\end{aligned}
$$

and

$$
\begin{aligned}
A_{k4} &= \lambda_k^{-2}\left(\widehat{\lambda}_k-\lambda_k\right)\langle\widehat{\phi}_k-\phi_k,\widehat{g}-g\rangle \\
&\leqslant \lambda_k^{-2}|\widehat{\lambda}_k-\lambda_k|\|\widehat{\phi}_k-\phi_k\|\|\widehat{g}-g\| \\
&= O_p(n^{-3/2}k^{\alpha+1}).
\end{aligned}
$$

Similarly, it is easy to see that $A_{k5}=O_p(n^{-3/2}k^{-2\alpha+1})$, $A_{k6}=O_p(n^{-2}k^{-2\alpha+1})$, $A_{k7}=O_p(n^{-1}k^{-2\alpha})$ and $A_{k8}=O_p(n^{-3/2}k^{-2\alpha})$. Thus, we observe that

$$
\begin{aligned}
& \widehat{b}_k(\rho_0)-b_k(\rho_0) \\
= & \lambda_k^{-1}\langle\phi_k,\widehat{g}-g\rangle+\lambda_k^{-1}\langle\widehat{\phi}_k-\phi_k,g\rangle-\lambda_k^{-2}\left(\widehat{\lambda}_k-\lambda_k\right)\langle\phi_k,g\rangle \\
& +O_p(n^{-1}k^{\alpha+1})+O_p(n^{-1}k^{\alpha}) \\
& +O_p(n^{-1}k^{\alpha+1})+O_p(n^{-3/2}k^{\alpha+1}) \\
& +O_p(n^{-3/2}k^{-2\alpha+1})+O_p(n^{-2}k^{-2\alpha+1}) \\
& +O_p(n^{-1}k^{-2\alpha})+O_p(n^{-3/2}k^{-2\alpha}) \\
= & \lambda_k^{-1}\langle\phi_k,\widehat{g}-g\rangle+\lambda_k^{-1}\langle\widehat{\phi}_k-\phi_k,g\rangle-\lambda_k^{-2}\left(\widehat{\lambda}_k-\lambda_k\right)\langle\phi_k,g\rangle+O_p(n^{-1}k^{\alpha+1}) \\
\equiv & M_1.
\end{aligned}
$$

Recall that

$$
\begin{aligned}
\widehat{g}(t) &= n^{-1}\sum_{i=1}^{n}\left(Y_i-\rho\sum_{j=1}^{n}w_{ij}Y_j\right)Z_i(t) \\
&= n^{-1}\sum_{i=1}^{n}Y_i^{*}Z_i(t),
\end{aligned}
$$

and

$$g(t) = E\{Y_i^* Z_i(t)\},$$

where Y_i^*'s are independent and identically distributed according to (2.6). Note that Condition (C1) implies that $E\{Y_i^* Z_i(t)\} < \infty$. By combining Lemma S2.4(b), Condition (C2) and Condition (C5), it follows that

$$\begin{aligned}
& M_1 \\
= \quad & n^{-1} \sum_{i=1}^{n} \lambda_k^{-1} \int \phi_k(t) \left[Y_i^* Z_i(t) - E\left\{Y_i^* Z_i(t)\right\}\right] dt \\
& + \lambda_k^{-1} \int E\left\{Y_i^* Z_i(t)\right\} \left\{ \sum_{\ell \neq k} (\lambda_k - \lambda_\ell)^{-1} \phi_\ell \langle (\widehat{G} - G)\phi_\ell, \phi_k \rangle + O_p(n^{-1} k^{\alpha+2} \log k) \right\} dt \\
& - \lambda_k^{-2} \left\{ \frac{1}{n} \sum_{i=1}^{n} (\xi_{ik}^2 - \lambda_k) + O_p(k/n) \right\} \langle \phi_k, g \rangle + O_p(n^{-1} k^{\alpha+1}) \\
= \quad & n^{-1} \sum_{i=1}^{n} \lambda_k^{-1} \int \phi_k(t) \left[Y_i^* Z_i(t) - E\left\{Y_i^* Z_i(t)\right\}\right] dt \\
& + \lambda_k^{-1} \int E\left\{Y_i^* Z_i(t)\right\} \left[\sum_{\ell:\ell \neq k} (\lambda_k - \lambda_\ell)^{-1} \phi_k(t) \right. \\
& \times \left. \iint \left\{ \frac{1}{n} \sum_{i=1}^{n} Z_i(u) Z_i(v) - G(u,v) \right\} \phi_k(u) \phi_\ell(v) du dv + O_p(n^{-1} k^{\alpha+2} \log k) \right] dt \\
& - \lambda_k^{-2} \left\{ \frac{1}{n} \sum_{i=1}^{n} (\xi_{ik}^2 - \lambda_k) + O_p(k/n) \right\} \int \phi_k(t) g(t) dt + O_p(n^{-1} k^{\alpha+1}) \\
= \quad & \frac{1}{n} \sum_{i=1}^{n} \lambda_k^{-1} \int \phi_k(t) \left[Y_i^* Z_i(t) - E\left\{Y_i^* Z_i(t)\right\}\right] dt \\
& + \frac{1}{n} \sum_{i=1}^{n} \lambda_k^{-1} \int E\left\{Y_i^* Z_i(t)\right\} \left[\sum_{\ell:\ell \neq k} (\lambda_k - \lambda_\ell)^{-1} \phi_k(t) \iint \left\{ Z_i(u) Z_i(v) - G(u,v) \right\} \right. \\
& \left. \times \phi_k(u) \phi_\ell(v) du dv \right] dt - \frac{1}{n} \sum_{i=1}^{n} \lambda_k^{-2} (\xi_{ik}^2 - \lambda_k) \int \phi_k(t) g(t) dt
\end{aligned}$$

$$+O_p(n^{-1}k^{2\alpha+2}\log k) + O_p(n^{-1}k^{2\alpha+1}) + O_p(n^{-1}k^{\alpha+1}).$$

Hence, we have

$$\widehat{b}_k(\rho_0) - b_k(\rho_0) \quad = \quad \frac{1}{n}\sum_{i=1}^{n}\varphi_{ik} + O_p(n^{-1}k^{2\alpha+2}\log k), \tag{S.27}$$

where

$$\begin{aligned}\varphi_{ik} \quad = \quad & \lambda_k^{-1}\int \phi_k(t)\left[Y_i^* Z_i(t) - E\left\{Y_i^* Z_i(t)\right\}\right]dt + \lambda_k^{-1}\int E\left\{Y_i^* Z_i(t)\right\} \\ & \times \left\{\sum_{\ell:\ell\neq k}(\lambda_k - \lambda_\ell)^{-1}\phi_k(t)\iint\left[Z_i(u)Z_i(v) - E\left\{Z_i(u)Z_i(v)\right\}\right]\phi_k(u)\phi_\ell(v)dudv\right\}dt \\ & -\lambda_k^{-2}(\xi_{ik}^2 - E\xi_{ik}^2)\int \phi_k(t)g(t)dt. \end{aligned} \tag{S.28}$$

The proof of $n^{1/4}$-consistency

To prove $\widehat{\rho}$ is $n^{1/4}$-consistent, by the technique of Fan and Li (2001), it is sufficient to show that, for any $\epsilon > 0$, there exists a constant $0 < C < \infty$, such that $Q_1\{\widehat{\mathbf{A}}; \widehat{\mathbf{b}}(\rho), \rho\}$ in (2.12) satisfies

$$\lim_{n\to\infty} \Pr\left[\inf_{|u|=C} Q_1\left\{\widehat{\mathbf{A}}; \widehat{\mathbf{b}}(\rho_0 + n^{-1/4}u), \rho_0 + n^{-1/4}u\right\} > Q_1\left\{\widehat{\mathbf{A}}; \widehat{\mathbf{b}}(\rho_0), \rho_0\right\}\right] \geqslant 1 - \epsilon. \tag{S.29}$$

To this end, we recall that from (2.14), for any ρ^*,

$$\begin{aligned}\frac{\mathrm{d}Q_1\{\widehat{\mathbf{A}}; \widehat{\mathbf{b}}(\rho), \rho\}}{\mathrm{d}\rho}\Big|_{\rho=\rho^*} \quad &= \quad \left[\frac{\partial Q_1\{\widehat{\mathbf{A}}; \widehat{\mathbf{b}}(\rho), \rho\}}{\partial \rho} + \frac{\partial Q_1\{\widehat{\mathbf{A}}; \widehat{\mathbf{b}}(\rho), \rho\}}{\partial \widehat{\mathbf{b}}(\rho)^{\mathrm{T}}}\frac{\partial \widehat{\mathbf{b}}(\rho)}{\partial \rho}\right]\Big|_{\rho=\rho^*} \\ &= \quad \frac{1}{n}\sum_{i=1}^{n} S_i\{\widehat{\mathbf{A}}; \widehat{\mathbf{b}}(\rho^*), \rho^*\},\end{aligned}$$

where $n^{-1}\sum_{i=1}^{n} S_i\{\widehat{\mathbf{A}}; \widehat{\mathbf{b}}(\rho^*), \rho^*\}$ is given in (2.15) with $\mathbf{b}$, ρ and $1/\sqrt{n}$ are replaced by $\widehat{\mathbf{b}}$, ρ^*

and $1/n$, respectively. Then, by Taylor's expansion and $|u| = C$, we have

$$\begin{aligned}
& Q_1\left\{\widehat{\mathbf{A}};\widehat{\mathbf{b}}(\rho_0+n^{-1/4}u),\rho_0+n^{-1/4}u\right\} - Q_1\left\{\widehat{\mathbf{A}};\widehat{\mathbf{b}}(\rho_0),\rho_0\right\} \\
= & \frac{\mathrm{d}Q_1\{\widehat{\mathbf{A}};\widehat{\mathbf{b}}(\rho_0),\rho_0\}}{\mathrm{d}\rho} n^{-1/4}u + \frac{\mathrm{d}^2Q_1\{\widehat{\mathbf{A}};\widehat{\mathbf{b}}(\rho_0),\rho_0\}}{\mathrm{d}\rho^2} 2^{-1}n^{-1/2}u^2 + o_p\left(n^{-1/2}u^2\right) \\
= & \left[\frac{1}{n}\sum_{i=1}^n S_i\{\widehat{\mathbf{A}};\widehat{\mathbf{b}}(\rho_0),\rho_0\}\right] n^{-1/4}u + \left[\frac{1}{n}\sum_{i=1}^n \frac{\mathrm{d}S_i\{\widehat{\mathbf{A}};\widehat{\mathbf{b}}(\rho_0),\rho_0\}}{\mathrm{d}\rho}\right] 2^{-1}n^{-1/2}u^2 + o_p(n^{-1/2}) \\
= & \left[\frac{1}{n}\sum_{i=1}^n S_i\{\widehat{\mathbf{A}};\widehat{\mathbf{b}}(\rho_0),\rho_0\}\right] n^{-1/4}u + \left[\frac{1}{n}\sum_{i=1}^n \frac{\mathrm{d}S_i\{\widehat{\mathbf{A}};\widehat{\mathbf{b}}(\rho_0),\rho_0\}}{\mathrm{d}\rho}\right] 2^{-1}n^{-1/2}C^2 + o_p(n^{-1/2}) \\
\geqslant & \left[\frac{1}{n}\sum_{i=1}^n \frac{\mathrm{d}S_i\{\widehat{\mathbf{A}};\widehat{\mathbf{b}}(\rho_0),\rho_0\}}{\mathrm{d}\rho}\right] 2^{-1}n^{-1/2}C^2 - \left|\frac{1}{n}\sum_{i=1}^n S_i\{\widehat{\mathbf{A}};\widehat{\mathbf{b}}(\rho_0),\rho_0\}\right| n^{-1/4}C + o_p(n^{-1/2}) \\
= & n^{-1/2}\left(\left[\frac{1}{n}\sum_{i=1}^n \frac{\mathrm{d}S_i\{\widehat{\mathbf{A}};\widehat{\mathbf{b}}(\rho_0),\rho_0\}}{\mathrm{d}\rho}\right] 2^{-1}C^2 - \left|n^{-3/4}\sum_{i=1}^n S_i\{\widehat{\mathbf{A}};\widehat{\mathbf{b}}(\rho_0),\rho_0\}\right| C\right) \\
& + o_p(n^{-1/2}). \qquad \text{(S.30)}
\end{aligned}$$

To establish (S.29), by (S.30), it suffices to show that

$$n^{-3/4}\sum_{i=1}^n S_i\{\widehat{\mathbf{A}};\widehat{\mathbf{b}}(\rho_0),\rho_0\} \quad = \quad O_p(1),$$

and

$$\frac{1}{n}\sum_{i=1}^n \frac{\mathrm{d}S_i\{\widehat{\mathbf{A}};\widehat{\mathbf{b}}(\rho_0),\rho_0\}}{\mathrm{d}\rho} \quad \xrightarrow{\text{Pr}} \quad c_*,$$

where c_* is a positive constant.

We now prove the above equations separately as follows. First, by mean value theorem,

$$\begin{aligned}
& n^{-3/4}\sum_{i=1}^n S_i\{\widehat{\mathbf{A}};\widehat{\mathbf{b}}(\rho_0),\rho_0\} \\
= & n^{-3/4}\sum_{i=1}^n S_i\{\widehat{\mathbf{A}};\mathbf{b}(\rho_0),\rho_0\} + n^{-3/4}\sum_{i=1}^n \frac{\partial S_i\{\widehat{\mathbf{A}};\mathbf{b}^*(\rho_0),\rho_0\}}{\partial \mathbf{b}^*(\rho_0)^{\mathrm{T}}}\left\{\widehat{\mathbf{b}}(\rho_0)-\mathbf{b}(\rho_0)\right\}
\end{aligned}$$

$$
\begin{aligned}
&= n^{-3/4}\sum_{i=1}^{n} S_i\{\mathbf{A};\mathbf{b}(\rho_0),\rho_0\} + \left[n^{-3/4}\sum_{i=1}^{n} S_i\{\widehat{\mathbf{A}};\mathbf{b}(\rho_0),\rho_0\} - n^{-3/4}\sum_{i=1}^{n} S_i\{\mathbf{A};\mathbf{b}(\rho_0),\rho_0\}\right] \\
&\quad + n^{-3/4}\sum_{i=1}^{n} \frac{\partial S_i\{\widehat{\mathbf{A}};\mathbf{b}(\rho_0),\rho_0\}}{\partial \mathbf{b}(\rho_0)^{\mathrm{T}}}\left\{\widehat{\mathbf{b}}(\rho_0) - \mathbf{b}(\rho_0)\right\} + o_p(n^{-1/4}) \\
&= n^{-3/4}\sum_{i=1}^{n} S_i\{\mathbf{A};\mathbf{b}(\rho_0),\rho_0\} + \left[n^{-3/4}\sum_{i=1}^{n} S_i\{\widehat{\mathbf{A}};\mathbf{b}(\rho_0),\rho_0\} - n^{-3/4}\sum_{i=1}^{n} S_i\{\mathbf{A};\mathbf{b}(\rho_0),\rho_0\}\right] \\
&\quad + \sum_{k=1}^{K}\left[n^{-1}\sum_{i=1}^{n}\frac{\partial S_i\{\mathbf{A};\mathbf{b}(\rho_0),\rho_0\}}{\partial b_k(\rho_0)}\right]\left(n^{-3/4}\sum_{j=1}^{n}\varphi_{jk}\right) + o_p(n^{-1/4}) \\
&= n^{-3/4}\sum_{i=1}^{n} S_i\{\mathbf{A};\mathbf{b}(\rho_0),\rho_0\} + \left[n^{-3/4}\sum_{i=1}^{n} S_i\{\widehat{\mathbf{A}};\mathbf{b}(\rho_0),\rho_0\} - n^{-3/4}\sum_{i=1}^{n} S_i\{\mathbf{A};\mathbf{b}(\rho_0),\rho_0\}\right] \\
&\quad + \sum_{k=1}^{K}\mathscr{B}_k(\mathbf{A})\left(n^{-3/4}\sum_{j=1}^{n}\varphi_{jk}\right) + o_p(1),
\end{aligned}
$$

where the second equality is by (S.34), the third equality is by (S.35), the last equality is due to (S.44) and $\mathscr{B}_k(\mathbf{A})$ is defined in (S.42). For the second term of above equation, by (S.2), we have

$$
n^{-3/4}\sum_{i=1}^{n} S_i\{\widehat{\mathbf{A}};\mathbf{b}(\rho_0),\rho_0\} - n^{-3/4}\sum_{i=1}^{n} S_i\{\mathbf{A};\mathbf{b}(\rho_0),\rho_0\} = O_p(n^{-1/4}).
$$

For the third term, by (S.46), we have

$$
\begin{aligned}
& E\left[n^{-3/4}\sum_{j=1}^{n}\left\{\sum_{k=1}^{K}\mathscr{B}_k(\mathbf{A})\varphi_{jk}\right\}\right]^2 \\
= \; & \mathrm{var}\left[n^{-3/4}\sum_{j=1}^{n}\left\{\sum_{k=1}^{K}\mathscr{B}_k(\mathbf{A})\varphi_{jk}\right\}\right] \\
= \; & O(1),
\end{aligned}
$$

where the last equality is by (S.36) and (S.43). Hence, we have

$$
n^{-3/4}\sum_{i=1}^{n} S_i\{\widehat{\mathbf{A}};\widehat{\mathbf{b}}(\rho_0),\rho_0\} = n^{-3/4}\sum_{i=1}^{n} S_i\{\mathbf{A};\mathbf{b}(\rho_0),\rho_0\} + O_p(1)
$$

$$
\begin{aligned}
&= O_p(n^{-1/4}) + O_p(1) \\
&= O_p(1),
\end{aligned}
$$

where the second equality is by (S.47) and (S.59). Second, by (S.4) with $\mathbf{b}(\rho)$ and ρ replaced by $\widehat{\mathbf{b}}(\rho_0)$ and ρ_0, respectively,

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\frac{\mathrm{d}S_i\{\widehat{\mathbf{A}};\widehat{\mathbf{b}}(\rho_0),\rho_0\}}{\mathrm{d}\rho_0} &= \frac{1}{n}\sum_{i=1}^{n}\frac{\mathrm{d}S_i\{\mathbf{A};\widehat{\mathbf{b}}(\rho_0),\rho_0\}}{\mathrm{d}\rho_0} + o_p\left(n^{-1/2}\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}\frac{\mathrm{d}S_i\{\mathbf{A};\mathbf{b}(\rho_0),\rho_0\}}{\mathrm{d}\rho_0} + o_p(1), \qquad \text{(S.31)}
\end{aligned}
$$

where the second inequality is due to (S.63). Then, by (S.31) and the proof of (S.65), we have

$$
\frac{1}{n}\sum_{i=1}^{n}\frac{\mathrm{d}S_i\{\widehat{\mathbf{A}};\widehat{\mathbf{b}}(\rho),\rho\}}{\mathrm{d}\rho} \mid \mathbf{A} \ \xrightarrow{\text{Pr}} \ c_*.
$$

Based on above results, for given $\mathbf{A}$, this establishes the (S.29).

The proof of asymptotic normality

For preparation, we recall from (2.11) that

$$
\widehat{b}_k(\rho) = n^{-1}\sum_{i=1}^{n}\widehat{\lambda}_k^{-1}\int\widehat{\phi}_k(t)\left(Y_i - \rho\sum_{j=1}^{n}w_{ij}Y_j\right)Z_i(t)dt.
$$

Note that $\widehat{b}_k(\rho)$ is a linear function of ρ, so

$$
\begin{aligned}
\frac{\partial\widehat{b}_k(\rho)}{\partial\rho} &= \partial\left\{n^{-1}\sum_{i=1}^{n}\widehat{\lambda}_k^{-1}\int\widehat{\phi}_k(t)\left(Y_i - \rho\sum_{j=1}^{n}w_{ij}Y_j\right)Z_i(t)dt\right\}\Big/\partial\rho \\
&= -n^{-1}\sum_{i=1}^{n}\widehat{\lambda}_k^{-1}\int\widehat{\phi}_k(t)\sum_{j=1}^{n}w_{ij}Y_jZ_i(t)dt
\end{aligned}
$$

and $\partial^2 \widehat{b}_k(\rho)/\partial \rho^2 = 0$. Since $\widehat{\rho}$ minimizes $Q_1\{\widehat{\mathbf{A}}; \widehat{\mathbf{b}}(\rho), \rho\}$ based on (2.12), it satisfies

$$\frac{\mathrm{d}Q_1\{\widehat{\mathbf{A}}; \widehat{\mathbf{b}}(\rho), \rho\}}{\mathrm{d}\rho}\Big|_{\rho=\widehat{\rho}} = \left[\frac{\partial Q_1\{\widehat{\mathbf{A}}; \widehat{\mathbf{b}}(\rho), \rho\}}{\partial \rho} + \frac{\partial Q_1\{\widehat{\mathbf{A}}; \widehat{\mathbf{b}}(\rho), \rho\}}{\partial \widehat{\mathbf{b}}(\rho)^{\mathrm{T}}} \frac{\partial \widehat{\mathbf{b}}(\rho)}{\partial \rho}\right]\Big|_{\rho=\widehat{\rho}} = 0.$$

Similar to (2.14), this is equivalently written as

$$n^{-3/4} \sum_{i=1}^{n} S_i\{\widehat{\mathbf{A}}; \widehat{\mathbf{b}}(\widehat{\rho}), \widehat{\rho}\} = 0,$$

where $S_i\{\widehat{\mathbf{A}}; \widehat{\mathbf{b}}(\widehat{\rho}), \widehat{\rho}\}$ is obtained from replacing $\mathbf{b}$ and ρ in formula (2.16) with $\widehat{\mathbf{b}}$ and $\widehat{\rho}$, respectively. Using Taylor's expansion,

$$\begin{aligned}
0 &= n^{-3/4} \sum_{i=1}^{n} S_i\{\widehat{\mathbf{A}}; \widehat{\mathbf{b}}(\widehat{\rho}), \widehat{\rho}\} \\
&= n^{-3/4} \sum_{i=1}^{n} S_i\{\widehat{\mathbf{A}}; \widehat{\mathbf{b}}(\rho_0), \rho_0\} + n^{-3/4} \sum_{i=1}^{n} \frac{\mathrm{d}S_i\{\widehat{\mathbf{A}}; \widehat{\mathbf{b}}(\rho^*), \rho^*\}}{\mathrm{d}\rho^*}(\widehat{\rho} - \rho_0) \\
&= n^{-3/4} \sum_{i=1}^{n} S_i\{\widehat{\mathbf{A}}; \mathbf{b}(\rho_0), \rho_0\} + n^{-3/4} \sum_{i=1}^{n} \frac{\partial S_i\{\widehat{\mathbf{A}}; \mathbf{b}^*(\rho_0), \rho_0\}}{\partial \mathbf{b}^*(\rho_0)^{\mathrm{T}}} \left\{\widehat{\mathbf{b}}(\rho_0) - \mathbf{b}(\rho_0)\right\} \\
&\quad + n^{-3/4} \sum_{i=1}^{n} \frac{\mathrm{d}S_i\{\widehat{\mathbf{A}}; \widehat{\mathbf{b}}(\rho^*), \rho^*\}}{\mathrm{d}\rho^*}(\widehat{\rho} - \rho_0) \\
&= n^{-3/4} \sum_{i=1}^{n} \left[S_i\{\widehat{\mathbf{A}}; \mathbf{b}(\rho_0), \rho_0\} + \sum_{k=1}^{K} \frac{\partial S_i\{\widehat{\mathbf{A}}; \mathbf{b}^*(\rho_0), \rho_0\}}{\partial b_k^*(\rho_0)} \left\{\widehat{b}_k(\rho_0) - b_k(\rho_0)\right\}\right] \\
&\quad + n^{-3/4} \sum_{i=1}^{n} \frac{\mathrm{d}S_i\{\widehat{\mathbf{A}}; \widehat{\mathbf{b}}(\rho^*), \rho^*\}}{\mathrm{d}\rho^*}(\widehat{\rho} - \rho_0), \qquad \text{(S.32)}
\end{aligned}$$

where ρ^* denotes the value between the ρ_0 and $\widehat{\rho}$ and $\mathbf{b}^*(\rho_0)$ between the $\mathbf{b}(\rho_0)$ and $\widehat{\mathbf{b}}(\rho_0)$, respectively. $S_i\{\widehat{\mathbf{A}}; \rho_0, \mathbf{b}^*(\rho_0)\}$ and $\partial S_i\{\widehat{\mathbf{A}}; \rho_0, \mathbf{b}^*(\rho_0)\}/\partial b_k^*(\rho_0)$ are given (2.16) and (2.17), respectively, with $\mathbf{b}(\rho)$ and ρ replaced by $\mathbf{b}^*(\rho_0)$ and ρ_0.

We calculate the expectation and variance of the first term of (S.32) below. Applying

Taylor's expansion and Condition (C6),

$$\frac{\partial S_i\{\widehat{\mathbf{A}};\mathbf{b}^*(\rho_0),\rho_0\}}{\partial b_k^*(\rho_0)} = \frac{\partial S_i\{\widehat{\mathbf{A}};\mathbf{b}(\rho_0),\rho_0\}}{\partial b_k(\rho_0)} + O_p\left\{\left|\widehat{b}_k(\rho_0)-b_k(\rho_0)\right|\right\}.$$

Recall that

$$\begin{aligned}
&\widehat{b}_k(\rho_0)-b_k(\rho_0)\\
=\;& \lambda_k^{-1}\langle\phi_k,\widehat{g}-g\rangle+\lambda_k^{-1}\langle\widehat{\phi}_k-\phi_k,g\rangle-\lambda_k^{-2}\left(\widehat{\lambda}_k-\lambda_k\right)\langle\phi_k,g\rangle+O_p(n^{-1}k^{\alpha+1}),
\end{aligned}$$

which, together with Condition (C1) and (C5), Lemma S2.1(c), Lemma S2.3 and Lemma S2.4(c) imply that

$$\begin{aligned}
&\left|\widehat{b}_k(\rho_0)-b_k(\rho_0)\right|\\
\leqslant\;& |\lambda_k^{-1}\langle\phi_k,\widehat{g}-g\rangle|+|\lambda_k^{-1}\langle\widehat{\phi}_k-\phi_k,g\rangle|+|\lambda_k^{-2}\left(\widehat{\lambda}_k-\lambda_k\right)\langle\phi_k,g\rangle|+O_p(n^{-1}k^{\alpha+1})\\
\leqslant\;& \lambda_k^{-1}\|\phi_k\|\|\widehat{g}-g\|+\lambda_k^{-1}\|\widehat{\phi}_k-\phi_k\|\|g\|+\lambda_k^{-2}|\widehat{\lambda}_k-\lambda_k||\langle\phi_k,g\rangle|+O_p(n^{-1}k^{\alpha+1})\\
=\;& O_p(k^{\alpha}n^{-1/2})+O_p(k^{\alpha+1}n^{-1/2})+O_p(k^{\alpha}n^{-1/2})+O_p(n^{-1}k^{\alpha+1})\\
=\;& O_p(k^{\alpha+1}n^{-1/2}), \qquad\qquad (S.33)
\end{aligned}$$

where the last second equality is because the condition $k=o\{n^{1/(2\alpha+2)}\}$ used in Lemma S2.4 naturally holds under Condition (C5). Based on above results and Condition (C5),

$$\begin{aligned}
& n^{-3/4}\sum_{i=1}^{n}\sum_{k=1}^{K}\frac{\partial S_i\{\widehat{\mathbf{A}};\mathbf{b}^*(\rho_0),\rho_0\}}{\partial b_k^*(\rho_0)}\left\{\widehat{b}_k(\rho_0)-b_k(\rho_0)\right\}\\
=\;& n^{-3/4}\sum_{i=1}^{n}\sum_{k=1}^{K}\frac{\partial S_i\{\widehat{\mathbf{A}};\mathbf{b}(\rho_0),\rho_0\}}{\partial b_k(\rho_0)}\left\{\widehat{b}_k(\rho_0)-b_k(\rho_0)\right\}+n^{-3/4}\sum_{i=1}^{n}\sum_{k=1}^{K}O_p\left[\left\{\widehat{b}_k(\rho_0)-b_k(\rho_0)\right\}^2\right]\\
=\;& n^{-3/4}\sum_{i=1}^{n}\sum_{k=1}^{K}\frac{\partial S_i\{\widehat{\mathbf{A}};\mathbf{b}(\rho_0),\rho_0\}}{\partial b_k(\rho_0)}\left\{\widehat{b}_k(\rho_0)-b_k(\rho_0)\right\}+n^{-3/4}\sum_{i=1}^{n}\sum_{k=1}^{K}O_p(k^{2\alpha+2}n^{-1})
\end{aligned}$$

$$= n^{-3/4}\sum_{i=1}^{n}\sum_{k=1}^{K}\frac{\partial S_i\{\widehat{\mathbf{A}};\mathbf{b}(\rho_0),\rho_0\}}{\partial b_k(\rho_0)}\left\{\widehat{b}_k(\rho_0)-b_k(\rho_0)\right\}+o_p(n^{-1/4}). \tag{S.34}$$

Based on (S.27), (S.3) and Condition (C5), we have

$$\begin{aligned}
& n^{-3/4}\sum_{i=1}^{n}\sum_{k=1}^{K}\frac{\partial S_i\{\widehat{\mathbf{A}};\mathbf{b}(\rho_0),\rho_0\}}{\partial b_k(\rho_0)}\left\{\widehat{b}_k(\rho_0)-b_k(\rho_0)\right\}\\
= \; & n^{-7/4}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{K}\frac{\partial S_i\{\widehat{\mathbf{A}};\mathbf{b}(\rho_0),\rho_0\}}{\partial b_k(\rho_0)}\varphi_{jk}+n^{-3/4}\sum_{i=1}^{n}\sum_{k=1}^{K}\frac{\partial S_i\{\mathbf{A};\mathbf{b}(\rho_0),\rho_0\}}{\partial b_k(\rho_0)}O_p(n^{-1}k^{2\alpha+2}\log k)\\
= \; & n^{-7/4}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{K}\frac{\partial S_i\{\widehat{\mathbf{A}};\mathbf{b}(\rho_0),\rho_0\}}{\partial b_k(\rho_0)}\varphi_{jk}+o_p(n^{-1/4})\\
= \; & \sum_{k=1}^{K}\left[n^{-1}\sum_{i=1}^{n}\frac{\partial S_i\{\widehat{\mathbf{A}};\mathbf{b}(\rho_0),\rho_0\}}{\partial b_k(\rho_0)}\right]\left(n^{-3/4}\sum_{j=1}^{n}\varphi_{jk}\right)+o_p(n^{-1/4})\\
= \; & \sum_{k=1}^{K}\left[n^{-1}\sum_{i=1}^{n}\frac{\partial S_i\{\mathbf{A};\mathbf{b}(\rho_0),\rho_0\}}{\partial b_k(\rho_0)}+o_p(n^{-1/4})\right]\left(n^{-3/4}\sum_{j=1}^{n}\varphi_{jk}\right)+o_p(n^{-1/4})\\
= \; & \sum_{k=1}^{K}\left[n^{-1}\sum_{i=1}^{n}\frac{\partial S_i\{\mathbf{A};\mathbf{b}(\rho_0),\rho_0\}}{\partial b_k(\rho_0)}\right]\left(n^{-3/4}\sum_{j=1}^{n}\varphi_{jk}\right)+o_p(n^{-1/4}),
\end{aligned} \tag{S.35}$$

where the second last equality is due to

$$\begin{aligned}
& \sup_{1\leqslant k\leqslant K}\left[\frac{1}{n}\sum_{i=1}^{n}\frac{\partial S_i\{\widehat{\mathbf{A}};\mathbf{b}(\rho_0),\rho_0\}}{\partial b_k(\rho_0)}-\frac{1}{n}\sum_{i=1}^{n}\frac{\partial S_i\{\mathbf{A};\mathbf{b}(\rho_0),\rho_0\}}{\partial b_k(\rho_0)}\right]^2\\
\leqslant \; & \left\|\frac{1}{n}\sum_{i=1}^{n}\frac{\partial S_i\{\widehat{\mathbf{A}};\mathbf{b}(\rho_0),\rho_0\}}{\partial \mathbf{b}(\rho_0)^{\mathrm{T}}}-\frac{1}{n}\sum_{i=1}^{n}\frac{\partial S_i\{\mathbf{A};\mathbf{b}(\rho_0),\rho_0\}}{\partial \mathbf{b}(\rho_0)^{\mathrm{T}}}\right\|^2\\
= \; & o_p(n^{-1/2})
\end{aligned}$$

following (S.3), and the last equality can be established because

$$\sum_{k=1}^{K}\left(n^{-3/4}\sum_{j=1}^{n}\varphi_{jk}\right) = O_p(1). \tag{S.36}$$

The proof of (S.36) is given in the Supplementary Material S3.11. In order to demonstrate the expectation of (S.35) equals to zero asymptotically, with $\widehat{\mathbf{A}}$ in (2.18) replaced by $\mathbf{A}$, we observe

$$
\begin{aligned}
& \frac{1}{n}\sum_{i=1}^{n}\frac{\partial S_i\{\mathbf{A};\mathbf{b}(\rho_0),\rho_0\}}{\partial \mathbf{b}(\rho_0)^{\mathrm{T}}} \\
= & \frac{1}{n}\left[-2\left\{\mathbf{Y}^{\mathrm{T}}\mathbf{M}(\rho_0)-\mathbf{b}(\rho_0)^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{S}(\rho_0)\right\}\left\{\frac{\partial \mathbf{D}(\rho_0)}{\partial\rho}\mathbf{D}(\rho_0)\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{A}\right\}\right. \\
& -\left\{\mathbf{Y}^{\mathrm{T}}\frac{\partial \mathbf{M}(\rho_0)}{\partial\rho}+\mathbf{b}(\rho_0)^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{W}-\frac{\partial \mathbf{b}(\rho_0)^{\mathrm{T}}}{\partial\rho}\mathbf{A}^{\mathrm{T}}\mathbf{S}(\rho_0)\right\}\left\{\mathbf{D}(\rho_0)^2\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{A}\right\} \\
& \left.+\left\{\mathbf{Y}^{\mathrm{T}}\mathbf{M}(\rho_0)-\mathbf{b}(\rho_0)^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{S}(\rho_0)\right\}\left\{\mathbf{D}(\rho_0)^2\mathbf{W}^{\mathrm{T}}\mathbf{A}\right\}\right] \\
= & \frac{1}{n}\left[-2\left\{\mathbf{Y}^{\mathrm{T}}\mathbf{M}(\rho_0)-\mathbf{b}(\rho_0)^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{S}(\rho_0)\right\}\left\{\frac{\partial \mathbf{D}(\rho_0)}{\partial\rho}\mathbf{D}(\rho_0)\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{A}\right\}\right. \\
& -\left\{\mathbf{Y}^{\mathrm{T}}\mathbf{S}(\rho_0)^{\mathrm{T}}\frac{\partial \mathbf{S}(\rho_0)}{\partial\rho}+\mathbf{b}(\rho_0)^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{W}\right\}\left\{\mathbf{D}(\rho_0)^2\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{A}\right\} \\
& -\left\{\mathbf{Y}^{\mathrm{T}}\frac{\partial \mathbf{S}(\rho_0)^{\mathrm{T}}}{\partial\rho}\mathbf{S}(\rho_0)-\frac{\partial \mathbf{b}(\rho_0)^{\mathrm{T}}}{\partial\rho}\mathbf{A}^{\mathrm{T}}\mathbf{S}(\rho_0)\right\}\left\{\mathbf{D}(\rho_0)^2\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{A}\right\} \\
& \left.+\left\{\mathbf{Y}^{\mathrm{T}}\mathbf{M}(\rho_0)-\mathbf{b}(\rho_0)^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{S}(\rho_0)\right\}\left\{\mathbf{D}(\rho_0)^2\mathbf{W}^{\mathrm{T}}\mathbf{A}\right\}\right] \\
= & \frac{1}{n}\left[2\left\{-\mathbf{Y}^{\mathrm{T}}\mathbf{M}(\rho_0)+\mathbf{b}(\rho_0)^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{S}(\rho_0)\right\}\left\{\frac{\partial \mathbf{D}(\rho_0)}{\partial\rho}\mathbf{D}(\rho_0)\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{A}\right\}\right. \\
& +\left\{\mathbf{Y}^{\mathrm{T}}\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{W}-\mathbf{b}(\rho_0)^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{W}\right\}\left\{\mathbf{D}(\rho_0)^2\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{A}\right\} \\
& +\left\{\mathbf{Y}^{\mathrm{T}}\mathbf{M}(\rho_0)-\mathbf{b}(\rho_0)^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{S}(\rho_0)\right\}\left\{\mathbf{D}(\rho_0)^2\mathbf{W}^{\mathrm{T}}\mathbf{A}\right\} \\
& \left.+\left\{\mathbf{Y}^{\mathrm{T}}\mathbf{W}^{\mathrm{T}}\mathbf{S}(\rho_0)+\frac{\partial \mathbf{b}(\rho_0)^{\mathrm{T}}}{\partial\rho}\mathbf{A}^{\mathrm{T}}\mathbf{S}(\rho_0)\right\}\left\{\mathbf{D}(\rho_0)^2\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{A}\right\}\right] \\
\equiv & \boldsymbol{\tau}_1^{\mathrm{T}}+\boldsymbol{\tau}_2^{\mathrm{T}}+\boldsymbol{\tau}_3^{\mathrm{T}}+\boldsymbol{\tau}_4^{\mathrm{T}}, \qquad \text{(S.37)}
\end{aligned}
$$

where $\boldsymbol{\tau}_\ell$'s are $K\times 1$ vectors with the kth element denoted by $\tau_{\ell k}$ for $\ell=1,2,3,4$, and $E(\boldsymbol{\tau}_\ell \mid \mathbf{A})=0$ for $\ell=1,2,3$. We now address first three terms of (S.37). For $\boldsymbol{\tau}_1$, note that

$$
\begin{aligned}
& \mathrm{var}(\tau_{1k}\mid \mathbf{A}) \\
= & E(\tau_{1k}^2\mid \mathbf{A})
\end{aligned}
$$

$$
\begin{aligned}
&\leqslant E(\|\boldsymbol{\tau}_1\|^2 \mid \mathbf{A})\\
&= E\left[4\frac{1}{n^2}\left\|\left\{-\mathbf{Y}^{\mathrm{T}}\mathbf{M}(\rho_0)+\mathbf{b}(\rho_0)^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{S}(\rho_0)\right\}\left\{\frac{\partial \mathbf{D}(\rho_0)}{\partial\rho}\mathbf{D}(\rho_0)\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{A}\right\}\right\|^2 \mid \mathbf{A}\right]\\
&= E\left[4\frac{1}{n^2}\left\{-\mathbf{Y}^{\mathrm{T}}\mathbf{M}(\rho_0)+\mathbf{b}(\rho_0)^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{S}(\rho_0)\right\}\left\{\frac{\partial \mathbf{D}(\rho_0)}{\partial\rho}\mathbf{D}(\rho_0)\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{A}\right\}\right.\\
&\qquad\left.\times\left\{\frac{\partial \mathbf{D}(\rho_0)}{\partial\rho}\mathbf{D}(\rho_0)\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{A}\right\}^{\mathrm{T}}\left\{-\mathbf{Y}^{\mathrm{T}}\mathbf{M}(\rho_0)+\mathbf{b}(\rho_0)^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{S}(\rho_0)\right\}^{\mathrm{T}} \mid \mathbf{A}\right]\\
&= E\left[4\frac{1}{n^2}\left\{-\mathbf{Y}^{\mathrm{T}}\mathbf{M}(\rho_0)+\mathbf{b}(\rho_0)^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{S}(\rho_0)\right\}\left\{\frac{\partial \mathbf{D}(\rho_0)}{\partial\rho}\mathbf{D}(\rho_0)\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{A}\right\}\right.\\
&\qquad\left.\times\left\{\mathbf{A}^{\mathrm{T}}\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)\frac{\partial \mathbf{D}(\rho_0)}{\partial\rho}\right\}\left\{-\mathbf{M}(\rho_0)\mathbf{Y}+\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{A}\mathbf{b}(\rho_0)\right\} \mid \mathbf{A}\right]\\
&= E\left[4\frac{1}{n^2}\left\{-\mathbf{Y}^{\mathrm{T}}\mathbf{S}(\rho_0)^{\mathrm{T}}+\mathbf{b}(\rho_0)^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\right\}\left\{\mathbf{S}(\rho_0)\frac{\partial \mathbf{D}(\rho_0)}{\partial\rho}\mathbf{D}(\rho_0)\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{A}\right\}\right.\\
&\qquad\left.\times\left\{\mathbf{A}^{\mathrm{T}}\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)\frac{\partial \mathbf{D}(\rho_0)}{\partial\rho}\mathbf{S}(\rho_0)^{\mathrm{T}}\right\}\left\{-\mathbf{S}(\rho_0)\mathbf{Y}+\mathbf{A}\mathbf{b}(\rho_0)\right\} \mid \mathbf{A}\right]\\
&= 4\frac{1}{n^2}E\left(\boldsymbol{\varepsilon}_n^{\mathrm{T}}\mathscr{L}_n\boldsymbol{\varepsilon}_n \mid \mathbf{A}\right),
\end{aligned}
$$

where

$$
\boldsymbol{\varepsilon}_n^{\mathrm{T}} = -\mathbf{Y}^{\mathrm{T}}\mathbf{S}(\rho_0)^{\mathrm{T}}+\mathbf{b}(\rho_0)^{\mathrm{T}}\mathbf{A}^{\mathrm{T}},
$$

and

$$
\mathscr{L}_n = \left\{\mathbf{S}(\rho_0)\frac{\partial \mathbf{D}(\rho_0)}{\partial\rho}\mathbf{D}(\rho_0)\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{A}\right\}\left\{\mathbf{A}^{\mathrm{T}}\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)\frac{\partial \mathbf{D}(\rho_0)}{\partial\rho}\mathbf{S}(\rho_0)^{\mathrm{T}}\right\}.
$$

We next check the order of $E(\boldsymbol{\varepsilon}_n^{\mathrm{T}}\mathscr{L}_n\boldsymbol{\varepsilon}_n \mid \mathbf{A})$. Together with $\boldsymbol{\varepsilon}_n \mid \mathbf{A} \sim \mathcal{N}(0,\sigma^2\mathbf{I}_n)$ and Lemma S2.7, we note that

$$
E(\boldsymbol{\varepsilon}_n^{\mathrm{T}}\mathscr{L}_n\boldsymbol{\varepsilon}_n \mid \mathbf{A}) = \sigma^2\mathrm{trace}\left(\mathscr{L}_n \mid \mathbf{A}\right),
$$

where $\mathrm{trace}(\mathscr{L}_n \mid \mathbf{A})$ denotes the calculation of $\mathrm{trace}(\mathscr{L}_n)$ for fixed $\mathbf{A}$. Below, we drop $\mathbf{A}$ in

$\mathrm{trace}(\mathscr{L}_n \mid \mathbf{A})$ for notational simplicity. Hence, we only need to check the order of $\mathrm{trace}(\mathscr{L}_n)$ for fixed $\mathbf{A}$. Since $\mathbf{S} = \mathbf{I} - \rho\mathbf{W}$, we have

$$\begin{aligned}
& \mathrm{trace}(\mathscr{L}_n) \\
= \; & \mathrm{trace}\left\{(\mathbf{I} - \rho_0\mathbf{W})\frac{\partial \mathbf{D}(\rho_0)}{\partial \rho}\mathbf{D}(\rho_0)(\mathbf{I} - \rho_0\mathbf{W}^{\mathrm{T}})\mathbf{A}\mathbf{A}^{\mathrm{T}}(\mathbf{I} - \rho_0\mathbf{W})\mathbf{D}(\rho_0)\frac{\partial \mathbf{D}(\rho_0)}{\partial \rho}(\mathbf{I} - \rho_0\mathbf{W}^{\mathrm{T}})\right\} \\
= \; & \mathrm{trace}(\mathbf{L}_{n1}) + \mathrm{trace}(\mathbf{L}_{n2}) + \mathrm{trace}(\mathbf{L}_{n3}) + \mathrm{trace}(\mathbf{L}_{n4}) + \mathrm{trace}(\mathbf{L}_{n5}),
\end{aligned}$$

where

$$\mathbf{L}_{n1} = \left\{\frac{\partial \mathbf{D}(\rho_0)}{\partial \rho}\mathbf{D}(\rho_0)\mathbf{A}\mathbf{A}^{\mathrm{T}}\mathbf{D}(\rho_0)\frac{\partial \mathbf{D}(\rho_0)}{\partial \rho}\right\},$$

$$\begin{aligned}
& \mathbf{L}_{n2} \\
= \; & -\rho_0\left\{\frac{\partial \mathbf{D}(\rho_0)}{\partial \rho}\mathbf{D}(\rho_0)\mathbf{A}\mathbf{A}^{\mathrm{T}}\mathbf{D}(\rho_0)\frac{\partial \mathbf{D}(\rho_0)}{\partial \rho}\mathbf{W}^{\mathrm{T}}\right\} - \rho_0\left\{\frac{\partial \mathbf{D}(\rho_0)}{\partial \rho}\mathbf{D}(\rho_0)\mathbf{A}\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{D}(\rho_0)\frac{\partial \mathbf{D}(\rho_0)}{\partial \rho}\right\} \\
& -\rho_0\left\{\frac{\partial \mathbf{D}(\rho_0)}{\partial \rho}\mathbf{D}(\rho_0)\mathbf{W}^{\mathrm{T}}\mathbf{A}\mathbf{A}^{\mathrm{T}}\mathbf{D}(\rho_0)\frac{\partial \mathbf{D}(\rho_0)}{\partial \rho}\right\} - \rho_0\left\{\mathbf{W}\frac{\partial \mathbf{D}(\rho_0)}{\partial \rho}\mathbf{D}(\rho_0)\mathbf{A}\mathbf{A}^{\mathrm{T}}\mathbf{D}(\rho_0)\frac{\partial \mathbf{D}(\rho_0)}{\partial \rho}\right\},
\end{aligned}$$

$$\begin{aligned}
\mathbf{L}_{n3} = \; & \rho_0^2\left\{(\mathbf{W})\frac{\partial \mathbf{D}(\rho_0)}{\partial \rho}\mathbf{D}(\rho_0)(\mathbf{W}^{\mathrm{T}})\mathbf{A}\mathbf{A}^{\mathrm{T}}\mathbf{D}(\rho_0)\frac{\partial \mathbf{D}(\rho_0)}{\partial \rho}\right\} \\
& +\rho_0^2\left\{(\mathbf{W})\frac{\partial \mathbf{D}(\rho_0)}{\partial \rho}\mathbf{D}(\rho_0)\mathbf{A}\mathbf{A}^{\mathrm{T}}(\mathbf{W})\mathbf{D}(\rho_0)\frac{\partial \mathbf{D}(\rho_0)}{\partial \rho}\right\} \\
& +\rho_0^2\left\{(\mathbf{W})\frac{\partial \mathbf{D}(\rho_0)}{\partial \rho}\mathbf{D}(\rho_0)\mathbf{A}\mathbf{A}^{\mathrm{T}}\mathbf{D}(\rho_0)\frac{\partial \mathbf{D}(\rho_0)}{\partial \rho}(\mathbf{W}^{\mathrm{T}})\right\} \\
& +\rho_0^2\left\{\frac{\partial \mathbf{D}(\rho_0)}{\partial \rho}\mathbf{D}(\rho_0)(\mathbf{W}^{\mathrm{T}})\mathbf{A}\mathbf{A}^{\mathrm{T}}(\mathbf{W})\mathbf{D}(\rho_0)\frac{\partial \mathbf{D}(\rho_0)}{\partial \rho}\right\} \\
& +\rho_0^2\left\{\frac{\partial \mathbf{D}(\rho_0)}{\partial \rho}\mathbf{D}(\rho_0)(\mathbf{W}^{\mathrm{T}})\mathbf{A}\mathbf{A}^{\mathrm{T}}\mathbf{D}(\rho_0)\frac{\partial \mathbf{D}(\rho_0)}{\partial \rho}(\mathbf{W}^{\mathrm{T}})\right\} \\
& +\rho_0^2\left\{\frac{\partial \mathbf{D}(\rho_0)}{\partial \rho}\mathbf{D}(\rho_0)\mathbf{A}\mathbf{A}^{\mathrm{T}}(\mathbf{W})\mathbf{D}(\rho_0)\frac{\partial \mathbf{D}(\rho_0)}{\partial \rho}(\mathbf{W}^{\mathrm{T}})\right\},
\end{aligned}$$

$$\begin{aligned}
\mathbf{L}_{n4} &= -\rho_0^3\left\{(\mathbf{W})\frac{\partial \mathbf{D}(\rho_0)}{\partial \rho}\mathbf{D}(\rho_0)(\mathbf{W}^{\mathrm{T}})\mathbf{A}\mathbf{A}^{\mathrm{T}}(\mathbf{W})\mathbf{D}(\rho_0)\frac{\partial \mathbf{D}(\rho_0)}{\partial \rho}\right\} \\
&\quad -\rho_0^3\left\{(\mathbf{W})\frac{\partial \mathbf{D}(\rho_0)}{\partial \rho}\mathbf{D}(\rho_0)(\mathbf{W}^{\mathrm{T}})\mathbf{A}\mathbf{A}^{\mathrm{T}}\mathbf{D}(\rho_0)\frac{\partial \mathbf{D}(\rho_0)}{\partial \rho}(\mathbf{W}^{\mathrm{T}})\right\} \\
&\quad -\rho_0^3\left\{(\mathbf{W})\frac{\partial \mathbf{D}(\rho_0)}{\partial \rho}\mathbf{D}(\rho_0)\mathbf{A}\mathbf{A}^{\mathrm{T}}(\mathbf{W})\mathbf{D}(\rho_0)\frac{\partial \mathbf{D}(\rho_0)}{\partial \rho}(\mathbf{W}^{\mathrm{T}})\right\} \\
&\quad -\rho_0^3\left\{\frac{\partial \mathbf{D}(\rho_0)}{\partial \rho}\mathbf{D}(\rho_0)(\mathbf{W}^{\mathrm{T}})\mathbf{A}\mathbf{A}^{\mathrm{T}}(\mathbf{W})\mathbf{D}(\rho_0)\frac{\partial \mathbf{D}(\rho_0)}{\partial \rho}(\mathbf{W}^{\mathrm{T}})\right\},
\end{aligned}$$

and

$$\mathbf{L}_{n5} = \rho^4\left\{\mathbf{W}\frac{\partial \mathbf{D}(\rho_0)}{\partial \rho}\mathbf{D}(\rho_0)\mathbf{W}^{\mathrm{T}}\mathbf{A}\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{D}(\rho_0)\frac{\partial \mathbf{D}(\rho_0)}{\partial \rho}\mathbf{W}^{\mathrm{T}}\right\}.$$

Denote the (i,i)th element of the diagonal matrix $\frac{\partial \mathbf{D}(\rho)}{\partial \rho}\mathbf{D}(\rho)$ by $\overline{\mathscr{D}}_{ii}(\rho)$ for $i = 1,\dots,n$. Through Remark 2, we know that $\overline{\mathscr{D}}_{ii}(\rho) = O(1)$. Together with $\sum_{k=1}^K \xi_{ik}^2 > 0$ and $E(\sum_{k=1}^K \xi_{ik}^2) = \sum_{k=1}^K \lambda_k = O(1)$ by the Condition (C2), we have

$$\sum_{k=1}^K \xi_{ik}^2 = O_p(1), \tag{S.38}$$

and from which we obtain

$$\begin{aligned}
&\mathrm{trace}(\mathbf{L}_{n1}) \\
&= \mathrm{trace}\left\{\frac{\partial \mathbf{D}(\rho_0)}{\partial \rho}\mathbf{D}(\rho_0)\mathbf{A}\mathbf{A}^{\mathrm{T}}\mathbf{D}(\rho_0)\frac{\partial \mathbf{D}(\rho_0)}{\partial \rho}\right\} \\
&= \mathrm{trace}\begin{pmatrix}
\overline{\mathscr{D}}_{11}(\rho_0)^2\sum_{k=1}^K \xi_{1k}^2 & \cdots & \overline{\mathscr{D}}_{11}(\rho_0)\overline{\mathscr{D}}_{nn}(\rho_0)\sum_{k=1}^K \xi_{1k}\xi_{nk} \\
\overline{\mathscr{D}}_{22}(\rho_0)\overline{\mathscr{D}}_{11}(\rho_0)\sum_{k=1}^K \xi_{2k}\xi_{1k} & \cdots & \overline{\mathscr{D}}_{22}(\rho_0)\overline{\mathscr{D}}_{nn}(\rho_0)\sum_{k=1}^K \xi_{2k}\xi_{nk} \\
 & & \\
\overline{\mathscr{D}}_{nn}(\rho_0)\overline{\mathscr{D}}_{11}(\rho_0)\sum_{k=1}^K \xi_{nk}\xi_{1k} & \cdots & \overline{\mathscr{D}}_{nn}(\rho_0)^2\sum_{k=1}^K \xi_{nk}^2
\end{pmatrix}
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{i=1}^{n} \overline{\mathscr{D}}_{ii}(\rho_0)^2 \sum_{k=1}^{K} \xi_{ik}^2 \\
&= \sum_{i=1}^{n} \overline{\mathscr{D}}_{ii}(\rho_0)^2 O_p(1) \\
&= O_p(n).
\end{aligned}
$$

We now investigate the first term of $\mathbf{L}_{n2}$. For $i \neq j$,

$$
\begin{aligned}
& E\left(\sum_{k=1}^{K} \xi_{jk}\xi_{ik}\right)^2 \\
= & \sum_{k=1}^{K} E\left(\xi_{jk}^2 \xi_{ik}^2\right) + \sum_{k_1 \neq k_2}^{K} E\left(\xi_{jk_1}\xi_{jk_2}\xi_{ik_1}\xi_{ik_2}\right) \\
= & \sum_{k=1}^{K} E\left(\xi_{jk}^2\right) E\left(\xi_{ik}^2\right) + \sum_{k_1 \neq k_2}^{K} E\left(\xi_{jk_1}\xi_{jk_2}\right) E\left(\xi_{ik_1}\xi_{ik_2}\right) \\
= & \sum_{k=1}^{K} \lambda_k^2 \\
= & O(1),
\end{aligned}
$$

where the last equality is due to the Condition (C2). Thus, for any $1 \leqslant i \neq j \leqslant n$, we get

$$
\sum_{k=1}^{K} \xi_{jk}\xi_{ik} = O_p(1). \tag{S.39}
$$

Combining (S.38) and (S.39), we have

$$
\operatorname{trace}\left[-\rho_0 \left\{ \frac{\partial \mathbf{D}(\rho_0)}{\partial \rho} \mathbf{D}(\rho_0) \mathbf{A}\mathbf{A}^{\mathrm{T}} \mathbf{D}(\rho_0) \frac{\partial \mathbf{D}(\rho_0)}{\partial \rho} \mathbf{W}^{\mathrm{T}} \right\}\right]
$$

$$
\begin{aligned}
&= -\rho_0 \times \text{trace}\begin{pmatrix}
\overline{\mathscr{D}}_{11}(\rho_0)\sum_{i=1}^n\sum_{k=1}^K \xi_{1k}\xi_{ik}w_{1i}\overline{\mathscr{D}}_{ii}(\rho_0) & \cdots & \overline{\mathscr{D}}_{11}(\rho_0)\sum_{i=1}^n\sum_{k=1}^K \xi_{1k}\xi_{ik}w_{ni}\overline{\mathscr{D}}_{ii}(\rho_0) \\
\overline{\mathscr{D}}_{22}(\rho_0)\sum_{i=1}^n\sum_{k=1}^K \xi_{2k}\xi_{ik}w_{1i}\overline{\mathscr{D}}_{ii}(\rho_0) & \cdots & \overline{\mathscr{D}}_{22}(\rho_0)\sum_{i=1}^n\sum_{k=1}^K \xi_{2k}\xi_{ik}w_{ni}\overline{\mathscr{D}}_{ii}(\rho_0) \\
 & & \\
\overline{\mathscr{D}}_{nn}(\rho_0)\sum_{i=1}^n\sum_{k=1}^K \xi_{nk}\xi_{ik}w_{1i}\overline{\mathscr{D}}_{ii}(\rho_0) & \cdots & \overline{\mathscr{D}}_{nn}(\rho_0)\sum_{i=1}^n\sum_{k=1}^K \xi_{nk}\xi_{ik}w_{ni}\overline{\mathscr{D}}_{ii}(\rho_0)
\end{pmatrix} \\
&= -\rho_0\sum_{j=1}^n \overline{\mathscr{D}}_{jj}(\rho_0)\sum_{i=1}^n\sum_{k=1}^K \xi_{jk}\xi_{ik}w_{ji}\overline{\mathscr{D}}_{ii}(\rho_0) \\
&= -\rho_0 O_p(1)\sum_{j=1}^n\sum_{i=1}^n w_{ji} \\
&= O_p(n),
\end{aligned}
$$

where the last equality is because $\sum_{i=1}^n w_{ji} = 1$. The same order can be similarly obtained for the other three terms of $\mathbf{L}_{n2}$ and the proof are omitted. For the first term of $\mathbf{L}_{n3}$, we have

$$
\begin{aligned}
&\rho_0^2\left\{(\mathbf{W})\frac{\partial\mathbf{D}(\rho_0)}{\partial\rho}\mathbf{D}(\rho_0)(\mathbf{W}^{\mathrm{T}})\mathbf{A}\mathbf{A}^{\mathrm{T}}\mathbf{D}(\rho_0)\frac{\partial\mathbf{D}(\rho_0)}{\partial\rho}\right\} \\
= \;&\rho_0^2\begin{pmatrix}
\sum_{i=1}^n w_{1i}^2\overline{\mathscr{D}}_{ii}(\rho_0) & \sum_{i=1}^n w_{1i}w_{2i}\overline{\mathscr{D}}_{ii}(\rho_0) & \cdots & \sum_{i=1}^n w_{1i}w_{ni}\overline{\mathscr{D}}_{ii}(\rho_0) \\
\sum_{i=1}^n w_{2i}w_{1i}\overline{\mathscr{D}}_{ii}(\rho_0) & \sum_{i=1}^n w_{2i}^2\overline{\mathscr{D}}_{ii}(\rho_0) & \cdots & \sum_{i=1}^n w_{2i}w_{ni}\overline{\mathscr{D}}_{ii}(\rho_0) \\
\vdots & \vdots & \ddots & \vdots \\
\sum_{i=1}^n w_{ni}w_{1i}\overline{\mathscr{D}}_{ii}(\rho_0) & \sum_{i=1}^n w_{ni}w_{2i}\overline{\mathscr{D}}_{ii}(\rho_0) & \cdots & \sum_{i=1}^n w_{ni}^2\overline{\mathscr{D}}_{ii}(\rho_0)
\end{pmatrix} \\
&\times\begin{pmatrix}
\sum_{k=1}^K \xi_{1k}^2\overline{\mathscr{D}}_{11}(\rho_0) & \sum_{k=1}^K \xi_{1k}\xi_{2k}\overline{\mathscr{D}}_{22}(\rho_0) & \cdots & \sum_{k=1}^K \xi_{1k}\xi_{nk}\overline{\mathscr{D}}_{nn}(\rho_0) \\
\sum_{k=1}^K \xi_{2k}\xi_{1k}\overline{\mathscr{D}}_{11}(\rho_0) & \sum_{k=1}^K \xi_{2k}^2\overline{\mathscr{D}}_{22}(\rho_0) & \cdots & \sum_{k=1}^K \xi_{2k}\xi_{nk}\overline{\mathscr{D}}_{nn}(\rho_0) \\
\vdots & \vdots & \ddots & \vdots \\
\sum_{k=1}^K \xi_{nk}\xi_{1k}\overline{\mathscr{D}}_{11}(\rho_0) & \sum_{k=1}^K \xi_{nk}\xi_{2k}\overline{\mathscr{D}}_{22}(\rho_0) & \cdots & \sum_{k=1}^K \xi_{nk}^2\overline{\mathscr{D}}_{nn}(\rho_0)
\end{pmatrix} \\
= \;&\rho_0^2\left[\sum_{j=1}^n\left\{\sum_{i=1}^n w_{\ell_1 i}w_{ji}\overline{\mathscr{D}}_{ii}(\rho_0)\right\}\left\{\sum_{k=1}^K \xi_{jk}\xi_{\ell_2 k}\overline{\mathscr{D}}_{\ell_2\ell_2}(\rho_0)\right\}\right]_{1\leqslant\ell_1,\ell_2\leqslant n}.
\end{aligned}
$$

Thus, by (S.39) and (S.24),

$$
\begin{aligned}
& \operatorname{trace}\left[\rho_0^2\left\{(\mathbf{W})\frac{\partial \mathbf{D}(\rho_0)}{\partial \rho}\mathbf{D}(\rho_0)(\mathbf{W}^{\mathrm{T}})\mathbf{A}\mathbf{A}^{\mathrm{T}}\mathbf{D}(\rho_0)\frac{\partial \mathbf{D}(\rho_0)}{\partial \rho}\right\}\right] \\
= \ & \rho_0^2\sum_{\ell=1}^{n}\sum_{j=1}^{n}\left\{\sum_{i=1}^{n} w_{\ell i}w_{ji}\overline{\mathscr{D}}_{ii}(\rho_0)\right\}\left\{\sum_{k=1}^{K}\xi_{jk}\xi_{\ell k}\overline{\mathscr{D}}_{\ell\ell}(\rho_0)\right\} \\
= \ & \rho_0^2\sum_{\ell=1}^{n}\sum_{j=1}^{n}\left\{\sum_{i=1}^{n} w_{\ell i}w_{ji}\overline{\mathscr{D}}_{ii}(\rho_0)\right\} O_p(1) \\
= \ & \rho_0^2\sum_{\ell=1}^{n}\sum_{j=1}^{n} o_p(n^{-1/2}) \\
= \ & o_p(n^{3/2}).
\end{aligned}
$$

Using the similar techniques, we can show that the order of other five terms of $\mathbf{L}_{n3}$ are also $o_p(n^{3/2})$. For $\mathbf{L}_{n4}$, we also focus on the first term since the other three terms are of the same order and can be dealt with similarly. Note that

$$
\begin{aligned}
& -\rho_0^3\left\{(\mathbf{W})\frac{\partial \mathbf{D}(\rho_0)}{\partial \rho}\mathbf{D}(\rho_0)(\mathbf{W}^{\mathrm{T}})\mathbf{A}\mathbf{A}^{\mathrm{T}}(\mathbf{W})\mathbf{D}(\rho_0)\frac{\partial \mathbf{D}(\rho_0)}{\partial \rho}\right\} \\
= \ & -\rho_0^3\begin{pmatrix}
\sum_{i=1}^{n} w_{1i}^2\overline{\mathscr{D}}_{ii}(\rho_0) & \sum_{i=1}^{n} w_{1i}w_{2i}\overline{\mathscr{D}}_{ii}(\rho_0) & \cdots & \sum_{i=1}^{n} w_{1i}w_{ni}\overline{\mathscr{D}}_{ii}(\rho_0) \\
\sum_{i=1}^{n} w_{2i}w_{1i}\overline{\mathscr{D}}_{ii}(\rho_0) & \sum_{i=1}^{n} w_{2i}^2\overline{\mathscr{D}}_{ii}(\rho_0) & \cdots & \sum_{i=1}^{n} w_{2i}w_{ni}\overline{\mathscr{D}}_{ii}(\rho_0) \\
\vdots & \vdots & \ddots & \vdots \\
\sum_{i=1}^{n} w_{ni}w_{1i}\overline{\mathscr{D}}_{ii}(\rho_0) & \sum_{i=1}^{n} w_{ni}w_{2i}\overline{\mathscr{D}}_{ii}(\rho_0) & \cdots & \sum_{i=1}^{n} w_{ni}^2\overline{\mathscr{D}}_{ii}(\rho_0)
\end{pmatrix} \\
& \times\begin{pmatrix}
\sum_{j=1}^{n}(\sum_{k=1}^{K}\xi_{1k}\xi_{jk})w_{j1}\overline{\mathscr{D}}_{11}(\rho_0) & \cdots & \sum_{j=1}^{n}(\sum_{k=1}^{K}\xi_{1k}\xi_{jk})w_{jn}\overline{\mathscr{D}}_{nn}(\rho_0) \\
\sum_{j=1}^{n}(\sum_{k=1}^{K}\xi_{2k}\xi_{jk})w_{j1}\overline{\mathscr{D}}_{11}(\rho_0) & \cdots & \sum_{j=1}^{n}(\sum_{k=1}^{K}\xi_{2k}\xi_{jk})w_{jn}\overline{\mathscr{D}}_{nn}(\rho_0) \\
\vdots & \ddots & \vdots \\
\sum_{j=1}^{n}(\sum_{k=1}^{K}\xi_{nk}\xi_{jk})w_{j1}\overline{\mathscr{D}}_{11}(\rho_0) & \cdots & \sum_{j=1}^{n}(\sum_{k=1}^{K}\xi_{nk}\xi_{jk})w_{jn}\overline{\mathscr{D}}_{nn}(\rho_0)
\end{pmatrix} \\
= \ & -\rho_0^3\left(\sum_{\ell=1}^{n}\sum_{i=1}^{n} w_{t_1 i}w_{\ell i}\overline{\mathscr{D}}_{ii}(\rho_0)\sum_{j=1}^{n}\sum_{k=1}^{K}\xi_{\ell k}\xi_{jk}w_{jt_2}\overline{\mathscr{D}}_{t_2t_2}(\rho_0)\right)_{1\leqslant t_1,t_2\leqslant n}.
\end{aligned}
$$

This implies that

$$\begin{aligned}
& \operatorname{trace}\left[-\rho_0^3\left\{(\mathbf{W})\frac{\partial \mathbf{D}(\rho_0)}{\partial \rho}\mathbf{D}(\rho_0)(\mathbf{W}^{\mathrm{T}})\mathbf{A}\mathbf{A}^{\mathrm{T}}(\mathbf{W})\mathbf{D}(\rho_0)\frac{\partial \mathbf{D}(\rho_0)}{\partial \rho}\right\}\right] \\
= & -\rho_0^3\sum_{t=1}^{n}\sum_{\ell=1}^{n}\sum_{i=1}^{n} w_{ti}w_{\ell i}\overline{\mathscr{D}}_{ii}(\rho_0)\sum_{j=1}^{n}\sum_{k=1}^{K}\xi_{\ell k}\xi_{jk}w_{jt}\overline{\mathscr{D}}_{tt}(\rho_0) \\
= & -\rho_0^3 O_p(1)\times\sum_{t=1}^{n}\sum_{\ell=1}^{n}\sum_{j=1}^{n}\sum_{i=1}^{n} w_{ti}w_{\ell i}w_{jt} \\
= & -\rho_0^3 O_p(1)o(n^{-1/2})\times\sum_{\ell=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{n} w_{jt} \\
= & -\rho_0^3 O_p(1)o(n^{-1/2})\times n^2 \\
= & \ o_p(n^{3/2}).
\end{aligned}$$

For $\mathbf{L}_{n5}$, we have

$$\begin{aligned}
& \mathbf{L}_{n5} \\
= & \ \rho_0^4\left\{\mathbf{W}\frac{\partial \mathbf{D}(\rho_0)}{\partial \rho}\mathbf{D}(\rho_0)\mathbf{W}^{\mathrm{T}}\mathbf{A}\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{D}(\rho_0)\frac{\partial \mathbf{D}(\rho_0)}{\partial \rho}\mathbf{W}^{\mathrm{T}}\right\} \\
= & \ \rho_0^4\begin{pmatrix}
\sum_{i=1}^n w_{1i}^2\overline{\mathscr{D}}_{ii}(\rho_0) & \sum_{i=1}^n w_{1i}w_{2i}\overline{\mathscr{D}}_{ii}(\rho_0) & \cdots & \sum_{i=1}^n w_{1i}w_{ni}\overline{\mathscr{D}}_{ii}(\rho_0) \\
\sum_{i=1}^n w_{2i}w_{1i}\overline{\mathscr{D}}_{ii}(\rho_0) & \sum_{i=1}^n w_{2i}^2\overline{\mathscr{D}}_{ii}(\rho_0) & \cdots & \sum_{i=1}^n w_{2i}w_{ni}\overline{\mathscr{D}}_{ii}(\rho_0) \\
\vdots & \vdots & \ddots & \vdots \\
\sum_{i=1}^n w_{ni}w_{1i}\overline{\mathscr{D}}_{ii}(\rho_0) & \sum_{i=1}^n w_{ni}w_{2i}\overline{\mathscr{D}}_{ii}(\rho_0) & \cdots & \sum_{i=1}^n w_{ni}^2\overline{\mathscr{D}}_{ii}(\rho_0)
\end{pmatrix} \\
& \times\begin{pmatrix}
\sum_{k=1}^K \xi_{1k}^2 & \sum_{k=1}^K \xi_{1k}\xi_{2k} & \cdots & \sum_{k=1}^K \xi_{1k}\xi_{nk} \\
\sum_{k=1}^K \xi_{2k}\xi_{1k} & \sum_{k=1}^K \xi_{2k}^2 & \cdots & \sum_{k=1}^K \xi_{2k}\xi_{nk} \\
 & & & \\
\sum_{k=1}^K \xi_{nk}\xi_{1k} & \sum_{k=1}^K \xi_{nk}\xi_{2k} & \cdots & \sum_{k=1}^K \xi_{nk}^2
\end{pmatrix}
\end{aligned}$$

$$
\times\begin{pmatrix}
\sum_{i=1}^n w_{1i}^2\overline{\mathscr{D}}_{ii}(\rho_0) & \sum_{i=1}^n w_{1i}w_{2i}\overline{\mathscr{D}}_{ii}(\rho_0) & \cdots & \sum_{i=1}^n w_{1i}w_{ni}\overline{\mathscr{D}}_{ii}(\rho_0) \\
\sum_{i=1}^n w_{2i}w_{1i}\overline{\mathscr{D}}_{ii}(\rho_0) & \sum_{i=1}^n w_{2i}^2\overline{\mathscr{D}}_{ii}(\rho_0) & \cdots & \sum_{i=1}^n w_{2i}w_{ni}\overline{\mathscr{D}}_{ii}(\rho_0) \\
\vdots & \vdots & \ddots & \vdots \\
\sum_{i=1}^n w_{ni}w_{1i}\overline{\mathscr{D}}_{ii}(\rho_0) & \sum_{i=1}^n w_{ni}w_{2i}\overline{\mathscr{D}}_{ii}(\rho_0) & \cdots & \sum_{i=1}^n w_{ni}^2\overline{\mathscr{D}}_{ii}(\rho_0)
\end{pmatrix}
$$

$$
= \rho_0^4\begin{pmatrix}
\sum_{i=1}^n w_{1i}^2\overline{\mathscr{D}}_{ii}(\rho_0) & \sum_{i=1}^n w_{1i}w_{2i}\overline{\mathscr{D}}_{ii}(\rho_0) & \cdots & \sum_{i=1}^n w_{1i}w_{ni}\overline{\mathscr{D}}_{ii}(\rho_0) \\
\sum_{i=1}^n w_{2i}w_{1i}\overline{\mathscr{D}}_{ii}(\rho_0) & \sum_{i=1}^n w_{2i}^2\overline{\mathscr{D}}_{ii}(\rho_0) & \cdots & \sum_{i=1}^n w_{2i}w_{ni}\overline{\mathscr{D}}_{ii}(\rho_0) \\
\vdots & \vdots & \ddots & \vdots \\
\sum_{i=1}^n w_{ni}w_{1i}\overline{\mathscr{D}}_{ii}(\rho_0) & \sum_{i=1}^n w_{ni}w_{2i}\overline{\mathscr{D}}_{ii}(\rho_0) & \cdots & \sum_{i=1}^n w_{ni}^2\overline{\mathscr{D}}_{ii}(\rho_0)
\end{pmatrix}
$$

$$
\times\begin{pmatrix}
\sum_{j=1}^n\sum_{k=1}^K \xi_{1k}\xi_{jk}\sum_{i=1}^n w_{ji}w_{1i}\overline{\mathscr{D}}_{ii}(\rho_0) & \cdots & \sum_{j=1}^n\sum_{k=1}^K \xi_{1k}\xi_{jk}\sum_{i=1}^n w_{ji}w_{ni}\overline{\mathscr{D}}_{ii}(\rho_0) \\
\sum_{j=1}^n\sum_{k=1}^K \xi_{2k}\xi_{jk}\sum_{i=1}^n w_{ji}w_{1i}\overline{\mathscr{D}}_{ii}(\rho_0) & \cdots & \sum_{j=1}^n\sum_{k=1}^K \xi_{2k}\xi_{jk}\sum_{i=1}^n w_{ji}w_{ni}\overline{\mathscr{D}}_{ii}(\rho_0) \\
\vdots & \ddots & \vdots \\
\sum_{j=1}^n\sum_{k=1}^K \xi_{nk}\xi_{jk}\sum_{i=1}^n w_{ji}w_{1i}\overline{\mathscr{D}}_{ii}(\rho_0) & \cdots & \sum_{j=1}^n\sum_{k=1}^K \xi_{nk}\xi_{jk}\sum_{i=1}^n w_{ji}w_{ni}\overline{\mathscr{D}}_{ii}(\rho_0)
\end{pmatrix}
$$

$$
= \rho_0^4\left(\sum_{\ell=1}^n\sum_{i_1=1}^n w_{t_1i_1}w_{\ell i_1}\overline{\mathscr{D}}_{i_1i_1}(\rho_0)\sum_{j=1}^n\sum_{k=1}^K \xi_{\ell k}\xi_{jk}\sum_{i_2=1}^n w_{ji_2}w_{t_2i_2}\overline{\mathscr{D}}_{i_2i_2}(\rho_0)\right)_{1\leqslant t_1,t_2\leqslant n}.
$$

Again, by (S.39) and (S.24),

$$
\begin{aligned}
& \operatorname{trace}(\mathbf{L}_{n5}) \\
= \ & \rho_0^4\sum_{t=1}^n\sum_{\ell=1}^n\sum_{i_1=1}^n w_{ti_1}w_{\ell i_1}\overline{\mathscr{D}}_{i_1i_1}(\rho_0)\sum_{j=1}^n\sum_{k=1}^K \xi_{\ell k}\xi_{jk}\sum_{i_2=1}^n w_{ji_2}w_{ti_2}\overline{\mathscr{D}}_{i_2i_2}(\rho_0) \\
= \ & \rho_0^4\sum_{t=1}^n\sum_{\ell=1}^n o(n^{-1/2})\sum_{j=1}^n O_p(1)o(n^{-1/2}) \\
= \ & o_p(n^2).
\end{aligned}
$$

Hence, according to above derivations, we have

$$
\begin{aligned}
E\left(\sup_{1\leqslant k\leqslant K}\tau_{1k}^2 \mid \mathbf{A}\right) &\leqslant E(\|\boldsymbol{\tau}_1\|^2 \mid \mathbf{A}) \\
&= 4\frac{1}{n^2}E(\boldsymbol{\varepsilon}_n^{\mathrm{T}}\mathscr{L}_n\boldsymbol{\varepsilon}_n \mid \mathbf{A}) \\
&= 4\frac{1}{n^2}\times o_p(n^2) \\
&= o_p(1), \qquad \text{(S.40)}
\end{aligned}
$$

and

$$
\begin{aligned}
\operatorname{var}(\tau_{1k}) &= E\{\operatorname{var}(\tau_{1k} \mid \mathbf{A})\} + \operatorname{var}\{E(\tau_{1k} \mid \mathbf{A})\} \\
&= E\left\{E(\tau_{1k}^2 \mid \mathbf{A})\right\} \\
&= o(1),
\end{aligned}
$$

uniformly for $k = 1, \ldots, K$. For the terms $\boldsymbol{\tau}_2$ and $\boldsymbol{\tau}_3$ of (S.37), similar derivations show that $\boldsymbol{\tau}_\ell = o_p(1)$ and $\sup_{1\leqslant k\leqslant K}\tau_{\ell k} = o_p(1)$ for $\ell = 2, 3$, so we omit the details. Now, incorporating (S.37) to (S.35), we can show that

$$
\begin{aligned}
&\sum_{k=1}^{K}\left[n^{-1}\sum_{i=1}^{n}\frac{\partial S_i\{\mathbf{A};\mathbf{b}(\rho_0),\rho_0\}}{\partial b_k(\rho_0)}\right]\left(n^{-3/4}\sum_{j=1}^{n}\varphi_{jk}\right) + o_p(n^{-1/4}) \\
= &\sum_{k=1}^{K}\{\tau_{4k} + o_p(1)\}\left(n^{-3/4}\sum_{j=1}^{n}\varphi_{jk}\right) + o_p(n^{-1/4}), \qquad \text{(S.41)}
\end{aligned}
$$

where τ_{4k} is the kth element of

$$
\frac{1}{n}\left\{\mathbf{Y}^{\mathrm{T}}\mathbf{W}^{\mathrm{T}}\mathbf{S}(\rho_0) + \frac{\partial\mathbf{b}(\rho_0)^{\mathrm{T}}}{\partial\rho}\mathbf{A}^{\mathrm{T}}\mathbf{S}(\rho_0)\right\}\left\{\mathbf{D}(\rho_0)^2\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{A}\right\}.
$$

By model (2.6), we have

$$\tau_{4k} \quad = \quad \mathscr{A}_k(\mathbf{A}) + \mathscr{B}_k(\mathbf{A}),$$

where

$$\mathscr{A}_k(\mathbf{A}) \quad = \quad \frac{1}{n}\boldsymbol{\varepsilon}^{\mathrm{T}}\mathbf{S}^{-\mathrm{T}}(\rho_0)\mathbf{W}^{\mathrm{T}}\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{A}\mathbf{e}_k,$$

and

$$\begin{aligned}\mathscr{B}_k(\mathbf{A}) \quad &= \quad \frac{1}{n}\mathbf{b}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{S}^{-\mathrm{T}}(\rho_0)\mathbf{W}^{\mathrm{T}}\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{A}\mathbf{e}_k \\ &\quad + \frac{1}{n}\frac{\partial \mathbf{b}(\rho_0)^{\mathrm{T}}}{\partial \rho}\mathbf{A}^{\mathrm{T}}\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{A}\mathbf{e}_k, \end{aligned} \tag{S.42}$$

where $\mathbf{e}_k$ denotes K dimension zero vector except that the kth element equals to one. Note that

$$\mathscr{A}_k(\mathbf{A}) = O_p(1/\sqrt{n}), \quad \text{and} \quad \mathscr{B}_k(\mathbf{A}) = O_p(1). \tag{S.43}$$

For presentational continuity, the proof of (S.43) is relegated to the Supplementary Material S3.12. This shows that $\tau_{4k} = \mathscr{B}_k(\mathbf{A}) + O_p(1/\sqrt{n})$. Together with (S.41) and (S.36), we have

$$\begin{aligned} & \sum_{k=1}^{K}\left[n^{-1}\sum_{i=1}^{n}\frac{\partial S_i\{\mathbf{A};\mathbf{b}(\rho_0),\rho_0\}}{\partial b_k(\rho_0)}\right]\left(n^{-3/4}\sum_{j=1}^{n}\varphi_{jk}\right) + o_p(n^{-1/4}) \\ = \quad & \sum_{k=1}^{K}\{\mathscr{B}_k(\mathbf{A}) + o_p(1)\}\left(n^{-3/4}\sum_{j=1}^{n}\varphi_{jk}\right) + o_p(n^{-1/4}) \\ = \quad & \sum_{k=1}^{K}\mathscr{B}_k(\mathbf{A})\left(n^{-3/4}\sum_{j=1}^{n}\varphi_{jk}\right) + o_p(1)\sum_{k=1}^{K}\left(n^{-3/4}\sum_{j=1}^{n}\varphi_{jk}\right) + o_p(n^{-1/4}) \end{aligned}$$

$$= \sum_{k=1}^{K} \mathscr{B}_k(\mathbf{A}) \left(n^{-3/4} \sum_{j=1}^{n} \varphi_{jk} \right) + o_p(1), \tag{S.44}$$

where the second equality is because $\sup_{1 \leqslant k \leqslant K} \tau_{\ell k} = o_p(1)$ for $\ell = 1, 2, 3$ derived from (S.40), so the first $o_p(1)$ after the first equal sign of (S.44) is uniform with respect to k. Thus, the first term of (S.32) can be expressed as

$$\begin{aligned}
& n^{-3/4} \sum_{i=1}^{n} \left[S_i\{\widehat{\mathbf{A}}; \mathbf{b}(\rho_0), \rho_0\} + \sum_{k=1}^{K} \frac{\partial S_i\{\widehat{\mathbf{A}}; \mathbf{b}^*(\rho_0), \rho_0\}}{\partial b_k^*(\rho_0)} \left\{ \widehat{b}_k(\rho_0) - b_k(\rho_0) \right\} \right] \\
= \;& n^{-3/4} \sum_{i=1}^{n} S_i\{\widehat{\mathbf{A}}; \mathbf{b}(\rho_0), \rho_0\} + n^{-3/4} \sum_{i=1}^{n} \sum_{k=1}^{K} \frac{\partial S_i\{\widehat{\mathbf{A}}; \mathbf{b}(\rho_0), \rho_0\}}{\partial b_k(\rho_0)} \left\{ \widehat{b}_k(\rho_0) - b_k(\rho_0) \right\} + o_p(n^{-1/4}) \\
= \;& n^{-3/4} \sum_{i=1}^{n} S_i\{\mathbf{A}; \mathbf{b}(\rho_0), \rho_0\} + n^{-1/4} \left[\frac{1}{\sqrt{n}} \sum_{i=1}^{n} S_i\{\widehat{\mathbf{A}}; \mathbf{b}(\rho_0), \rho_0\} - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} S_i\{\mathbf{A}; \mathbf{b}(\rho_0), \rho_0\} \right] \\
& + \sum_{k=1}^{K} \left[\frac{1}{n} \sum_{i=1}^{n} \frac{\partial S_i\{\mathbf{A}; \mathbf{b}(\rho_0), \rho_0\}}{\partial b_k(\rho_0)} \right] \left(n^{-3/4} \sum_{j=1}^{n} \varphi_{jk} \right) + o_p(n^{-1/4}) \\
= \;& n^{-3/4} \sum_{i=1}^{n} S_i\{\mathbf{A}; \mathbf{b}(\rho_0), \rho_0\} + n^{-1/4} \left[\frac{1}{\sqrt{n}} \sum_{i=1}^{n} S_i\{\widehat{\mathbf{A}}; \mathbf{b}(\rho_0), \rho_0\} - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} S_i\{\mathbf{A}; \mathbf{b}(\rho_0), \rho_0\} \right] \\
& + \sum_{k=1}^{K} \mathscr{B}_k(\mathbf{A}) \left(n^{-3/4} \sum_{j=1}^{n} \varphi_{jk} \right) + o_p(1) \\
= \;& n^{-3/4} \sum_{i=1}^{n} S_i\{\mathbf{A}; \mathbf{b}(\rho_0), \rho_0\} + O_p(n^{-1/4}) + \sum_{k=1}^{K} \mathscr{B}_k(\mathbf{A}) \left(n^{-3/4} \sum_{j=1}^{n} \varphi_{jk} \right) + o_p(1) \\
= \;& n^{-3/4} \sum_{i=1}^{n} S_i\{\mathbf{A}; \mathbf{b}(\rho_0), \rho_0\} + \sum_{k=1}^{K} \mathscr{B}_k(\mathbf{A}) \left(n^{-3/4} \sum_{j=1}^{n} \varphi_{jk} \right) + o_p(1),
\end{aligned} \tag{S.45}$$

where the first equality is due to (S.34), the second equality is because (S.35), the third equality follows from (S.44), and the last second equality is obtained by (S.2). Now, we will focus on the first two leading terms of (S.45). For the second term of (S.45), with $E(\varphi_{jk} \mid \mathbf{A}) = 0$ from (S.28), it is easy to see that

$$E\left\{ \sum_{k=1}^{K} \mathscr{B}_k(\mathbf{A}) \left(n^{-3/4} \sum_{j=1}^{n} \varphi_{jk} \right) \right\}$$

$$
\begin{aligned}
&= E\left[\sum_{k=1}^{K}\mathscr{B}_k(\mathbf{A})E\left\{\left(n^{-3/4}\sum_{j=1}^{n}\varphi_{jk}\right)\mid\mathbf{A}\right\}\right]\\
&= 0. \qquad\qquad (\mathrm{S}.46)
\end{aligned}
$$

For the first term of (S.45), according to (2.10) with $\widehat{\mathbf{A}}$ replaced by $\mathbf{A}$, it is easy to see that $S_i\{\mathbf{A};\mathbf{b}(\rho),\rho\}$ can be expressed as

$$
S_i\{\mathbf{A};\mathbf{b}(\rho),\rho\} = \{Y_i - E\left(Y_i|\mathbf{Y}_{-i},\mathbf{A}\right)\} f\left(\mathbf{Y}_{-i},\mathbf{A}\right),
$$

where $f\left(\mathbf{Y}_{-i},\mathbf{A}\right)$ denotes a function of $\mathbf{Y}_{-i}$ and $\mathbf{A}$. Thus,

$$
\begin{aligned}
E\left[S_i\{\mathbf{A};\mathbf{b}(\rho),\rho\}\right] &= E\left(E\left[S_i\{\mathbf{A};\mathbf{b}(\rho),\rho\}\mid\mathbf{Y}_{-i},\mathbf{A}\right]\right)\\
&= E\left[f\left(\mathbf{Y}_{-i},\mathbf{A}\right)E\left\{Y_i - E\left(Y_i|\mathbf{Y}_{-i},\mathbf{A}\right)|\mathbf{Y}_{-i},\mathbf{A}\right\}\right]\\
&= 0. \qquad\qquad (\mathrm{S}.47)
\end{aligned}
$$

Therefore, by (S.46), (S.47) and (S.45), we have

$$
\begin{aligned}
& E\left(n^{-3/4}\sum_{i=1}^{n}\left[S_i\{\widehat{\mathbf{A}};\mathbf{b}(\rho_0),\rho_0\} + \sum_{k=1}^{K}\frac{\partial S_i\{\widehat{\mathbf{A}};\mathbf{b}^*(\rho_0),\rho_0\}}{\partial b_k^*(\rho_0)}\left\{\widehat{b}_k(\rho_0)-b_k(\rho_0)\right\}\right]\right)\\
&= o(1). \qquad\qquad (\mathrm{S}.48)
\end{aligned}
$$

This demonstrates the expectation of first term of (S.32) equals to zero asymptotically.

Next, we investigate the variance of the first two terms of (S.45). First, for the second term of (S.45), according to the form of (S.28), we have $E(\varphi_{jk_1}\varphi_{j'k_2}\mid\mathbf{A}) = E(\varphi_{jk_1}\mid\mathbf{A})E(\varphi_{j'k_2}\mid\mathbf{A})$ for any $j\neq j'$ and $k_1,k_2=1,\ldots,K$, and φ_{jk}'s are i.i.d for $j=1,\ldots,n$.

Then,

$$\begin{aligned}
& \operatorname{var}\left[n^{-3/4}\sum_{j=1}^{n}\left\{\sum_{k=1}^{K}\mathscr{B}_k(\mathbf{A})\varphi_{jk}\right\}\right] \\
= \;& E\left(\operatorname{var}\left[n^{-3/4}\sum_{j=1}^{n}\left\{\sum_{k=1}^{K}\mathscr{B}_k(\mathbf{A})\varphi_{jk}\right\} \mid \mathbf{A}\right]\right) \\
= \;& E\left\{E\left(\left[n^{-1/4}\cdot n^{-1/2}\sum_{j=1}^{n}\left\{\sum_{k=1}^{K}\mathscr{B}_k(\mathbf{A})\varphi_{jk}\right\}\right]^2 \mid \mathbf{A}\right)\right\} \\
= \;& n^{-1/2}E\left(E\left[\frac{1}{n}\sum_{j=1}^{n}\left\{\sum_{k=1}^{K}\mathscr{B}_k(\mathbf{A})\varphi_{jk}\right\}^2 \mid \mathbf{A}\right]\right) \\
& + n^{-1/2}E\left(E\left[\frac{1}{n}\sum_{j\neq j'}\left\{\sum_{k_1=1}^{K}\sum_{k_2=1}^{K}\mathscr{B}_{k_1}(\mathbf{A})\mathscr{B}_{k_2}(\mathbf{A})\varphi_{jk_1}\varphi_{j'k_2}\right\} \mid \mathbf{A}\right]\right) \\
= \;& n^{-1/2}E\left(E\left[\left\{\sum_{k=1}^{K}\mathscr{B}_k(\mathbf{A})\varphi_{jk}\right\}^2 \mid \mathbf{A}\right]\right).
\end{aligned} \tag{S.49}$$

Second, for the first term of (S.45), we have

$$\begin{aligned}
& \operatorname{var}\left[n^{-3/4}\sum_{i=1}^{n}S_i\{\mathbf{A};\mathbf{b}(\rho_0),\rho_0\}\right] \\
= \;& E\left[n^{-3/4}\sum_{i=1}^{n}S_i\{\mathbf{A};\mathbf{b}(\rho_0),\rho_0\}\right]^2 \\
= \;& n^{-1/2}E\left[n^{-1/2}\sum_{i=1}^{n}S_i\{\mathbf{A};\mathbf{b}(\rho_0),\rho_0\}\right]^2.
\end{aligned} \tag{S.50}$$

We now further rewrite $n^{-1/2}\sum_{i=1}^{n}S_i\{\mathbf{A};\mathbf{b}(\rho_0),\rho_0\}$ as a quadratic form plus a linear form, then we use Lemma S2.9 to calculate (S.50). Using (2.15) with $\widehat{\mathbf{A}}$ and $\widehat{\mathbf{b}}(\rho)$ replaced by $\mathbf{A}$ and $\mathbf{b}(\rho_0)$, respectively, we have

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}S_i\{\mathbf{A};\mathbf{b}(\rho_0),\rho_0\}$$

$$
\begin{aligned}
&= \frac{1}{\sqrt{n}} \left\{\mathbf{D}(\rho_0)\mathbf{M}(\rho_0)\mathbf{Y} - \mathbf{D}(\rho_0)\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{A}\mathbf{b}(\rho_0)\right\}^{\mathrm{T}} \frac{\partial \left[\mathbf{D}(\rho_0)\mathbf{S}(\rho_0)^{\mathrm{T}}\left\{\mathbf{S}(\rho_0)\mathbf{Y} - \mathbf{A}\mathbf{b}(\rho_0)\right\}\right]}{\partial \rho} \\
&= \frac{1}{\sqrt{n}} \left\{\mathbf{D}(\rho_0)\mathbf{M}(\rho_0)\mathbf{Y} - \mathbf{D}(\rho_0)\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{A}\mathbf{b}(\rho_0)\right\}^{\mathrm{T}} \\
&\quad \times \left[\frac{\partial \mathbf{D}(\rho_0)}{\partial \rho}\mathbf{S}(\rho_0)^{\mathrm{T}}\left\{\mathbf{S}(\rho_0)\mathbf{Y} - \mathbf{A}\mathbf{b}(\rho_0)\right\} + \mathbf{D}(\rho)\frac{\partial \mathbf{S}(\rho_0)^{\mathrm{T}}}{\partial \rho}\left\{\mathbf{S}(\rho_0)\mathbf{Y} - \mathbf{A}\mathbf{b}(\rho_0)\right\}\right. \\
&\quad \left. + \mathbf{D}(\rho_0)\mathbf{S}(\rho_0)^{\mathrm{T}}\frac{\partial \mathbf{S}(\rho_0)}{\partial \rho}\mathbf{Y} - \mathbf{D}(\rho_0)\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{A}\frac{\partial \mathbf{b}(\rho_0)}{\partial \rho}\right] \\
&= \frac{1}{\sqrt{n}} \left\{\mathbf{D}(\rho_0)\mathbf{M}(\rho_0)\mathbf{Y} - \mathbf{D}(\rho_0)\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{A}\mathbf{b}(\rho_0)\right\}^{\mathrm{T}} \\
&\quad \times \left[\frac{\partial \mathbf{D}(\rho_0)}{\partial \rho}\mathbf{S}(\rho_0)^{\mathrm{T}}\left\{\mathbf{S}(\rho_0)\mathbf{Y} - \mathbf{A}\mathbf{b}(\rho_0)\right\} + \mathbf{D}(\rho)\frac{\partial \mathbf{S}(\rho_0)^{\mathrm{T}}}{\partial \rho}\left\{\mathbf{S}(\rho_0)\mathbf{Y} - \mathbf{A}\mathbf{b}(\rho_0)\right\}\right. \\
&\quad \left. + \mathbf{D}(\rho_0)\mathbf{S}(\rho_0)^{\mathrm{T}}\frac{\partial \mathbf{S}(\rho_0)}{\partial \rho}\mathbf{S}(\rho_0)^{-1}\mathbf{S}(\rho_0)\mathbf{Y} - \mathbf{D}(\rho_0)\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{A}\frac{\partial \mathbf{b}(\rho_0)}{\partial \rho}\right] \\
&= \frac{1}{\sqrt{n}} \left\{\mathbf{D}(\rho_0)\mathbf{M}(\rho_0)\mathbf{Y} - \mathbf{D}(\rho_0)\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{A}\mathbf{b}(\rho_0)\right\}^{\mathrm{T}} \\
&\quad \times \left[\frac{\partial \mathbf{D}(\rho_0)}{\partial \rho}\mathbf{S}(\rho_0)^{\mathrm{T}}\left\{\mathbf{S}(\rho_0)\mathbf{Y} - \mathbf{A}\mathbf{b}(\rho_0)\right\} + \mathbf{D}(\rho)\frac{\partial \mathbf{S}(\rho_0)^{\mathrm{T}}}{\partial \rho}\left\{\mathbf{S}(\rho_0)\mathbf{Y} - \mathbf{A}\mathbf{b}(\rho_0)\right\}\right. \\
&\quad + \mathbf{D}(\rho_0)\mathbf{S}(\rho_0)^{\mathrm{T}}\frac{\partial \mathbf{S}(\rho_0)}{\partial \rho}\mathbf{S}(\rho_0)^{-1}\left\{\mathbf{S}(\rho_0)\mathbf{Y} - \mathbf{A}\mathbf{b}(\rho_0)\right\} \\
&\quad \left. + \mathbf{D}(\rho_0)\mathbf{S}(\rho_0)^{\mathrm{T}}\frac{\partial \mathbf{S}(\rho_0)}{\partial \rho}\mathbf{S}(\rho_0)^{-1}\mathbf{A}\mathbf{b}(\rho_0) - \mathbf{D}(\rho_0)\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{A}\frac{\partial \mathbf{b}(\rho_0)}{\partial \rho}\right] \\
&= \frac{1}{\sqrt{n}} \left\{\mathbf{S}(\rho_0)\mathbf{Y} - \mathbf{A}\mathbf{b}(\rho_0)\right\}^{\mathrm{T}}\mathbf{S}(\rho_0)\mathbf{D}(\rho_0) \\
&\quad \times \left\{\frac{\partial \mathbf{D}(\rho_0)}{\partial \rho}\mathbf{S}(\rho_0)^{\mathrm{T}} + \mathbf{D}(\rho)\frac{\partial \mathbf{S}(\rho_0)^{\mathrm{T}}}{\partial \rho} + \mathbf{D}(\rho_0)\mathbf{S}(\rho_0)^{\mathrm{T}}\frac{\partial \mathbf{S}(\rho_0)}{\partial \rho}\mathbf{S}(\rho_0)^{-1}\right\}\left\{\mathbf{S}(\rho_0)\mathbf{Y} - \mathbf{A}\mathbf{b}(\rho_0)\right\} \\
&\quad + \frac{1}{\sqrt{n}}\left\{\mathbf{S}(\rho_0)\mathbf{Y} - \mathbf{A}\mathbf{b}(\rho_0)\right\}^{\mathrm{T}}\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\mathbf{S}(\rho_0)^{\mathrm{T}}\left\{\frac{\partial \mathbf{S}(\rho_0)}{\partial \rho}\mathbf{S}(\rho_0)^{-1}\mathbf{A}\mathbf{b}(\rho_0) - \mathbf{A}\frac{\partial \mathbf{b}(\rho_0)}{\partial \rho}\right\} \\
&= \frac{1}{\sqrt{n}}\boldsymbol{\varepsilon}_n^{\mathrm{T}}\mathbf{M}_n\boldsymbol{\varepsilon}_n + \frac{1}{\sqrt{n}}\boldsymbol{\varepsilon}_n^{\mathrm{T}}\mathbf{U}_n, \qquad \text{(S.51)}
\end{aligned}
$$

where

$$
\begin{aligned}
\mathbf{M}_n &= \mathbf{S}(\rho_0)\mathbf{D}(\rho_0)\frac{\partial \mathbf{D}(\rho_0)}{\partial \rho}\mathbf{S}(\rho_0)^{\mathrm{T}} + \mathbf{S}(\rho_0)\mathbf{D}(\rho)^2\frac{\partial \mathbf{S}(\rho_0)^{\mathrm{T}}}{\partial \rho} \\
&\quad + \mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\mathbf{S}(\rho_0)^{\mathrm{T}}\frac{\partial \mathbf{S}(\rho_0)}{\partial \rho}\mathbf{S}(\rho_0)^{-1}, \qquad \text{(S.52)}
\end{aligned}
$$

$$\mathbf{U}_n = \mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\mathbf{S}(\rho_0)^{\mathrm{T}}\left\{\frac{\partial \mathbf{S}(\rho_0)}{\partial \rho}\mathbf{S}(\rho_0)^{-1}\mathbf{A}\mathbf{b}(\rho_0) - \mathbf{A}\frac{\partial \mathbf{b}(\rho_0)}{\partial \rho}\right\}, \tag{S.53}$$

and

$$\boldsymbol{\varepsilon}_n = \mathbf{S}(\rho_0)\mathbf{Y} - \mathbf{A}\mathbf{b}(\rho_0),$$

and $\boldsymbol{\varepsilon}_n \mid \mathbf{A} \sim \mathcal{N}_n(\mathbf{0}, \sigma^2\mathbf{I}_n)$.

Next we check $\mathrm{trace}(\mathbf{M}_n^2)$, $\mathrm{trace}(\mathbf{M}_n\mathbf{M}_n^{\mathrm{T}})$, $\mathrm{trace}\left\{\mathrm{diag}^2(\mathbf{M}_n)\right\}$, and $\mathbf{U}_n^{\mathrm{T}}\mathbf{U}_n$, respectively. By $\mathrm{trace}(\mathbf{AB}) = \mathrm{trace}(\mathbf{BA})$ for any compatible dimension matrices $\mathbf{A}$ and $\mathbf{B}$, it follows that

$$\begin{aligned}
& \mathrm{trace}(\mathbf{M}_n^2) \\
= \; & \mathrm{trace}\left[\left\{\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)\frac{\partial \mathbf{D}(\rho_0)}{\partial \rho}\mathbf{S}(\rho_0)^{\mathrm{T}} + \mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\frac{\partial \mathbf{S}(\rho_0)^{\mathrm{T}}}{\partial \rho}\right.\right. \\
& \left.\left. + \mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\mathbf{S}(\rho_0)^{\mathrm{T}}\frac{\partial \mathbf{S}(\rho_0)}{\partial \rho}\mathbf{S}(\rho_0)^{-1}\right\}^2\right] \\
= \; & \mathrm{trace}\left\{\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)\frac{\partial \mathbf{D}(\rho_0)}{\partial \rho}\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)\frac{\partial \mathbf{D}(\rho_0)}{\partial \rho}\mathbf{S}(\rho_0)^{\mathrm{T}}\right\} \\
& + \mathrm{trace}\left\{\mathbf{S}(\rho_0)\mathbf{D}(\rho)^2\frac{\partial \mathbf{S}(\rho_0)^{\mathrm{T}}}{\partial \rho}\mathbf{S}(\rho_0)\mathbf{D}(\rho)^2\frac{\partial \mathbf{S}(\rho_0)^{\mathrm{T}}}{\partial \rho}\right\} \\
& + \mathrm{trace}\left\{\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\mathbf{S}(\rho_0)^{\mathrm{T}}\frac{\partial \mathbf{S}(\rho_0)}{\partial \rho}\mathbf{D}(\rho_0)^2\mathbf{S}(\rho_0)^{\mathrm{T}}\frac{\partial \mathbf{S}(\rho_0)}{\partial \rho}\mathbf{S}(\rho_0)^{-1}\right\} \\
& + 2\mathrm{trace}\left\{\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)\frac{\partial \mathbf{D}(\rho_0)}{\partial \rho}\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{S}(\rho_0)\mathbf{D}(\rho)^2\frac{\partial \mathbf{S}(\rho_0)^{\mathrm{T}}}{\partial \rho}\right\} \\
& + 2\mathrm{trace}\left\{\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)\frac{\partial \mathbf{D}(\rho_0)}{\partial \rho}\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\mathbf{S}(\rho_0)^{\mathrm{T}}\frac{\partial \mathbf{S}(\rho_0)}{\partial \rho}\mathbf{S}(\rho_0)^{-1}\right\} \\
& + 2\mathrm{trace}\left\{\mathbf{S}(\rho_0)\mathbf{D}(\rho)^2\frac{\partial \mathbf{S}(\rho_0)^{\mathrm{T}}}{\partial \rho}\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\mathbf{S}(\rho_0)^{\mathrm{T}}\frac{\partial \mathbf{S}(\rho_0)}{\partial \rho}\mathbf{S}(\rho_0)^{-1}\right\} \\
= \; & \mathrm{trace}\left\{\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)\frac{\partial \mathbf{D}(\rho_0)}{\partial \rho}\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)\frac{\partial \mathbf{D}(\rho_0)}{\partial \rho}\mathbf{S}(\rho_0)^{\mathrm{T}}\right\} \\
& + \mathrm{trace}\left\{\mathbf{S}(\rho_0)\mathbf{D}(\rho)^2\frac{\partial \mathbf{S}(\rho_0)^{\mathrm{T}}}{\partial \rho}\mathbf{S}(\rho_0)\mathbf{D}(\rho)^2\frac{\partial \mathbf{S}(\rho_0)^{\mathrm{T}}}{\partial \rho}\right\}
\end{aligned}$$

$$
\begin{aligned}
& +\mathrm{trace}\left\{\mathbf{D}(\rho_0)^2\mathbf{S}(\rho_0)^{\mathrm{T}}\frac{\partial\mathbf{S}(\rho_0)}{\partial\rho}\mathbf{D}(\rho_0)^2\mathbf{S}(\rho_0)^{\mathrm{T}}\frac{\partial\mathbf{S}(\rho_0)}{\partial\rho}\right\} \\
& +2\mathrm{trace}\left\{\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)\frac{\partial\mathbf{D}(\rho_0)}{\partial\rho}\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{S}(\rho_0)\mathbf{D}(\rho)^2\frac{\partial\mathbf{S}(\rho_0)^{\mathrm{T}}}{\partial\rho}\right\} \\
& +2\mathrm{trace}\left\{\mathbf{D}(\rho_0)^2\mathbf{S}(\rho_0)^{\mathrm{T}}\frac{\partial\mathbf{S}(\rho_0)}{\partial\rho}\mathbf{D}(\rho_0)\frac{\partial\mathbf{D}(\rho_0)}{\partial\rho}\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{S}(\rho_0)\right\} \\
& +2\mathrm{trace}\left\{\mathbf{D}(\rho_0)^2\mathbf{S}(\rho_0)^{\mathrm{T}}\frac{\partial\mathbf{S}(\rho_0)}{\partial\rho}\mathbf{D}(\rho_0)^2\frac{\partial\mathbf{S}(\rho_0)^{\mathrm{T}}}{\partial\rho}\mathbf{S}(\rho_0)\right\} \\
= \; & \mathscr{T}_1. \qquad \text{(S.54)}
\end{aligned}
$$

Recall that $w_{ii}=0$, $\sum_{\ell=1}^n w_{i\ell}=1$. Let $\overline{\mathbf{D}}(\rho)\equiv\frac{\partial\mathbf{D}(\rho)}{\partial\rho}\mathbf{D}(\rho)$, then the diagonal elements of the $\overline{\mathbf{D}}(\rho)$ were denoted by $\overline{\mathscr{D}}_{ii}(\rho)$ and $\sup_{1\leqslant i\leqslant n}\overline{\mathscr{D}}_{ii}(\rho)\leqslant C_1$ for some constant C_1. For the first term of $\mathscr{T}_1$, by (S.24) and the Condition (C9), we note that

$$
\begin{aligned}
& \mathrm{trace}\left\{\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)\frac{\partial\mathbf{D}(\rho_0)}{\partial\rho}\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)\frac{\partial\mathbf{D}(\rho_0)}{\partial\rho}\mathbf{S}(\rho_0)^{\mathrm{T}}\right\} \\
= \; & \mathrm{trace}\left\{(\mathbf{I}_n-\rho_0\mathbf{W})\overline{\mathbf{D}}(\rho_0)(\mathbf{I}_n-\rho_0\mathbf{W}^{\mathrm{T}})(\mathbf{I}_n-\rho_0\mathbf{W})\overline{\mathbf{D}}(\rho_0)(\mathbf{I}_n-\rho_0\mathbf{W}^{\mathrm{T}})\right\} \\
= \; & \mathrm{trace}\left\{(\mathbf{I}_n-\rho_0\mathbf{W})\overline{\mathbf{D}}(\rho_0)(\mathbf{I}_n-\rho_0\mathbf{W}-\rho_0\mathbf{W}^{\mathrm{T}}+\rho_0^2\mathbf{W}^{\mathrm{T}}\mathbf{W})\overline{\mathbf{D}}(\rho_0)(\mathbf{I}_n-\rho_0\mathbf{W}^{\mathrm{T}})\right\} \\
= \; & \mathrm{trace}\left\{\overline{\mathbf{D}}(\rho_0)(\mathbf{I}_n-\rho_0\mathbf{W}-\rho_0\mathbf{W}^{\mathrm{T}}+\rho_0^2\mathbf{W}^{\mathrm{T}}\mathbf{W})\overline{\mathbf{D}}(\rho_0)\right\} \\
& +\mathrm{trace}\left\{\overline{\mathbf{D}}(\rho_0)(\mathbf{I}_n-\rho_0\mathbf{W}-\rho_0\mathbf{W}^{\mathrm{T}}+\rho_0^2\mathbf{W}^{\mathrm{T}}\mathbf{W})\overline{\mathbf{D}}(\rho_0)(-\rho_0\mathbf{W}^{\mathrm{T}})\right\} \\
& +\mathrm{trace}\left\{(-\rho_0\mathbf{W})\overline{\mathbf{D}}(\rho_0)(\mathbf{I}_n-\rho_0\mathbf{W}-\rho_0\mathbf{W}^{\mathrm{T}}+\rho_0^2\mathbf{W}^{\mathrm{T}}\mathbf{W})\overline{\mathbf{D}}(\rho_0)\right\} \\
& +\mathrm{trace}\left\{(-\rho_0\mathbf{W})\overline{\mathbf{D}}(\rho_0)(\mathbf{I}_n-\rho_0\mathbf{W}-\rho_0\mathbf{W}^{\mathrm{T}}+\rho_0^2\mathbf{W}^{\mathrm{T}}\mathbf{W})\overline{\mathbf{D}}(\rho_0)(-\rho_0\mathbf{W}^{\mathrm{T}})\right\} \\
= \; & \sum_{i=1}^n\overline{\mathscr{D}}_{ii}^2(\rho_0)+\rho_0^2\sum_{j=1}^n\sum_{i=1}^n w_{ij}^2\overline{\mathscr{D}}_{jj}^2(\rho_0) \\
& +\rho_0^2\sum_{j=1}^n\sum_{i=1}^n w_{ji}^2\overline{\mathscr{D}}_{jj}(\rho_0)\overline{\mathscr{D}}_{ii}(\rho_0)+\rho_0^2\sum_{j=1}^n\sum_{i=1}^n w_{ij}w_{ji}\overline{\mathscr{D}}_{jj}(\rho_0)\overline{\mathscr{D}}_{ii}(\rho_0) \\
& -\rho_0^3\sum_{\ell=1}^n\sum_{j=1}^n\sum_{i=1}^n w_{i\ell}w_{ij}w_{\ell j}\overline{\mathscr{D}}_{\ell\ell}(\rho_0)\overline{\mathscr{D}}_{jj}(\rho_0) \\
& +\rho_0^2\sum_{j=1}^n\sum_{i=1}^n w_{ji}w_{ij}\overline{\mathscr{D}}_{ii}(\rho_0)\overline{\mathscr{D}}_{jj}(\rho_0)+\rho_0^2\sum_{j=1}^n\sum_{i=1}^n w_{ji}^2\overline{\mathscr{D}}_{ii}(\rho_0)\overline{\mathscr{D}}_{jj}(\rho_0)
\end{aligned}
$$

$$
\begin{aligned}
& -\rho_0^3 \sum_{\ell=1}^n \sum_{j=1}^n \sum_{i=1}^n w_{\ell i} w_{ji} w_{j\ell} \overline{\mathscr{D}}_{ii}(\rho_0) \overline{\mathscr{D}}_{\ell\ell}(\rho_0) \\
& +\rho_0^2 \sum_{j=1}^n \sum_{i=1}^n w_{ji}^2 \overline{\mathscr{D}}_{ii}^2(\rho_0) - \rho_0^3 \sum_{\ell=1}^n \sum_{j=1}^n \sum_{i=1}^n w_{\ell i} w_{ij} w_{\ell j} \overline{\mathscr{D}}_{ii}(\rho_0) \overline{\mathscr{D}}_{jj}(\rho_0) \\
& -\rho_0^3 \sum_{\ell=1}^n \sum_{j=1}^n \sum_{i=1}^n w_{\ell i} w_{ji} w_{\ell j} \overline{\mathscr{D}}_{ii}(\rho_0) \overline{\mathscr{D}}_{jj}(\rho_0) + \rho_0^4 \sum_{\ell=1}^n \sum_{j=1}^n \left(\sum_{i=1}^n w_{\ell i} \overline{\mathscr{D}}_{ii}(\rho_0) w_{ji} \right) \left(\sum_{i=1}^n w_{ji} \overline{\mathscr{D}}_{ii}(\rho_0) w_{\ell i} \right) \\
\leqslant \; & nC_1^2 + \rho_0^2 \sum_{j=1}^n \sum_{i=1}^n w_{ij} \frac{C}{\sqrt{n}} C_1^2 + \rho_0^2 \sum_{j=1}^n \sum_{i=1}^n w_{ji} \frac{C}{\sqrt{n}} C_1^2 + \rho_0^2 \sum_{j=1}^n \sum_{i=1}^n w_{ij} \frac{C}{\sqrt{n}} C_1^2 \\
& +|\rho_0|^3 \sum_{\ell=1}^n \sum_{i=1}^n w_{i\ell} \left(\sum_{j=1}^n w_{ij} w_{\ell j} \right) C_1^2 + \rho_0^2 \sum_{j=1}^n \sum_{i=1}^n w_{ji} w_{ij} C_1^2 \\
& +\rho_0^2 \sum_{j=1}^n \sum_{i=1}^n w_{ji}^2 C_1^2 + |\rho_0|^3 \sum_{\ell=1}^n \sum_{j=1}^n \left(\sum_{i=1}^n w_{\ell i} w_{ji} \right) w_{j\ell} C_1^2 \\
& +\rho_0^2 \sum_{j=1}^n \sum_{i=1}^n w_{ji}^2 C_1^2 + |\rho_0|^3 \sum_{\ell=1}^n \sum_{i=1}^n w_{\ell i} \left(\sum_{j=1}^n w_{ij} w_{\ell j} \right) C_1^2 \\
& +|\rho_0|^3 \sum_{\ell=1}^n \sum_{j=1}^n \left(\sum_{i=1}^n w_{\ell i} w_{ji} \right) w_{\ell j} C_1^2 + \rho_0^4 \sum_{\ell=1}^n \sum_{j=1}^n \left(\sum_{i=1}^n w_{\ell i} w_{ji} \right) \left(\sum_{i=1}^n w_{ji} w_{\ell i} \right) C_1^2 \\
\leqslant \; & nC_1^2 + \rho_0^2 \sum_{i=1}^n \frac{C}{\sqrt{n}} C_1^2 + \rho_0^2 \sum_{j=1}^n \frac{C}{\sqrt{n}} C_1^2 + \rho_0^2 \sum_{i=1}^n \frac{C}{\sqrt{n}} C_1^2 \\
& +|\rho_0|^3 \sum_{i=1}^n \sum_{\ell=1}^n w_{i\ell} \left(\frac{C}{\sqrt{n}} \right) C_1^2 + \rho_0^2 \sum_{j=1}^n \sum_{i=1}^n w_{ji} \frac{C}{\sqrt{n}} C_1^2 \\
& +\rho_0^2 \sum_{j=1}^n \sum_{i=1}^n w_{ji} \frac{C}{\sqrt{n}} C_1^2 + |\rho_0|^3 \sum_{j=1}^n \sum_{\ell=1}^n w_{j\ell} \left(\frac{C}{\sqrt{n}} \right) C_1^2 \\
& +\rho_0^2 \sum_{j=1}^n \sum_{i=1}^n w_{ji} \frac{C}{\sqrt{n}} C_1^2 + |\rho_0|^3 \sum_{\ell=1}^n \sum_{i=1}^n w_{\ell i} \left(\frac{C}{\sqrt{n}} \right) C_1^2 \\
& +|\rho_0|^3 \sum_{\ell=1}^n \sum_{j=1}^n w_{\ell j} \left(\frac{C}{\sqrt{n}} \right) C_1^2 + \rho_0^4 \sum_{\ell=1}^n \sum_{j=1}^n \left(\frac{C}{\sqrt{n}} \right)^2 C_1^2 \\
= \; & C_1^2 n + 6\rho_0^2 C C_1^2 \sqrt{n} + 4|\rho_0|^3 C C_1^2 \sqrt{n} + \rho_0^4 C^2 C_1^2 n.
\end{aligned}
$$

For the other terms of $\mathscr{T}_1$, similar derivations lead to the same order $O(n)$, so we omit the details. Hence, there exists some constant κ_1 such that

$$\frac{1}{n}\text{trace}(\mathbf{M}_n^2) \quad \leqslant \quad \kappa_1.$$

By trace $\left\{(\mathbf{A}+\mathbf{B}+\mathbf{C})(\mathbf{A}^{\mathrm{T}}+\mathbf{B}^{\mathrm{T}}+\mathbf{C}^{\mathrm{T}})\right\} \leqslant 3\text{trace}\left(\mathbf{A}\mathbf{A}^{\mathrm{T}}+\mathbf{B}\mathbf{B}^{\mathrm{T}}+\mathbf{C}\mathbf{C}^{\mathrm{T}}\right)$ for any $n\times n$ matrices $\mathbf{A},\mathbf{B},\mathbf{C}$, it is also noticed from (S.52) that

$$\begin{aligned}
& \text{trace}(\mathbf{M}_n\mathbf{M}_n^{\mathrm{T}}) \\
= \;& \text{trace}\left[\left\{\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)\frac{\partial\mathbf{D}(\rho_0)}{\partial\rho}\mathbf{S}(\rho_0)^{\mathrm{T}}+\mathbf{S}(\rho_0)\mathbf{D}(\rho)^2\frac{\partial\mathbf{S}(\rho_0)^{\mathrm{T}}}{\partial\rho}\right.\right. \\
& \qquad \left. +\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\mathbf{S}(\rho_0)^{\mathrm{T}}\frac{\partial\mathbf{S}(\rho_0)}{\partial\rho}\mathbf{S}(\rho_0)^{-1}\right\} \\
& \times\left\{\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)\frac{\partial\mathbf{D}(\rho_0)}{\partial\rho}\mathbf{S}(\rho_0)^{\mathrm{T}}+\mathbf{S}(\rho_0)\mathbf{D}(\rho)^2\frac{\partial\mathbf{S}(\rho_0)^{\mathrm{T}}}{\partial\rho}\right. \\
& \qquad \left.\left. +\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\mathbf{S}(\rho_0)^{\mathrm{T}}\frac{\partial\mathbf{S}(\rho_0)}{\partial\rho}\mathbf{S}(\rho_0)^{-1}\right\}^{\mathrm{T}}\right] \\
\leqslant \;& 3\text{trace}\left\{\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)\frac{\partial\mathbf{D}(\rho_0)}{\partial\rho}\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{S}(\rho_0)\frac{\partial\mathbf{D}(\rho_0)}{\partial\rho}\mathbf{D}(\rho_0)\mathbf{S}(\rho_0)^{\mathrm{T}}\right\} \\
& +3\text{trace}\left\{\mathbf{S}(\rho_0)\mathbf{D}(\rho)^2\frac{\partial\mathbf{S}(\rho_0)^{\mathrm{T}}}{\partial\rho}\frac{\partial\mathbf{S}(\rho_0)}{\partial\rho}\mathbf{D}(\rho)^2\mathbf{S}(\rho_0)^{\mathrm{T}}\right\} \\
& +3\text{trace}\left\{\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\mathbf{S}(\rho_0)^{\mathrm{T}}\frac{\partial\mathbf{S}(\rho_0)}{\partial\rho}\mathbf{S}(\rho_0)^{-1}\mathbf{S}(\rho_0)^{-\mathrm{T}}\frac{\partial\mathbf{S}(\rho_0)^{\mathrm{T}}}{\partial\rho}\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\mathbf{S}(\rho_0)^{\mathrm{T}}\right\} \\
= \;& \mathscr{T}_2.
\end{aligned} \tag{S.55}$$

For the first two terms of $\mathscr{T}_2$ in (S.55), similar to the treatment of $\mathscr{T}_1$ in (S.54), it is easy to show that they are of order $O(n)$, and the details are omitted. So, we only need to tackle the third term of (S.55). For brevity, we denote the third term by $\mathscr{T}_{2,3}$. Since matrices $\{\partial\mathbf{S}(\rho_0)^{\mathrm{T}}\partial\rho\}\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\mathbf{S}(\rho_0)^{\mathrm{T}}\{\partial\mathbf{S}(\rho_0)\partial\rho\}$ and $\mathbf{S}(\rho_0)^{-1}\mathbf{S}(\rho_0)^{-\mathrm{T}}$ are both

$n \times n$ positive semidefinite, by Lemma S2.8, we have

$$
\begin{aligned}
&\mathscr{T}_{2,3} \\
= \;& \operatorname{trace}\left\{\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\mathbf{S}(\rho_0)^{\mathrm{T}}\frac{\partial \mathbf{S}(\rho_0)}{\partial \rho}\mathbf{S}(\rho_0)^{-1}\mathbf{S}(\rho_0)^{-\mathrm{T}}\frac{\partial \mathbf{S}(\rho_0)^{\mathrm{T}}}{\partial \rho}\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\mathbf{S}(\rho_0)^{\mathrm{T}}\right\} \\
= \;& \operatorname{trace}\left\{\mathbf{S}(\rho_0)^{-1}\mathbf{S}(\rho_0)^{-\mathrm{T}}\frac{\partial \mathbf{S}(\rho_0)^{\mathrm{T}}}{\partial \rho}\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\mathbf{S}(\rho_0)^{\mathrm{T}}\frac{\partial \mathbf{S}(\rho_0)}{\partial \rho}\right\} \\
\leqslant \;& \lambda_{\max}\left\{\mathbf{S}(\rho_0)^{-1}\mathbf{S}(\rho_0)^{-\mathrm{T}}\right\} \\
& \times \operatorname{trace}\left\{\frac{\partial \mathbf{S}(\rho_0)^{\mathrm{T}}}{\partial \rho}\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\mathbf{S}(\rho_0)^{\mathrm{T}}\frac{\partial \mathbf{S}(\rho_0)}{\partial \rho}\right\}. \qquad \text{(S.56)}
\end{aligned}
$$

For the sake of notation, let

$$
\mathscr{T}_{2,3a} = \lambda_{\max}\left\{\mathbf{S}(\rho_0)^{-1}\mathbf{S}(\rho_0)^{-\mathrm{T}}\right\},
$$

and

$$
\mathscr{T}_{2,3b} = \operatorname{trace}\left\{\frac{\partial \mathbf{S}(\rho_0)^{\mathrm{T}}}{\partial \rho}\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\mathbf{S}(\rho_0)^{\mathrm{T}}\frac{\partial \mathbf{S}(\rho_0)}{\partial \rho}\right\}.
$$

Note that $\mathbf{M} = \mathbf{S}^{\mathrm{T}}\mathbf{S}$ and for some $C_1 \geqslant 0$ and $C_2 > 0$, we have $C_1 \leqslant \lambda_{\min}\{\mathbf{M}(\rho)\} \leqslant \lambda_{\max}\{\mathbf{M}(\rho)\} \leqslant C_2$ from the Condition (C8), which implies that $C_2^{-1} \leqslant \lambda_{\max}\left\{\mathbf{S}(\rho_0)^{-1}\mathbf{S}(\rho_0)^{-\mathrm{T}}\right\} \leqslant C_1^{-1}$. Hence, $\mathscr{T}_{2,3a}$ is bounded. For the second term of (S.56), recall that $|\rho| < 1$ and $D_{ii}(\rho) = 1/(1 + \rho^2 \sum_{j=1}^{n} w_{ji}^2) \leqslant 1$ by Remark 2, then we have

$$
\begin{aligned}
&\mathscr{T}_{2,3b} \\
= \;& \operatorname{trace}\left\{\frac{\partial \mathbf{S}(\rho_0)^{\mathrm{T}}}{\partial \rho}\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\mathbf{S}(\rho_0)^{\mathrm{T}}\frac{\partial \mathbf{S}(\rho_0)}{\partial \rho}\right\} \\
= \;& \operatorname{trace}\left\{(-\mathbf{W}^{\mathrm{T}})(\mathbf{I} - \rho_0\mathbf{W})\mathbf{D}^2(\rho_0)\left(\mathbf{I} - \rho_0\mathbf{W} - \rho_0\mathbf{W}^{\mathrm{T}} + \rho_0^2\mathbf{W}^{\mathrm{T}}\mathbf{W}\right)\right. \\
& \left.\times \mathbf{D}^2(\rho_0)(\mathbf{I} - \rho_0\mathbf{W}^{\mathrm{T}})(-\mathbf{W})\right\}
\end{aligned}
$$

$$
\begin{aligned}
&\leqslant \operatorname{trace}\left\{(\mathbf{W}^{\mathrm{T}})(\mathbf{I}+|\rho_0|\mathbf{W})\mathbf{D}^2(\rho_0)\left(\mathbf{I}+|\rho_0|\mathbf{W}+|\rho_0|\mathbf{W}^{\mathrm{T}}+\rho_0^2\mathbf{W}^{\mathrm{T}}\mathbf{W}\right)\right. \\
&\qquad \left.\times\mathbf{D}^2(\rho_0)(\mathbf{I}+|\rho_0|\mathbf{W}^{\mathrm{T}})(\mathbf{W})\right\} \\
&\leqslant \operatorname{trace}\left\{\mathbf{W}^{\mathrm{T}}(\mathbf{I}+\mathbf{W})\left(\mathbf{I}+\mathbf{W}+\mathbf{W}^{\mathrm{T}}+\mathbf{W}^{\mathrm{T}}\mathbf{W}\right)(\mathbf{I}+\mathbf{W}^{\mathrm{T}})\mathbf{W}\right\} \\
&= \mathscr{T}_{2,3b}^{*}+\mathscr{T}_{2,3b}^{**}+\mathscr{T}_{2,3b}^{***}, \qquad \text{(S.57)}
\end{aligned}
$$

where

$$
\begin{aligned}
\mathscr{T}_{2,3b}^{*} &= \operatorname{trace}\left(\mathbf{W}^{\mathrm{T}}\mathbf{W}\right)+\operatorname{trace}\left(\mathbf{W}^{\mathrm{T}}\mathbf{W}\mathbf{W}\right)+\operatorname{trace}\left(\mathbf{W}^{\mathrm{T}}\mathbf{W}^{\mathrm{T}}\mathbf{W}\right) \\
&\quad +\operatorname{trace}\left(\mathbf{W}^{\mathrm{T}}\mathbf{W}^{\mathrm{T}}\mathbf{W}\right)+\operatorname{trace}\left(\mathbf{W}^{\mathrm{T}}\mathbf{W}\mathbf{W}\right),
\end{aligned}
$$

$$
\begin{aligned}
&\mathscr{T}_{2,3b}^{**} \\
&= \operatorname{trace}\left(\mathbf{W}^{\mathrm{T}}\mathbf{W}^{\mathrm{T}}\mathbf{W}\mathbf{W}\right)+\operatorname{trace}\left(\mathbf{W}^{\mathrm{T}}\mathbf{W}\mathbf{W}^{\mathrm{T}}\mathbf{W}\right)+\operatorname{trace}\left(\mathbf{W}^{\mathrm{T}}\mathbf{W}^{\mathrm{T}}\mathbf{W}^{\mathrm{T}}\mathbf{W}\right) \\
&\quad +\operatorname{trace}\left(\mathbf{W}^{\mathrm{T}}\mathbf{W}\mathbf{W}\mathbf{W}\right)+\operatorname{trace}\left(\mathbf{W}^{\mathrm{T}}\mathbf{W}\mathbf{W}^{\mathrm{T}}\mathbf{W}\right)+\operatorname{trace}\left(\mathbf{W}^{\mathrm{T}}\mathbf{W}\mathbf{W}^{\mathrm{T}}\mathbf{W}\right),
\end{aligned}
$$

and

$$
\begin{aligned}
\mathscr{T}_{2,3b}^{***} &= \operatorname{trace}\left(\mathbf{W}^{\mathrm{T}}\mathbf{W}\mathbf{W}^{\mathrm{T}}\mathbf{W}\mathbf{W}\right)+\operatorname{trace}\left(\mathbf{W}^{\mathrm{T}}\mathbf{W}^{\mathrm{T}}\mathbf{W}\mathbf{W}^{\mathrm{T}}\mathbf{W}\right) \\
&\quad +\operatorname{trace}\left(\mathbf{W}^{\mathrm{T}}\mathbf{W}\mathbf{W}\mathbf{W}^{\mathrm{T}}\mathbf{W}\right)+\operatorname{trace}\left(\mathbf{W}^{\mathrm{T}}\mathbf{W}\mathbf{W}^{\mathrm{T}}\mathbf{W}^{\mathrm{T}}\mathbf{W}\right) \\
&\quad +\operatorname{trace}\left(\mathbf{W}^{\mathrm{T}}\mathbf{W}\mathbf{W}^{\mathrm{T}}\mathbf{W}\mathbf{W}^{\mathrm{T}}\mathbf{W}\right). \qquad \text{(S.58)}
\end{aligned}
$$

Below, we first derive the orders of first and last terms of (S.58) since these terms have five and six $\mathbf{W}$'s. For the first term of (S.58), by Cauchy-Schwarz inequality, Lemma S2.8, it follows that

$$
\begin{aligned}
& \operatorname{trace}\left\{\left(\mathbf{W}^{\mathrm{T}}\mathbf{W}\right)\left(\mathbf{W}^{\mathrm{T}}\mathbf{W}\mathbf{W}\right)\right\} \\
\leqslant\ & \left\{\operatorname{trace}\left(\mathbf{W}^{\mathrm{T}}\mathbf{W}\mathbf{W}^{\mathrm{T}}\mathbf{W}\right)\right\}^{1/2}\left\{\operatorname{trace}\left(\mathbf{W}^{\mathrm{T}}\mathbf{W}\cdot\mathbf{W}\mathbf{W}^{\mathrm{T}}\mathbf{W}^{\mathrm{T}}\mathbf{W}\right)\right\}^{1/2} \\
=\ & \left\{\operatorname{trace}\left(\mathbf{W}^{\mathrm{T}}\mathbf{W}\mathbf{W}^{\mathrm{T}}\mathbf{W}\right)\right\}^{1/2}\left\{\operatorname{trace}\left(\mathbf{W}\mathbf{W}^{\mathrm{T}}\cdot\mathbf{W}^{\mathrm{T}}\mathbf{W}\mathbf{W}^{\mathrm{T}}\mathbf{W}\right)\right\}^{1/2} \\
\leqslant\ & \left\{\lambda_{\max}(\mathbf{W}^{\mathrm{T}}\mathbf{W})\operatorname{trace}\left(\mathbf{W}^{\mathrm{T}}\mathbf{W}\right)\right\}^{1/2}\left\{\lambda_{\max}(\mathbf{W}\mathbf{W}^{\mathrm{T}})\operatorname{trace}\left(\mathbf{W}^{\mathrm{T}}\mathbf{W}\mathbf{W}^{\mathrm{T}}\mathbf{W}\right)\right\}^{1/2} \\
\leqslant\ & \left\{\lambda_{\max}(\mathbf{W}\mathbf{W}^{\mathrm{T}})\operatorname{trace}\left(\mathbf{W}^{\mathrm{T}}\mathbf{W}\right)\right\}^{1/2}\left\{\lambda_{\max}(\mathbf{W}\mathbf{W}^{\mathrm{T}})\lambda_{\max}(\mathbf{W}\mathbf{W}^{\mathrm{T}})\operatorname{trace}\left(\mathbf{W}^{\mathrm{T}}\mathbf{W}\right)\right\}^{1/2} \\
=\ & \left\{\lambda_{\max}(\mathbf{W}\mathbf{W}^{\mathrm{T}})\right\}^{3/2}\operatorname{trace}\left(\mathbf{W}^{\mathrm{T}}\mathbf{W}\right) \\
=\ & O(n),
\end{aligned}
$$

where the last equality is due to the Condition (C8) and $\sum_{\ell=1}^{n} w_{i\ell} = 1$ for any $1 \leqslant i \leqslant n$. Analogously, for the last term of (S.58), it follows that

$$
\begin{aligned}
\operatorname{trace}\left(\mathbf{W}^{\mathrm{T}}\mathbf{W}\mathbf{W}^{\mathrm{T}}\mathbf{W}\mathbf{W}^{\mathrm{T}}\mathbf{W}\right) &\leqslant \lambda_{\max}(\mathbf{W}^{\mathrm{T}}\mathbf{W})\operatorname{trace}\left(\mathbf{W}^{\mathrm{T}}\mathbf{W}\mathbf{W}^{\mathrm{T}}\mathbf{W}\right) \\
&\leqslant \left\{\lambda_{\max}(\mathbf{W}^{\mathrm{T}}\mathbf{W})\right\}^{2}\operatorname{trace}\left(\mathbf{W}^{\mathrm{T}}\mathbf{W}\right) \\
&= O(n).
\end{aligned}
$$

The other three terms of (S.58) are of the same order as the first term, so they can be dealt with similarly. Hence, we have $\mathscr{T}_{2,3b}^{***} = O(n)$. For the terms $\mathscr{T}_{2,3b}^{*}$ and $\mathscr{T}_{2,3b}^{**}$ in (S.57), similar to the above illustration, we know immediately that $\mathscr{T}_{2,3b}^{*} = O(n)$ and $\mathscr{T}_{2,3b}^{**} = O(n)$. Therefore, combining (S.55), (S.56), (S.57) and the above results, we conclude that there exists some constant κ_2 such that

$$
\frac{1}{n}\operatorname{trace}(\mathbf{M}_n\mathbf{M}_n^{\mathrm{T}}) \leqslant \kappa_2.
$$

It is easy to notice that $\mathrm{trace}\left\{\mathrm{diag}^2(\mathbf{M}_n)\right\} \leqslant \mathrm{trace}\left(\mathbf{M}_n\mathbf{M}_n^{\mathrm{T}}\right)$ by which we also have

$$\frac{1}{n}\mathrm{trace}\left\{\mathrm{diag}^2(\mathbf{M}_n)\right\} \leqslant \kappa_2.$$

Below, we calculate $\mathbf{U}_n^{\mathrm{T}}\mathbf{U}_n$. By (S.53), we have

$$\begin{aligned}
&\mathbf{U}_n^{\mathrm{T}}\mathbf{U}_n \\
\leqslant\ & 2\mathbf{b}(\rho_0)^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{S}(\rho_0)^{-\mathrm{T}}\frac{\partial\mathbf{S}(\rho_0)^{\mathrm{T}}}{\partial\rho}\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\mathbf{S}(\rho_0)^{\mathrm{T}}\frac{\partial\mathbf{S}(\rho_0)}{\partial\rho}\mathbf{S}(\rho_0)^{-1}\mathbf{A}\mathbf{b}(\rho_0) \\
&+2\frac{\partial\mathbf{b}(\rho_0)^{\mathrm{T}}}{\partial\rho}\mathbf{A}^{\mathrm{T}}\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{A}\frac{\partial\mathbf{b}(\rho_0)}{\partial\rho} \\
\leqslant\ & 2\lambda_{\max}\left\{\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{S}(\rho_0)\right\}\lambda_{\max}\left\{\mathbf{D}^4(\rho_0)\right\}\lambda_{\max}\left\{\mathbf{S}(\rho_0)\mathbf{S}(\rho_0)^{\mathrm{T}}\right\}\lambda_{\max}\left\{\frac{\partial\mathbf{S}(\rho_0)^{\mathrm{T}}}{\partial\rho}\frac{\partial\mathbf{S}(\rho_0)}{\partial\rho}\right\} \\
&\quad\times\lambda_{\max}\left\{\mathbf{S}(\rho_0)^{-\mathrm{T}}\mathbf{S}(\rho_0)^{-1}\right\}\lambda_{\max}\left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)\|\mathbf{b}(\rho_0)\|^2 \\
&+2\lambda_{\max}\left\{\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{S}(\rho_0)\right\}\lambda_{\max}\left\{\mathbf{D}^4(\rho_0)\right\}\lambda_{\max}\left\{\mathbf{S}(\rho_0)\mathbf{S}(\rho_0)^{\mathrm{T}}\right\}\lambda_{\max}\left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)\left\|\frac{\partial\mathbf{b}(\rho_0)}{\partial\rho}\right\|^2 \\
=\ & O(n),
\end{aligned}$$

where the last equality is due to the Conditions (C7) and (C8). Following the previous results, by (S.51) and Lemma S2.9, we have

$$\begin{aligned}
&n^{-1/2}E\left[n^{-1/2}\sum_{i=1}^{n}S_i\{\mathbf{A};\mathbf{b}(\rho_0),\rho_0\}\right]^2 \\
=\ & \frac{1}{\sqrt{n}}\left[\frac{\sigma^4}{n}\left\{\mathrm{trace}(\mathbf{M}_n\mathbf{M}_n^{\mathrm{T}})+\mathrm{trace}\left(\mathbf{M}_n^2\right)\right\}\right. \\
&\left.+\frac{1}{n}\mathrm{trace}\left\{\mathrm{diag}^2(\mathbf{M}_n)\right\}\left\{E(\varepsilon_1^4)-3\sigma^4\right\}+\frac{\sigma^2}{n}\mathbf{U}_n^{\mathrm{T}}\mathbf{U}_n\right] \\
=\ & O\left(\frac{1}{\sqrt{n}}\right). \qquad\qquad (\mathrm{S}.59)
\end{aligned}$$

We further obtain

$$
\begin{aligned}
& \operatorname{var}\left(n^{-3/4}\sum_{i=1}^{n}\left[S_i\{\widehat{\mathbf{A}};\mathbf{b}(\rho_0),\rho_0\}+\sum_{k=1}^{K}\frac{\partial S_i\{\widehat{\mathbf{A}};\mathbf{b}^*(\rho_0),\rho_0\}}{\partial b_k^*(\rho_0)}\left\{\widehat{b}_k(\rho_0)-b_k(\rho_0)\right\}\right]\right) \\
\leqslant\ & E\left(n^{-3/4}\sum_{i=1}^{n}\left[S_i\{\widehat{\mathbf{A}};\mathbf{b}(\rho_0),\rho_0\}+\sum_{k=1}^{K}\frac{\partial S_i\{\widehat{\mathbf{A}};\mathbf{b}^*(\rho_0),\rho_0\}}{\partial b_k^*(\rho_0)}\left\{\widehat{b}_k(\rho_0)-b_k(\rho_0)\right\}\right]\right)^2 \\
=\ & E\left[n^{-3/4}\sum_{i=1}^{n}S_i\{\mathbf{A};\mathbf{b}(\rho_0),\rho_0\}+\sum_{k=1}^{K}\mathscr{B}_k(\mathbf{A})\left(n^{-3/4}\sum_{j=1}^{n}\varphi_{jk}\right)+o_p(1)\right]^2 \\
\leqslant\ & 2E\left[n^{-3/4}\sum_{i=1}^{n}S_i\{\mathbf{A};\mathbf{b}(\rho_0),\rho_0\}\right]^2+2E\left\{\sum_{k=1}^{K}\mathscr{B}_k(\mathbf{A})\left(n^{-3/4}\sum_{j=1}^{n}\varphi_{jk}\right)\right\}^2+o(1) \\
=\ & 2n^{-1/2}E\left(E\left[\left\{\sum_{k=1}^{K}\mathscr{B}_k(\mathbf{A})\varphi_{jk}\right\}^2\mid\mathbf{A}\right]\right)+O\left(\frac{1}{\sqrt{n}}\right)+o(1) \\
=\ & 2n^{-1/2}E\left[\left\{\sum_{k=1}^{K}\mathscr{B}_k(\mathbf{A})\varphi_{jk}\right\}^2\right]+o(1) \\
\equiv\ & 2\Sigma_1+o(1),
\end{aligned}
\tag{S.60}
$$

where the first equality is by (S.45), the second equality is by (S.59) and (S.49). Hence, by (S.48) and (S.60), we get the expectation and variance of the first term of (S.32).

Consequently, we rewrite (S.32) as follows

$$
\begin{aligned}
& n^{1/4}\left(\widehat{\rho}-\rho_0\right) \\
=\ & n^{-3/4}\sum_{i=1}^{n}\left[S_i\{\widehat{\mathbf{A}};\mathbf{b}(\rho_0),\rho_0\}+\sum_{k=1}^{K}\frac{\partial S_i\{\widehat{\mathbf{A}};\mathbf{b}^*(\rho_0),\rho_0\}}{\partial b_k^*(\rho_0)}\right. \\
& \left.\times\left\{\widehat{b}_k(\rho_0)-b_k(\rho_0)\right\}\right]\Big/\left[\frac{1}{n}\sum_{i=1}^{n}\frac{\mathrm{d}S_i\{\widehat{\mathbf{A}};\widehat{\mathbf{b}}(\rho^*),\rho^*\}}{\mathrm{d}\rho^*}\right] \\
=\ & \left[n^{-3/4}\sum_{i=1}^{n}S_i\{\mathbf{A};\mathbf{b}(\rho_0),\rho_0\}+\sum_{k=1}^{K}\mathscr{B}_k(\mathbf{A})\left(n^{-3/4}\sum_{j=1}^{n}\varphi_{jk}\right)\right. \\
& \left.+o_p(1)\right]\Big/\left[\frac{1}{n}\sum_{i=1}^{n}\frac{\mathrm{d}S_i\{\widehat{\mathbf{A}};\widehat{\mathbf{b}}(\rho^*),\rho^*\}}{\mathrm{d}\rho^*}\right]
\end{aligned}
$$

$$
\begin{aligned}
&= \left\{ n^{-3/4} \sum_{i=1}^{n} \sum_{k=1}^{K} \mathscr{B}_k(\mathbf{A}) \varphi_{ik} + O_p\left(n^{-1/4}\right) + o_p(1) \right\} \Big/ \left[\frac{1}{n} \sum_{i=1}^{n} \frac{\mathrm{d} S_i\{\widehat{\mathbf{A}}; \widehat{\mathbf{b}}(\rho^*), \rho^*\}}{\mathrm{d}\rho^*} \right] \\
&= \left\{ n^{-3/4} \sum_{i=1}^{n} \sum_{k=1}^{K} \mathscr{B}_k(\mathbf{A}) \varphi_{ik} + o_p(1) \right\} \Big/ \left[\frac{1}{n} \sum_{i=1}^{n} \frac{\mathrm{d} S_i\{\widehat{\mathbf{A}}; \widehat{\mathbf{b}}(\rho^*), \rho^*\}}{\mathrm{d}\rho^*} \right], \qquad \text{(S.61)}
\end{aligned}
$$

where the second equality is by (S.45) and the last second equality is due to (S.47), (S.50) and (S.59).

We start with the denominator of (S.61). We will show that, for given $\mathbf{A}$,

$$
\frac{1}{n} \sum_{i=1}^{n} \frac{\mathrm{d} S_i\{\widehat{\mathbf{A}}; \widehat{\mathbf{b}}(\rho^*), \rho^*\}}{\mathrm{d}\rho^*} \mid \mathbf{A} \xrightarrow{\mathrm{Pr}} \text{some constant.} \qquad \text{(S.62)}
$$

Note that ρ^* denotes the value between the ρ_0 and $\widehat{\rho}$, and $\mathbf{b}^*(\rho_0)$ between $\mathbf{b}(\rho_0)$ and $\widehat{\mathbf{b}}(\rho_0)$. By (S.4), with $\mathbf{b}(\rho)$ and ρ replaced by $\widehat{\mathbf{b}}(\rho^*)$ and ρ^*, respectively and Taylor's expansion, we have

$$
\begin{aligned}
& \frac{1}{n} \sum_{i=1}^{n} \frac{\mathrm{d} S_i\{\widehat{\mathbf{A}}; \widehat{\mathbf{b}}(\rho^*), \rho^*\}}{\mathrm{d}\rho^*} \\
=\ & \frac{1}{n} \sum_{i=1}^{n} \frac{\mathrm{d} S_i\{\mathbf{A}; \widehat{\mathbf{b}}(\rho^*), \rho^*\}}{\mathrm{d}\rho^*} + o_p(1) \\
=\ & \frac{1}{n} \sum_{i=1}^{n} \frac{\mathrm{d} S_i\{\mathbf{A}; \widehat{\mathbf{b}}(\rho_0), \rho_0\}}{\mathrm{d}\rho_0} + \frac{1}{n} \sum_{i=1}^{n} \frac{\mathrm{d}^2 S_i\{\mathbf{A}; \widehat{\mathbf{b}}(\rho_0), \rho_0\}}{\mathrm{d}\rho_0^2} (\rho^* - \rho_0) + o_p(1),
\end{aligned}
$$

where the last equality is because $\widehat{\rho}$ is $n^{1/4}$-consistent. We now handle the first two terms of the above separately. Note that $\|\widehat{\mathbf{b}}(\rho_0) - \mathbf{b}(\rho_0)\| = O_p(K^{\alpha+3/2} n^{-1/2})$ implied by $|\widehat{b}_k(\rho_0) - b_k(\rho_0)| = O_p(k^{\alpha+1} n^{-1/2})$ from (S.33). Together with the Conditions (C5) and (C6), we have

$$
\begin{aligned}
& \frac{1}{n} \sum_{i=1}^{n} \frac{\mathrm{d} S_i\{\mathbf{A}; \widehat{\mathbf{b}}(\rho_0), \rho_0\}}{\mathrm{d}\rho_0} \\
=\ & \frac{1}{n} \sum_{i=1}^{n} \frac{\mathrm{d} S_i\{\mathbf{A}; \mathbf{b}(\rho_0), \rho_0\}}{\mathrm{d}\rho_0} + \left\{ \partial \left[\frac{1}{n} \sum_{i=1}^{n} \frac{\mathrm{d} S_i\{\mathbf{A}; \mathbf{b}(\rho_0), \rho_0\}}{\mathrm{d}\rho_0} \right] \Big/ \partial \mathbf{b}(\rho_0)^{\mathrm{T}} \right\} \left\{ \widehat{\mathbf{b}}(\rho_0) - \mathbf{b}(\rho_0) \right\}
\end{aligned}
$$

$$
\begin{aligned}
& +o_p\left(\left\|\widehat{\mathbf{b}}(\rho_0)-\mathbf{b}(\rho_0)\right\|\right) \\
= \ & \frac{1}{n}\sum_{i=1}^{n}\frac{\mathrm{d}S_i\{\mathbf{A};\mathbf{b}(\rho_0),\rho_0\}}{\mathrm{d}\rho_0}+O_p\left(\left\|\frac{1}{n}\sum_{i=1}^{n}\left(\partial\left[\frac{\mathrm{d}S_i\{\mathbf{A};\mathbf{b}(\rho_0),\rho_0\}}{\mathrm{d}\rho_0}\right]\Big/\partial\mathbf{b}(\rho_0)^{\mathrm{T}}\right)\right\|\cdot\left\|\widehat{\mathbf{b}}(\rho_0)-\mathbf{b}(\rho_0)\right\|\right) \\
& +o_p\left(\left\|\widehat{\mathbf{b}}(\rho_0)-\mathbf{b}(\rho_0)\right\|\right) \\
= \ & \frac{1}{n}\sum_{i=1}^{n}\frac{\mathrm{d}S_i\{\mathbf{A};\mathbf{b}(\rho_0),\rho_0\}}{\mathrm{d}\rho_0}+O_p(K^{\alpha+3/2}n^{-1/2})+o_p(K^{\alpha+3/2}n^{-1/2}) \\
= \ & \frac{1}{n}\sum_{i=1}^{n}\frac{\mathrm{d}S_i\{\mathbf{A};\mathbf{b}(\rho_0),\rho_0\}}{\mathrm{d}\rho_0}+o_p(1), \qquad \text{(S.63)}
\end{aligned}
$$

where the inequality is due to Cauchy-Schwarz inequality. Similarly, we have

$$
\begin{aligned}
& \frac{1}{n}\sum_{i=1}^{n}\frac{\mathrm{d}^2S_i\{\mathbf{A};\widehat{\mathbf{b}}(\rho_0),\rho_0\}}{\mathrm{d}\rho_0^2}(\rho^*-\rho_0) \\
= \ & (\rho^*-\rho_0)\left[\frac{1}{n}\sum_{i=1}^{n}\frac{\mathrm{d}^2S_i\{\mathbf{A};\mathbf{b}(\rho_0),\rho_0\}}{\mathrm{d}\rho_0^2}\right. \\
& \left.+\left\{\frac{1}{n}\sum_{i=1}^{n}\partial\left[\frac{\mathrm{d}^2S_i\{\mathbf{A};\mathbf{b}(\rho_0),\rho_0\}}{\mathrm{d}\rho_0^2}\right]\Big/\partial\mathbf{b}(\rho_0)^{\mathrm{T}}\right\}\left\{\widehat{\mathbf{b}}(\rho_0)-\mathbf{b}(\rho_0)\right\}+o_p\left(\left\|\widehat{\mathbf{b}}(\rho_0)-\mathbf{b}(\rho_0)\right\|\right)\right] \\
= \ & O_p(n^{-1/4})\left\{O_p(1)+O_p(K^{\alpha+3/2}n^{-1/2})+o_p(K^{\alpha+3/2}n^{-1/2})\right\} \\
= \ & o_p(1),
\end{aligned}
$$

Hence, based on above results, we have

$$
\frac{1}{n}\sum_{i=1}^{n}\frac{\mathrm{d}S_i\{\widehat{\mathbf{A}};\widehat{\mathbf{b}}(\rho^*),\rho^*\}}{\mathrm{d}\rho^*} = \frac{1}{n}\sum_{i=1}^{n}\frac{\mathrm{d}S_i\{\mathbf{A};\mathbf{b}(\rho_0),\rho_0\}}{\mathrm{d}\rho_0}+o_p(1). \qquad \text{(S.64)}
$$

We next show that

$$
\frac{1}{n}\sum_{i=1}^{n}\frac{\mathrm{d}S_i\{\mathbf{A};\mathbf{b}(\rho_0),\rho_0\}}{\mathrm{d}\rho_0}\mid\mathbf{A}\ \xrightarrow{\mathrm{Pr}}\ \text{some constant}. \qquad \text{(S.65)}
$$

Similar to the derivative of (S.51), we have

$$
\begin{aligned}
& \frac{1}{\sqrt{n}}\sum_{i=1}^{n} S_i\{\mathbf{A};\mathbf{b}(\rho_0),\rho_0\} \\
= \ & \frac{1}{\sqrt{n}}\left\{\mathbf{D}(\rho_0)\mathbf{M}(\rho_0)\mathbf{Y}-\mathbf{D}(\rho_0)\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{A}\mathbf{b}(\rho_0)\right\}^{\mathrm{T}} \\
& \times\left[\frac{\partial\mathbf{D}(\rho_0)}{\partial\rho}\mathbf{S}(\rho_0)^{\mathrm{T}}\left\{\mathbf{S}(\rho_0)\mathbf{Y}-\mathbf{A}\mathbf{b}(\rho_0)\right\}+\mathbf{D}(\rho)\frac{\partial\mathbf{S}(\rho_0)^{\mathrm{T}}}{\partial\rho}\left\{\mathbf{S}(\rho_0)\mathbf{Y}-\mathbf{A}\mathbf{b}(\rho_0)\right\}\right. \\
& +\mathbf{D}(\rho_0)\mathbf{S}(\rho_0)^{\mathrm{T}}\frac{\partial\mathbf{S}(\rho_0)}{\partial\rho}\mathbf{S}(\rho_0)^{-1}\left\{\mathbf{S}(\rho_0)\mathbf{Y}-\mathbf{A}\mathbf{b}(\rho_0)\right\} \\
& \left.+\mathbf{D}(\rho_0)\mathbf{S}(\rho_0)^{\mathrm{T}}\frac{\partial\mathbf{S}(\rho_0)}{\partial\rho}\mathbf{S}(\rho_0)^{-1}\mathbf{A}\mathbf{b}(\rho_0)-\mathbf{D}(\rho_0)\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{A}\frac{\partial\mathbf{b}(\rho_0)}{\partial\rho}\right],
\end{aligned}
$$

from which we obtain that

$$
\frac{1}{n}\sum_{i=1}^{n}\frac{\mathrm{d}S_i\{\mathbf{A};\mathbf{b}(\rho_0),\rho_0\}}{\mathrm{d}\rho_0} = \frac{1}{n}\dot{\mathbf{F}}^{\mathrm{T}}\dot{\mathbf{F}}+\frac{1}{n}\mathbf{F}^{\mathrm{T}}\ddot{\mathbf{F}}, \tag{S.66}
$$

where

$$
\mathbf{F} = \mathbf{D}(\rho_0)\mathbf{S}(\rho_0)^{\mathrm{T}}\boldsymbol{\varepsilon}_n,
$$

$$
\begin{aligned}
\dot{\mathbf{F}} = & \left[\frac{\partial\mathbf{D}(\rho_0)}{\partial\rho}\mathbf{S}(\rho_0)^{\mathrm{T}}\boldsymbol{\varepsilon}_n+\mathbf{D}(\rho)\frac{\partial\mathbf{S}(\rho_0)^{\mathrm{T}}}{\partial\rho}\boldsymbol{\varepsilon}_n+\mathbf{D}(\rho_0)\mathbf{S}(\rho_0)^{\mathrm{T}}\frac{\partial\mathbf{S}(\rho_0)}{\partial\rho}\mathbf{S}(\rho_0)^{-1}\boldsymbol{\varepsilon}_n\right. \\
& \left.+\mathbf{D}(\rho_0)\mathbf{S}(\rho_0)^{\mathrm{T}}\frac{\partial\mathbf{S}(\rho_0)}{\partial\rho}\mathbf{S}(\rho_0)^{-1}\mathbf{A}\mathbf{b}(\rho_0)-\mathbf{D}(\rho_0)\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{A}\frac{\partial\mathbf{b}(\rho_0)}{\partial\rho}\right],
\end{aligned}
$$

$$
\begin{aligned}
\ddot{\mathbf{F}} = & \frac{\partial^2\mathbf{D}(\rho_0)}{\partial\rho^2}\mathbf{S}(\rho_0)^{\mathrm{T}}\boldsymbol{\varepsilon}_n \\
& +2\frac{\partial\mathbf{D}(\rho_0)}{\partial\rho}\left\{\frac{\partial\mathbf{S}(\rho_0)^{\mathrm{T}}}{\partial\rho}\boldsymbol{\varepsilon}_n+\mathbf{S}(\rho_0)^{\mathrm{T}}\frac{\partial\mathbf{S}(\rho_0)}{\partial\rho}\mathbf{Y}\right\}+\mathbf{D}(\rho_0)\frac{\partial^2\mathbf{M}(\rho_0)}{\partial\rho^2}\mathbf{Y} \\
& -2\frac{\partial\mathbf{D}(\rho_0)}{\partial\rho}\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{A}\frac{\partial\mathbf{b}(\rho_0)}{\partial\rho}-2\mathbf{D}(\rho_0)\frac{\partial\mathbf{S}(\rho_0)^{\mathrm{T}}}{\partial\rho}\mathbf{A}\frac{\partial\mathbf{b}(\rho_0)}{\partial\rho},
\end{aligned}
$$

$$\boldsymbol{\varepsilon}_n \quad = \quad \mathbf{S}(\rho_0)\mathbf{Y} - \mathbf{A}\mathbf{b}(\rho_0),$$

and $\boldsymbol{\varepsilon}_n \mid \mathbf{A} \sim \mathcal{N}(0, \sigma^2\mathbf{I}_n)$. To establish (S.65), it suffices to show that

$$\frac{1}{n}\dot{\mathbf{F}}^{\mathrm{T}}\dot{\mathbf{F}} \mid \mathbf{A} \quad \overset{\mathrm{Pr}}{\to} \quad c_*, \tag{S.67}$$

and

$$\frac{1}{n}\mathbf{F}^{\mathrm{T}}\ddot{\mathbf{F}} \mid \mathbf{A} \quad \overset{\mathrm{Pr}}{\to} \quad 0, \tag{S.68}$$

where c_* is some positive constant. We first prove (S.67). Let

$$\dot{\mathbf{F}} \quad = \quad \mathbf{G}\boldsymbol{\varepsilon}_n + \mathbf{h}(\mathbf{A}),$$

where

$$\mathbf{G} \quad = \quad \frac{\partial \mathbf{D}(\rho_0)}{\partial \rho}\mathbf{S}(\rho_0)^{\mathrm{T}} + \mathbf{D}(\rho)\frac{\partial \mathbf{S}(\rho_0)^{\mathrm{T}}}{\partial \rho} + \mathbf{D}(\rho_0)\mathbf{S}(\rho_0)^{\mathrm{T}}\frac{\partial \mathbf{S}(\rho_0)}{\partial \rho}\mathbf{S}(\rho_0)^{-1},$$

and

$$\mathbf{h}(\mathbf{A}) \quad = \quad \mathbf{D}(\rho_0)\mathbf{S}(\rho_0)^{\mathrm{T}}\frac{\partial \mathbf{S}(\rho_0)}{\partial \rho}\mathbf{S}(\rho_0)^{-1}\mathbf{A}\mathbf{b}(\rho_0) - \mathbf{D}(\rho_0)\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{A}\frac{\partial \mathbf{b}(\rho_0)}{\partial \rho}.$$

Then, by Lemma S2.9, we have

$$\begin{aligned} & \mathrm{var}\left(\frac{1}{n}\dot{\mathbf{F}}^{\mathrm{T}}\dot{\mathbf{F}} \mid \mathbf{A}\right) \\ = \quad & \frac{1}{n^2}\mathrm{var}\left\{\boldsymbol{\varepsilon}_n^{\mathrm{T}}\mathbf{G}^{\mathrm{T}}\mathbf{G}\boldsymbol{\varepsilon}_n + \boldsymbol{\varepsilon}_n^{\mathrm{T}}\mathbf{G}^{\mathrm{T}}\mathbf{h}(\mathbf{A}) + \mathbf{h}(\mathbf{A})^{\mathrm{T}}\mathbf{G}\boldsymbol{\varepsilon}_n + \mathbf{h}(\mathbf{A})^{\mathrm{T}}\mathbf{h}(\mathbf{A}) \mid \mathbf{A}\right\} \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{n^2}\mathrm{var}\left\{\boldsymbol{\varepsilon}_n^{\mathrm{T}}\mathbf{G}^{\mathrm{T}}\mathbf{G}\boldsymbol{\varepsilon}_n + 2\mathbf{h}(\mathbf{A})^{\mathrm{T}}\mathbf{G}\boldsymbol{\varepsilon}_n \mid \mathbf{A}\right\} \\
&= \frac{1}{n^2}2\sigma^4\mathrm{trace}(\mathbf{G}^{\mathrm{T}}\mathbf{G}\mathbf{G}^{\mathrm{T}}\mathbf{G}) + \frac{1}{n^2}\mathrm{trace}\left[\{\mathrm{diag}(\mathbf{G}^{\mathrm{T}}\mathbf{G})\}^2\right]\left\{E(\varepsilon_1^4) - 3\sigma^4\right\} \\
&\quad + \frac{1}{n^2}4\sigma^2\mathbf{h}(\mathbf{A})^{\mathrm{T}}\mathbf{G}\mathbf{G}^{\mathrm{T}}\mathbf{h}(\mathbf{A}). \qquad \text{(S.69)}
\end{aligned}$$

We now check each term of (S.69). For the first term, by Lemma S2.8(a), we note that

$$\frac{1}{n^2}2\sigma^4\mathrm{trace}(\mathbf{G}^{\mathrm{T}}\mathbf{G}\mathbf{G}^{\mathrm{T}}\mathbf{G}) \leqslant \frac{1}{n^2}2\sigma^4\mathrm{trace}(\mathbf{G}^{\mathrm{T}}\mathbf{G})\lambda_{\max}(\mathbf{G}^{\mathrm{T}}\mathbf{G}).$$

Since for any matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ with compatible dimensions, we have

$$\lambda_{\max}\{(\mathbf{A}+\mathbf{B}+\mathbf{C})^{\mathrm{T}}(\mathbf{A}+\mathbf{B}+\mathbf{C})\} \leqslant 3\lambda_{\max}(\mathbf{A}^{\mathrm{T}}\mathbf{A}) + 3\lambda_{\max}(\mathbf{B}^{\mathrm{T}}\mathbf{B}) + 3\lambda_{\max}(\mathbf{C}^{\mathrm{T}}\mathbf{C}),$$

from which, by the Condition (C8) and Remark 2, we have

$$\begin{aligned}
&\lambda_{\max}\left(\mathbf{G}^{\mathrm{T}}\mathbf{G}\right) \\
= \;& \lambda_{\max}\left[\left\{\frac{\partial\mathbf{D}(\rho_0)}{\partial\rho}\mathbf{S}(\rho_0)^{\mathrm{T}} + \mathbf{D}(\rho)\frac{\partial\mathbf{S}(\rho_0)^{\mathrm{T}}}{\partial\rho} + \mathbf{D}(\rho_0)\mathbf{S}(\rho_0)^{\mathrm{T}}\frac{\partial\mathbf{S}(\rho_0)}{\partial\rho}\mathbf{S}(\rho_0)^{-1}\right\}^{\mathrm{T}}\right. \\
&\left.\times\left\{\frac{\partial\mathbf{D}(\rho_0)}{\partial\rho}\mathbf{S}(\rho_0)^{\mathrm{T}} + \mathbf{D}(\rho)\frac{\partial\mathbf{S}(\rho_0)^{\mathrm{T}}}{\partial\rho} + \mathbf{D}(\rho_0)\mathbf{S}(\rho_0)^{\mathrm{T}}\frac{\partial\mathbf{S}(\rho_0)}{\partial\rho}\mathbf{S}(\rho_0)^{-1}\right\}\right] \\
\leqslant \;& 3\lambda_{\max}\left[\mathbf{S}(\rho_0)\left\{\frac{\partial\mathbf{D}(\rho_0)}{\partial\rho}\right\}^2\mathbf{S}(\rho_0)^{\mathrm{T}}\right] + 3\lambda_{\max}\left\{\frac{\partial\mathbf{S}(\rho_0)}{\partial\rho}\mathbf{D}(\rho)^2\frac{\partial\mathbf{S}(\rho_0)^{\mathrm{T}}}{\partial\rho}\right\} \\
&+3\lambda_{\max}\left\{\mathbf{S}(\rho_0)^{-\mathrm{T}}\frac{\partial\mathbf{S}(\rho_0)^{\mathrm{T}}}{\partial\rho}\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\mathbf{S}(\rho_0)^{\mathrm{T}}\frac{\partial\mathbf{S}(\rho_0)}{\partial\rho}\mathbf{S}(\rho_0)^{-1}\right\} \\
\leqslant \;& 3\lambda_{\max}\left[\left\{\frac{\partial\mathbf{D}(\rho_0)}{\partial\rho}\right\}^2\right]\lambda_{\max}\left\{\mathbf{S}(\rho_0)\mathbf{S}(\rho_0)^{\mathrm{T}}\right\} + 3\lambda_{\max}\left\{\mathbf{D}(\rho)^2\right\}\lambda_{\max}\left(\mathbf{W}\mathbf{W}^{\mathrm{T}}\right) \\
&+3\lambda_{\max}\left\{\mathbf{D}(\rho_0)^2\right\}\lambda_{\max}\left\{\mathbf{S}(\rho_0)\mathbf{S}(\rho_0)^{\mathrm{T}}\right\}\lambda_{\max}\left(\mathbf{W}^{\mathrm{T}}\mathbf{W}\right)\left[\lambda_{\min}\left\{\mathbf{S}(\rho_0)\mathbf{S}(\rho_0)^{\mathrm{T}}\right\}\right]^{-1} \\
\leqslant \;& C,
\end{aligned}$$

for some constant C, where the last equality is because the Conditions (C7) (C8) and Remark 2, and the second inequality is based on the fact that $\lambda_{\max}\left(\mathbf{B}^{\mathrm{T}}\mathbf{A}\mathbf{B}\right) \leqslant \lambda_{\max}(\mathbf{A})\lambda_{\max}\left(\mathbf{B}^{\mathrm{T}}\mathbf{B}\right)$, where $\mathbf{A}$ and $\mathbf{B}$ are two matrices, and $\mathbf{A}$ is positive semidefinite matrix.

Similarly,

$$
\begin{aligned}
& \frac{1}{n^2}\mathrm{trace}\left(\mathbf{G}^{\mathrm{T}}\mathbf{G}\right) \\
= \; & \frac{1}{n^2}\mathrm{trace}\left[\left\{\frac{\partial \mathbf{D}(\rho_0)}{\partial\rho}\mathbf{S}(\rho_0)^{\mathrm{T}} + \mathbf{D}(\rho)\frac{\partial \mathbf{S}(\rho_0)^{\mathrm{T}}}{\partial\rho} + \mathbf{D}(\rho_0)\mathbf{S}(\rho_0)^{\mathrm{T}}\frac{\partial \mathbf{S}(\rho_0)}{\partial\rho}\mathbf{S}(\rho_0)^{-1}\right\}^{\mathrm{T}}\right. \\
& \quad \left.\times\left\{\frac{\partial \mathbf{D}(\rho_0)}{\partial\rho}\mathbf{S}(\rho_0)^{\mathrm{T}} + \mathbf{D}(\rho)\frac{\partial \mathbf{S}(\rho_0)^{\mathrm{T}}}{\partial\rho} + \mathbf{D}(\rho_0)\mathbf{S}(\rho_0)^{\mathrm{T}}\frac{\partial \mathbf{S}(\rho_0)}{\partial\rho}\mathbf{S}(\rho_0)^{-1}\right\}\right] \\
\leqslant \; & 3\frac{1}{n^2}\mathrm{trace}\left\{\mathbf{S}(\rho_0)\frac{\partial \mathbf{D}(\rho_0)}{\partial\rho}\frac{\partial \mathbf{D}(\rho_0)}{\partial\rho}\mathbf{S}(\rho_0)^{\mathrm{T}}\right\} \\
& +3\frac{1}{n^2}\mathrm{trace}\left\{\frac{\partial \mathbf{S}(\rho_0)}{\partial\rho}\mathbf{D}(\rho)^2\frac{\partial \mathbf{S}(\rho_0)^{\mathrm{T}}}{\partial\rho}\right\} \\
& +3\frac{1}{n^2}\mathrm{trace}\left\{\mathbf{S}(\rho_0)^{-\mathrm{T}}\frac{\partial \mathbf{S}(\rho_0)^{\mathrm{T}}}{\partial\rho}\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\mathbf{S}(\rho_0)^{\mathrm{T}}\frac{\partial \mathbf{S}(\rho_0)}{\partial\rho}\mathbf{S}(\rho_0)^{-1}\right\} \\
= \; & 3\frac{1}{n^2}\mathrm{trace}\left[\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{S}(\rho_0)\left\{\frac{\partial \mathbf{D}(\rho_0)}{\partial\rho}\right\}^2\right] \\
& +3\frac{1}{n^2}\mathrm{trace}\left\{\frac{\partial \mathbf{S}(\rho_0)^{\mathrm{T}}}{\partial\rho}\frac{\partial \mathbf{S}(\rho_0)}{\partial\rho}\mathbf{D}(\rho)^2\right\} \\
& +3\frac{1}{n^2}\mathrm{trace}\left\{\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\mathbf{S}(\rho_0)^{\mathrm{T}}\frac{\partial \mathbf{S}(\rho_0)}{\partial\rho}\mathbf{S}(\rho_0)^{-1}\mathbf{S}(\rho_0)^{-\mathrm{T}}\frac{\partial \mathbf{S}(\rho_0)^{\mathrm{T}}}{\partial\rho}\right\} \\
\leqslant \; & 3\frac{1}{n^2}\mathrm{trace}\left\{\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{S}(\rho_0)\right\}\lambda_{\max}\left[\left\{\frac{\partial \mathbf{D}(\rho_0)}{\partial\rho}\right\}^2\right] \\
& +3\frac{1}{n^2}\mathrm{trace}\left\{\frac{\partial \mathbf{S}(\rho_0)^{\mathrm{T}}}{\partial\rho}\frac{\partial \mathbf{S}(\rho_0)}{\partial\rho}\right\}\lambda_{\max}\left\{\mathbf{D}(\rho)^2\right\} \\
& +3\frac{1}{n^2}\lambda_{\max}\left\{\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\mathbf{S}(\rho_0)^{\mathrm{T}}\right\}\mathrm{trace}\left\{\frac{\partial \mathbf{S}(\rho_0)}{\partial\rho}\mathbf{S}(\rho_0)^{-1}\mathbf{S}(\rho_0)^{-\mathrm{T}}\frac{\partial \mathbf{S}(\rho_0)^{\mathrm{T}}}{\partial\rho}\right\} \\
\leqslant \; & 3\frac{1}{n^2}\mathrm{trace}\left\{\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{S}(\rho_0)\right\}\lambda_{\max}\left[\left\{\frac{\partial \mathbf{D}(\rho_0)}{\partial\rho}\right\}^2\right] \\
& +3\frac{1}{n^2}\mathrm{trace}\left(\mathbf{W}^{\mathrm{T}}\mathbf{W}\right)\lambda_{\max}\left\{\mathbf{D}(\rho)^2\right\} \\
& +3\frac{1}{n^2}\lambda_{\max}\left\{\mathbf{D}(\rho_0)^2\right\}\lambda_{\max}\left\{\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{S}(\rho_0)\right\}\mathrm{trace}\left\{\mathbf{S}(\rho_0)^{-1}\mathbf{S}(\rho_0)^{-\mathrm{T}}\frac{\partial \mathbf{S}(\rho_0)^{\mathrm{T}}}{\partial\rho}\frac{\partial \mathbf{S}(\rho_0)}{\partial\rho}\right\}
\end{aligned}
$$

$$
\begin{aligned}
&\leqslant 3\frac{1}{n^2}\text{trace}\left\{\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{S}(\rho_0)\right\}\lambda_{\max}^2\left\{\frac{\partial \mathbf{D}(\rho_0)}{\partial\rho}\right\} \\
&\quad +3\frac{1}{n^2}\text{trace}\left(\mathbf{W}^{\mathrm{T}}\mathbf{W}\right)\lambda_{\max}\left\{\mathbf{D}(\rho)^2\right\} \\
&\quad +3\frac{1}{n^2}\lambda_{\max}\left\{\mathbf{D}(\rho_0)^2\right\}\lambda_{\max}\left\{\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{S}(\rho_0)\right\}\lambda_{\max}\left\{\mathbf{S}(\rho_0)^{-1}\mathbf{S}(\rho_0)^{-\mathrm{T}}\right\}\text{trace}\left(\mathbf{W}^{\mathrm{T}}\mathbf{W}\right).
\end{aligned}
$$

Note that, by $\sum_{i=1}^n w_{ji} = 1$ for any $1 \leqslant i, j \leqslant n$,

$$
\begin{aligned}
\text{trace}\left\{\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{S}(\rho_0)\right\} &\leqslant 2\text{trace}(\mathbf{I}_n) + 2\rho_0^2\text{trace}(\mathbf{W}^{\mathrm{T}}\mathbf{W}) \\
&= 2n + 2\rho_0^2\sum_{j=1}^n\sum_{i=1}^n w_{ij}^2 \\
&\leqslant 2n + 2\rho_0^2\sum_{j=1}^n\sum_{i=1}^n w_{ij} \\
&\leqslant (2+2\rho_0^2)n. \qquad\qquad (\text{S.70})
\end{aligned}
$$

Thus, based on above derivation, we have

$$
\begin{aligned}
&\frac{1}{n^2}\text{trace}\left(\mathbf{G}^{\mathrm{T}}\mathbf{G}\right) \\
&\leqslant 3\frac{1}{n^2}\text{trace}\left\{\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{S}(\rho_0)\right\}\lambda_{\max}^2\left\{\frac{\partial \mathbf{D}(\rho_0)}{\partial\rho}\right\} \\
&\quad +3\frac{1}{n^2}\text{trace}\left(\mathbf{W}^{\mathrm{T}}\mathbf{W}\right)\lambda_{\max}\left\{\mathbf{D}(\rho)^2\right\} \\
&\quad +3\frac{1}{n^2}\lambda_{\max}\left\{\mathbf{D}(\rho_0)^2\right\}\lambda_{\max}\left\{\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{S}(\rho_0)\right\}\lambda_{\max}\left\{\mathbf{S}(\rho_0)^{-1}\mathbf{S}(\rho_0)^{-\mathrm{T}}\right\}\text{trace}\left(\mathbf{W}^{\mathrm{T}}\mathbf{W}\right) \\
&\leqslant \frac{6}{n}(1+\rho_0^2)\lambda_{\max}^2\left\{\frac{\partial \mathbf{D}(\rho_0)}{\partial\rho}\right\} + 3\frac{1}{n}\lambda_{\max}\left\{\mathbf{D}(\rho)^2\right\} \\
&\quad +3\frac{1}{n}\lambda_{\max}\left\{\mathbf{D}(\rho_0)^2\right\}\lambda_{\max}\left\{\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{S}(\rho_0)\right\}/\lambda_{\min}\left\{\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{S}(\rho_0)\right\} \\
&= O\left(\frac{1}{n}\right).
\end{aligned}
$$

Therefore, we have

$$\begin{aligned}\frac{1}{n^2}2\sigma^4\text{trace}(\mathbf{G}^{\mathrm{T}}\mathbf{G}\mathbf{G}^{\mathrm{T}}\mathbf{G}) &\leqslant \frac{1}{n^2}2\sigma^4\text{trace}(\mathbf{G}^{\mathrm{T}}\mathbf{G})\lambda_{\max}(\mathbf{G}^{\mathrm{T}}\mathbf{G}) \\ &= O\left(\frac{1}{n}\right),\end{aligned} \tag{S.71}$$

and

$$\begin{aligned}\frac{1}{n^2}\text{trace}\left[\{\text{diag}(\mathbf{G}^{\mathrm{T}}\mathbf{G})\}^2\right]\left\{E(\varepsilon_1^4)-3\sigma^4\right\} &\leqslant \frac{1}{n^2}\text{trace}(\mathbf{G}^{\mathrm{T}}\mathbf{G}\mathbf{G}^{\mathrm{T}}\mathbf{G})\left\{E(\varepsilon_1^4)-3\sigma^4\right\} \\ &= O\left(\frac{1}{n}\right).\end{aligned} \tag{S.72}$$

For the third term of (S.69), we have

$$\begin{aligned}&4\sigma^2\frac{1}{n^2}\mathbf{h}(\mathbf{A})^{\mathrm{T}}\mathbf{G}\mathbf{G}^{\mathrm{T}}\mathbf{h}(\mathbf{A}) \\ \leqslant\;& 4\sigma^2\frac{1}{n^2}\lambda_{\max}\left(\mathbf{G}\mathbf{G}^{\mathrm{T}}\right)\mathbf{h}(\mathbf{A})^{\mathrm{T}}\mathbf{h}(\mathbf{A}) \\ \leqslant\;& 8\sigma^2\frac{C}{n^2}\times\left\{\mathbf{b}(\rho_0)^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{S}(\rho_0)^{-\mathrm{T}}\frac{\partial\mathbf{S}(\rho_0)^{\mathrm{T}}}{\partial\rho}\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\mathbf{S}(\rho_0)^{\mathrm{T}}\frac{\partial\mathbf{S}(\rho_0)}{\partial\rho}\mathbf{S}(\rho_0)^{-1}\mathbf{A}\mathbf{b}(\rho_0)\right. \\ &\left.+\frac{\partial\mathbf{b}(\rho_0)^{\mathrm{T}}}{\partial\rho}\mathbf{A}^{\mathrm{T}}\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{A}\frac{\partial\mathbf{b}(\rho_0)}{\partial\rho}\right\} \\ \leqslant\;& 8\sigma^2\frac{C}{n}\times\left[\lambda_{\max}^2\{\mathbf{D}(\rho_0)\}\lambda_{\max}\left\{\mathbf{S}(\rho_0)\mathbf{S}(\rho_0)^{\mathrm{T}}\right\}\lambda_{\max}\left(\mathbf{W}^{\mathrm{T}}\mathbf{W}\right)\right. \\ &\times\lambda_{\max}\left\{\mathbf{S}(\rho_0)^{-\mathrm{T}}\mathbf{S}(\rho_0)^{-1}\right\}\lambda_{\max}\left(\frac{1}{n}\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)\|\mathbf{b}(\rho_0)\|^2 \\ &\left.+\lambda_{\max}^2\{\mathbf{D}(\rho_0)\}\lambda_{\max}\left\{\mathbf{S}(\rho_0)\mathbf{S}(\rho_0)^{\mathrm{T}}\right\}\lambda_{\max}\left(\frac{1}{n}\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)\left\|\frac{\partial\mathbf{b}(\rho_0)}{\partial\rho}\right\|^2\right] \\ =\;& O\left(\frac{1}{n}\right).\end{aligned} \tag{S.73}$$

where the last equality is also due to the Condition (C7), (C8) and Remark 2. Combining

(S.69), (S.71),(S.72) and (S.73), we conclude that

$$\begin{aligned} \mathrm{var}\left(\frac{1}{n}\dot{\mathbf{F}}^{\mathrm{T}}\dot{\mathbf{F}} \mid \mathbf{A}\right) &= O\left(\frac{1}{n}\right) \\ &= o(1), \end{aligned}$$

for given $\mathbf{A}$. By the law of large numbers, this implies (S.67). We next prove (S.68).

By the expression of $\mathbf{F}$ and $\ddot{\mathbf{F}}$ and (2.6), we have

$$\begin{aligned} \mathbf{F}^{\mathrm{T}}\ddot{\mathbf{F}} &= \boldsymbol{\varepsilon}_n^{\mathrm{T}}\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)\frac{\partial^2\mathbf{D}(\rho_0)}{\partial\rho^2}\mathbf{S}(\rho_0)^{\mathrm{T}}\boldsymbol{\varepsilon}_n \\ &\quad +2\boldsymbol{\varepsilon}_n^{\mathrm{T}}\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)\frac{\partial\mathbf{D}(\rho_0)}{\partial\rho}\left\{\frac{\partial\mathbf{S}(\rho_0)^{\mathrm{T}}}{\partial\rho}\boldsymbol{\varepsilon}_n + \mathbf{S}(\rho_0)^{\mathrm{T}}\frac{\partial\mathbf{S}(\rho_0)}{\partial\rho}\mathbf{Y}\right\} \\ &\quad +\boldsymbol{\varepsilon}_n^{\mathrm{T}}\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\frac{\partial^2\mathbf{M}(\rho_0)}{\partial\rho^2}\mathbf{Y} \\ &\quad -2\boldsymbol{\varepsilon}_n^{\mathrm{T}}\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)\frac{\partial\mathbf{D}(\rho_0)}{\partial\rho}\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{A}\frac{\partial\mathbf{b}(\rho_0)}{\partial\rho} \\ &\quad -2\boldsymbol{\varepsilon}_n^{\mathrm{T}}\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\frac{\partial\mathbf{S}(\rho_0)^{\mathrm{T}}}{\partial\rho}\mathbf{A}\frac{\partial\mathbf{b}(\rho_0)}{\partial\rho} \\ &= \boldsymbol{\varepsilon}_n^{\mathrm{T}}\left\{\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)\frac{\partial^2\mathbf{D}(\rho_0)}{\partial\rho^2}\mathbf{S}(\rho_0)^{\mathrm{T}} + 2\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)\frac{\partial\mathbf{D}(\rho_0)}{\partial\rho}\frac{\partial\mathbf{S}(\rho_0)^{\mathrm{T}}}{\partial\rho}\right\}\boldsymbol{\varepsilon}_n \\ &\quad +\boldsymbol{\varepsilon}_n^{\mathrm{T}}\left\{2\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)\frac{\partial\mathbf{D}(\rho_0)}{\partial\rho}\mathbf{S}(\rho_0)^{\mathrm{T}}\frac{\partial\mathbf{S}(\rho_0)}{\partial\rho} + \mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\frac{\partial^2\mathbf{M}(\rho_0)}{\partial\rho^2}\right\}\mathbf{Y} \\ &\quad -2\boldsymbol{\varepsilon}_n^{\mathrm{T}}\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)\frac{\partial\mathbf{D}(\rho_0)}{\partial\rho}\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{A}\frac{\partial\mathbf{b}(\rho_0)}{\partial\rho} \\ &\quad -2\boldsymbol{\varepsilon}_n^{\mathrm{T}}\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\frac{\partial\mathbf{S}(\rho_0)^{\mathrm{T}}}{\partial\rho}\mathbf{A}\frac{\partial\mathbf{b}(\rho_0)}{\partial\rho} \\ &= \boldsymbol{\varepsilon}_n^{\mathrm{T}}\mathbf{G}_2\boldsymbol{\varepsilon}_n + \boldsymbol{\varepsilon}_n^{\mathrm{T}}\mathbf{h}_2(\mathbf{A}), \end{aligned} \tag{S.74}$$

where

$$\begin{aligned} \mathbf{G}_2 &= \mathbf{S}(\rho_0)\mathbf{D}(\rho_0)\frac{\partial^2\mathbf{D}(\rho_0)}{\partial\rho^2}\mathbf{S}(\rho_0)^{\mathrm{T}} + 2\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)\frac{\partial\mathbf{D}(\rho_0)}{\partial\rho}\frac{\partial\mathbf{S}(\rho_0)^{\mathrm{T}}}{\partial\rho} \\ &\quad + \left\{2\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)\frac{\partial\mathbf{D}(\rho_0)}{\partial\rho}\mathbf{S}(\rho_0)^{\mathrm{T}}\frac{\partial\mathbf{S}(\rho_0)}{\partial\rho} + \mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\frac{\partial^2\mathbf{M}(\rho_0)}{\partial\rho^2}\right\}\mathbf{S}^{-1}(\rho_0), \end{aligned}$$

and

$$
\begin{aligned}
\mathbf{h}_2(\mathbf{A}) &= \left\{2\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)\frac{\partial \mathbf{D}(\rho_0)}{\partial \rho}\mathbf{S}(\rho_0)^{\mathrm{T}}\frac{\partial \mathbf{S}(\rho_0)}{\partial \rho} + \mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\frac{\partial^2 \mathbf{M}(\rho_0)}{\partial \rho^2}\right\}\mathbf{S}(\rho_0)^{-1}\mathbf{A}\mathbf{b}(\rho_0) \\
&\quad -2\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)\frac{\partial \mathbf{D}(\rho_0)}{\partial \rho}\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{A}\frac{\partial \mathbf{b}(\rho_0)}{\partial \rho} \\
&\quad -2\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\frac{\partial \mathbf{S}(\rho_0)^{\mathrm{T}}}{\partial \rho}\mathbf{A}\frac{\partial \mathbf{b}(\rho_0)}{\partial \rho}.
\end{aligned}
$$

By Lemma S2.9 and Lemma S2.8(c), we have

$$
\begin{aligned}
&\operatorname{var}\left(\frac{1}{n}\mathbf{F}^{\mathrm{T}}\ddot{\mathbf{F}} \mid \mathbf{A}\right) \\
=\;& \frac{1}{n^2}\operatorname{var}\left\{\boldsymbol{\varepsilon}_n^{\mathrm{T}}\mathbf{G}_2\boldsymbol{\varepsilon}_n + \boldsymbol{\varepsilon}_n^{\mathrm{T}}\mathbf{h}_2(\mathbf{A})\right\} \\
=\;& \frac{1}{n^2}\sigma^4\left\{\operatorname{trace}(\mathbf{G}_2\mathbf{G}_2^{\mathrm{T}}) + \operatorname{trace}\left(\mathbf{G}_2^2\right)\right\} + \frac{1}{n^2}\operatorname{trace}\{\operatorname{diag}^2(\mathbf{G}_2)\}\left\{E(\varepsilon_1^4) - 3\sigma^4\right\} \\
&+\frac{1}{n^2}\sigma^2\mathbf{h}_2(\mathbf{A})^{\mathrm{T}}\mathbf{h}_2(\mathbf{A}) \\
\leqslant\;& \frac{2}{n^2}\sigma^4\operatorname{trace}(\mathbf{G}_2\mathbf{G}_2^{\mathrm{T}}) + \frac{1}{n^2}\operatorname{trace}\left[\{\operatorname{diag}(\mathbf{G}_2)\}^2\right]\left\{E(\varepsilon_1^4) - 3\sigma^4\right\} \\
&+\frac{1}{n^2}\sigma^2\mathbf{h}_2(\mathbf{A})^{\mathrm{T}}\mathbf{h}_2(\mathbf{A}) \\
\leqslant\;& \frac{1}{n^2}\operatorname{trace}(\mathbf{G}_2\mathbf{G}_2^{\mathrm{T}})\left\{2\sigma^4 + \left|E(\varepsilon_1^4) - 3\sigma^4\right|\right\} + \frac{1}{n^2}\sigma^2\mathbf{h}_2(\mathbf{A})^{\mathrm{T}}\mathbf{h}_2(\mathbf{A}).
\end{aligned} \tag{S.75}
$$

Note that

$$
\begin{aligned}
&\frac{1}{n^2}\operatorname{trace}(\mathbf{G}_2\mathbf{G}_2^{\mathrm{T}}) \\
\leqslant\;& 4\frac{1}{n^2}\operatorname{trace}\left\{\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)\frac{\partial^2 \mathbf{D}(\rho_0)}{\partial \rho^2}\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{S}(\rho_0)\frac{\partial^2 \mathbf{D}(\rho_0)}{\partial \rho^2}\mathbf{D}(\rho_0)\mathbf{S}(\rho_0)^{\mathrm{T}}\right\} \\
&+16\frac{1}{n^2}\operatorname{trace}\left\{\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)\frac{\partial \mathbf{D}(\rho_0)}{\partial \rho}\frac{\partial \mathbf{S}(\rho_0)^{\mathrm{T}}}{\partial \rho}\frac{\partial \mathbf{S}(\rho_0)}{\partial \rho}\frac{\partial \mathbf{D}(\rho_0)}{\partial \rho}\mathbf{D}(\rho_0)\mathbf{S}(\rho_0)^{\mathrm{T}}\right\} \\
&+16\frac{1}{n^2}\operatorname{trace}\left\{\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)\frac{\partial \mathbf{D}(\rho_0)}{\partial \rho}\mathbf{S}(\rho_0)^{\mathrm{T}}\frac{\partial \mathbf{S}(\rho_0)}{\partial \rho}\mathbf{S}^{-1}(\rho_0)\right. \\
&\quad\left.\times\mathbf{S}^{-\mathrm{T}}(\rho_0)\frac{\partial \mathbf{S}(\rho_0)^{\mathrm{T}}}{\partial \rho}\mathbf{S}(\rho_0)\frac{\partial \mathbf{D}(\rho_0)}{\partial \rho}\mathbf{D}(\rho_0)\mathbf{S}(\rho_0)^{\mathrm{T}}\right\}
\end{aligned}
$$

$$
\begin{aligned}
& +4\frac{1}{n^2}\text{trace}\left\{\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\frac{\partial^2\mathbf{M}(\rho_0)}{\partial\rho^2}\mathbf{S}^{-1}(\rho_0)\mathbf{S}^{-\mathrm{T}}(\rho_0)\frac{\partial^2\mathbf{M}(\rho_0)}{\partial\rho^2}\mathbf{D}(\rho_0)^2\mathbf{S}(\rho_0)^{\mathrm{T}}\right\} \\
\leqslant\ & 4\frac{1}{n^2}\lambda_{\max}\left\{\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{S}(\rho_0)\right\}\lambda^2_{\max}\left\{\mathbf{D}(\rho_0)\right\}\lambda^2_{\max}\left\{\frac{\partial^2\mathbf{D}(\rho_0)}{\partial\rho^2}\right\}\text{trace}\left\{\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{S}(\rho_0)\right\} \\
& +16\frac{1}{n^2}\lambda_{\max}\left\{\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{S}(\rho_0)\right\}\lambda^2_{\max}\left\{\mathbf{D}(\rho_0)\right\}\lambda^2_{\max}\left\{\frac{\partial\mathbf{D}(\rho_0)}{\partial\rho}\right\}\text{trace}\left(\mathbf{W}^{\mathrm{T}}\mathbf{W}\right) \\
& +16\frac{1}{n^2}\lambda^2_{\max}\left\{\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{S}(\rho_0)\right\}\lambda^2_{\max}\left\{\mathbf{D}(\rho_0)\right\}\lambda^2_{\max}\left\{\frac{\partial\mathbf{D}(\rho_0)}{\partial\rho}\right\}\lambda^{-1}_{\min}\left\{\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{S}(\rho_0)\right\}\text{trace}\left(\mathbf{W}\mathbf{W}^{\mathrm{T}}\right) \\
& +16\frac{1}{n^2}\lambda_{\max}\left\{\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{S}(\rho_0)\right\}\lambda^4_{\max}\left\{\mathbf{D}(\rho_0)\right\}\lambda^{-1}_{\min}\left\{\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{S}(\rho_0)\right\}\lambda_{\max}\left(\mathbf{W}^{\mathrm{T}}\mathbf{W}\right)\text{trace}\left(\mathbf{W}^{\mathrm{T}}\mathbf{W}\right) \\
=\ & O\left(\frac{1}{n}\right),
\end{aligned}
$$

where the last equality is by the Condition (C8) and (S.70).

Similarly,

$$
\begin{aligned}
& \frac{1}{n^2}\mathbf{h}_2(\mathbf{A})^{\mathrm{T}}\mathbf{h}_2(\mathbf{A}) \\
=\ & \frac{1}{n^2}\left[\left\{2\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)\frac{\partial\mathbf{D}(\rho_0)}{\partial\rho}\mathbf{S}(\rho_0)^{\mathrm{T}}\frac{\partial\mathbf{S}(\rho_0)}{\partial\rho}+\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\frac{\partial^2\mathbf{M}(\rho_0)}{\partial\rho^2}\right\}\mathbf{S}(\rho_0)^{-1}\mathbf{A}\mathbf{b}(\rho_0)\right. \\
& -2\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)\frac{\partial\mathbf{D}(\rho_0)}{\partial\rho}\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{A}\frac{\partial\mathbf{b}(\rho_0)}{\partial\rho} \\
& \left.-2\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\frac{\partial\mathbf{S}(\rho_0)^{\mathrm{T}}}{\partial\rho}\mathbf{A}\frac{\partial\mathbf{b}(\rho_0)}{\partial\rho}\right]^{\mathrm{T}} \\
& \times\left[\left\{2\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)\frac{\partial\mathbf{D}(\rho_0)}{\partial\rho}\mathbf{S}(\rho_0)^{\mathrm{T}}\frac{\partial\mathbf{S}(\rho_0)}{\partial\rho}+\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\frac{\partial^2\mathbf{M}(\rho_0)}{\partial\rho^2}\right\}\mathbf{S}(\rho_0)^{-1}\mathbf{A}\mathbf{b}(\rho_0)\right. \\
& -2\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)\frac{\partial\mathbf{D}(\rho_0)}{\partial\rho}\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{A}\frac{\partial\mathbf{b}(\rho_0)}{\partial\rho} \\
& \left.-2\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\frac{\partial\mathbf{S}(\rho_0)^{\mathrm{T}}}{\partial\rho}\mathbf{A}\frac{\partial\mathbf{b}(\rho_0)}{\partial\rho}\right] \\
\leqslant\ & \frac{16}{n^2}\mathbf{b}(\rho_0)^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{S}(\rho_0)^{-\mathrm{T}}\frac{\partial\mathbf{S}(\rho_0)^{\mathrm{T}}}{\partial\rho}\mathbf{S}(\rho_0)\frac{\partial\mathbf{D}(\rho_0)}{\partial\rho}\mathbf{D}(\rho_0)\mathbf{S}(\rho_0)^{\mathrm{T}} \\
& \times\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)\frac{\partial\mathbf{D}(\rho_0)}{\partial\rho}\mathbf{S}(\rho_0)^{\mathrm{T}}\frac{\partial\mathbf{S}(\rho_0)}{\partial\rho}\mathbf{S}(\rho_0)^{-1}\mathbf{A}\mathbf{b}(\rho_0) \\
& +\frac{4}{n^2}\mathbf{b}(\rho_0)^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{S}(\rho_0)^{-\mathrm{T}}\frac{\partial^2\mathbf{M}(\rho_0)^{\mathrm{T}}}{\partial\rho^2}\mathbf{D}(\rho_0)^2\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\frac{\partial^2\mathbf{M}(\rho_0)}{\partial\rho^2}\mathbf{S}(\rho_0)^{-1}\mathbf{A}\mathbf{b}(\rho_0) \\
& +\frac{16}{n^2}\frac{\partial\mathbf{b}(\rho_0)^{\mathrm{T}}}{\partial\rho}\mathbf{A}^{\mathrm{T}}\mathbf{S}(\rho_0)\frac{\partial\mathbf{D}(\rho_0)}{\partial\rho}\mathbf{D}(\rho_0)\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)\frac{\partial\mathbf{D}(\rho_0)}{\partial\rho}\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{A}\frac{\partial\mathbf{b}(\rho_0)}{\partial\rho}
\end{aligned}
$$

$$
\begin{aligned}
&+\frac{16}{n^2}\frac{\partial \mathbf{b}(\rho_0)^{\mathrm{T}}}{\partial \rho}\mathbf{A}^{\mathrm{T}}\frac{\partial \mathbf{S}(\rho_0)}{\partial \rho}\mathbf{D}(\rho_0)^2\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\frac{\partial \mathbf{S}(\rho_0)^{\mathrm{T}}}{\partial \rho}\mathbf{A}\frac{\partial \mathbf{b}(\rho_0)}{\partial \rho}\\
\leqslant\ & \frac{16}{n}\lambda_{\max}\left\{\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{S}(\rho_0)\right\}\lambda_{\max}^2\left\{\mathbf{D}(\rho_0)\right\}\lambda_{\max}^2\left\{\frac{\partial \mathbf{D}(\rho_0)}{\partial \rho}\right\}\lambda_{\max}\left\{\mathbf{S}(\rho_0)\mathbf{S}(\rho_0)^{\mathrm{T}}\right\}\\
&\lambda_{\max}\left(\mathbf{W}^{\mathrm{T}}\mathbf{W}\right)\lambda_{\min}^{-1}\left\{\mathbf{S}(\rho_0)\mathbf{S}(\rho_0)^{\mathrm{T}}\right\}\lambda_{\max}\left(\frac{1}{n}\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)\left\|\mathbf{b}(\rho_0)\right\|^2\\
&+\frac{16}{n}\lambda_{\max}\left\{\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{S}(\rho_0)\right\}\lambda_{\max}^4\left\{\mathbf{D}(\rho_0)\right\}\lambda_{\max}\left(\mathbf{W}^{\mathrm{T}}\mathbf{W}\right)\\
&\times\lambda_{\max}\left(\mathbf{W}\mathbf{W}^{\mathrm{T}}\right)\lambda_{\min}^{-1}\left\{\mathbf{S}(\rho_0)\mathbf{S}(\rho_0)^{\mathrm{T}}\right\}\lambda_{\max}\left(\frac{1}{n}\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)\left\|\mathbf{b}(\rho_0)\right\|^2\\
&+\frac{16}{n}\lambda_{\max}\left\{\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{S}(\rho_0)\right\}\lambda_{\max}^2\left\{\mathbf{D}(\rho_0)\right\}\lambda_{\max}^2\left\{\frac{\partial \mathbf{D}(\rho_0)}{\partial \rho}\right\}\\
&\quad\times\lambda_{\max}\left\{\mathbf{S}(\rho_0)\mathbf{S}(\rho_0)^{\mathrm{T}}\right\}\lambda_{\max}\left(\frac{1}{n}\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)\left\|\frac{\partial \mathbf{b}(\rho_0)}{\partial \rho}\right\|^2\\
&+\frac{16}{n}\lambda_{\max}\left\{\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{S}(\rho_0)\right\}\lambda_{\max}^4\left\{\mathbf{D}(\rho_0)\right\}\lambda_{\max}\left(\mathbf{W}\mathbf{W}^{\mathrm{T}}\right)\lambda_{\max}\left(\frac{1}{n}\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)\left\|\frac{\partial \mathbf{b}(\rho_0)}{\partial \rho}\right\|^2\\
=\ & O\left(\frac{1}{n}\right).
\end{aligned}
$$

Based on above results and (S.75), we have

$$
\operatorname{var}\left(\frac{1}{n}\mathbf{F}^{\mathrm{T}}\ddot{\mathbf{F}} \mid \mathbf{A}\right) = O\left(\frac{1}{n}\right).
$$

We also verify that $E(\mathbf{F}\ddot{\mathbf{F}}) = 0$ in the Supplementary Material S3.13. Then we obtain (S.68). Therefore, together with (S.67), (S.68) and (S.66), we conclude that

$$
\frac{1}{n}\sum_{i=1}^{n}\frac{\mathrm{d}S_i\{\mathbf{A};\mathbf{b}(\rho_0),\rho_0\}}{\mathrm{d}\rho_0} \mid \mathbf{A} \quad \xrightarrow{\mathrm{Pr}} \quad c_*, \tag{S.76}
$$

where c_* is defined in (S.67). Combining (S.64) and (S.76), we complete the proof of (S.62).

We next show that the asymptotic normality of numerator of (S.61). Note that

$$
n^{-3/4}\sum_{i=1}^{n}\sum_{k=1}^{K}\mathscr{B}_k(\mathbf{A})\varphi_{ik} + o_p(1) = n^{-3/4}\sum_{i=1}^{n}\mathscr{C}_i(\mathbf{A}) + o_p(1),
$$

where $\mathscr{C}_i(\mathbf{A}) = \sum_{k=1}^{K} \mathscr{B}_k(\mathbf{A})\varphi_{ik}$, the expression of $\mathscr{B}_k(\mathbf{A})$ and φ_{ik} refer to (S.42) and (S.28). Note that $\mathscr{C}_i(\mathbf{A})$'s are i.i.d for $i = 1, \ldots, n$ for given $\mathbf{A}$, $E\{n^{-3/4}\sum_{i=1}^{n} \mathscr{C}_i(\mathbf{A})\} = 0$ from (S.46) and $\mathrm{var}\{n^{-3/4}\sum_{i=1}^{n} \mathscr{C}_i(\mathbf{A})\} = \Sigma_1$, where Σ_1 is defined as (S.60). Then, by the central limit theorem, we have $n^{-3/4}\sum_{i=1}^{n} \mathscr{C}_i(\mathbf{A}) \xrightarrow{\mathrm{d}} \mathcal{N}(0, \Sigma_1)$ as $n \to \infty$.

Therefore, according to above results, (S.62) and(S.61), we have, for given $\mathbf{A}$,

$$n^{1/4}\left(\widehat{\rho} - \rho_0\right) \xrightarrow{\mathrm{d}} \mathcal{N}(0, \Sigma),$$

where $\Sigma = c_*^{-2}\Sigma_1$. □

S3.11 Proof of (S.36)

To prove the (S.36), we first verify the following equation.

$$E\left[\iint \{Z_i(u)Z_i(v) - G(u,v)\}\,\phi_k(s)\phi_\ell(t)dsdt\right]^2 = O(\lambda_k\lambda_\ell). \tag{S.77}$$

Since

$$\begin{aligned} E(\xi_{ik}\xi_{i\ell}) &= \iint G(s,t)\phi_k(s)\phi_\ell(t)dsdt \\ &= \begin{cases} \lambda_k, & \text{if } k = \ell, \\ 0, & \text{if } k \neq \ell, \end{cases} \end{aligned}$$

we have, for $k \neq \ell$,

$$\begin{aligned} &\iint \{Z_i(s)Z_i(t) - G(s,t)\}\,\phi_k(s)\phi_\ell(t)dsdt \\ = &\int Z_i(s)\phi_k(s)ds \int Z_i(t)\phi_\ell(t)dt \end{aligned}$$

$$= \quad \xi_{ik}\xi_{i\ell},$$

and by Condition (C1),

$$\begin{aligned} E\{(\xi_{ik}\xi_{i\ell})^2\} &\leqslant \{E(\xi_{ik}^4)E(\xi_{i\ell}^4)\}^{1/2} \\ &\leqslant C\lambda_k\lambda_\ell. \end{aligned}$$

Hence (S.77) follows. This leads to

$$\iint \{Z_i(u)Z_i(v) - G(u,v)\}\,\phi_k(s)\phi_\ell(t)dsdt = O_p(\lambda_k^{1/2}\lambda_\ell^{1/2}),$$

from which and combined with (S.14), we have

$$\sum_{\ell:\ell\neq k}(\lambda_k-\lambda_\ell)^{-1}\iint \{Z_i(u)Z_i(v) - G(u,v)\}\,\phi_k(u)\phi_\ell(v)dudv = O_p(k\log k). \tag{S.78}$$

Next we prove (S.36). Recall that $g(t) = E\{Y_i^* Z_i(t)\}$, $\int \phi_k^2(t)dt = 1$, and according to (S.28), $E\varphi_{jk} = 0$, φ_{jk}'s are i.i.d for $j = 1,\ldots,n$. Further incorporating Condition (C4), Lemma S2.3 and (S.78), we have

$$\begin{aligned} &\operatorname{var}\left(n^{-3/4}\sum_{j=1}^n \varphi_{jk}\right) \\ =\ & E\left(n^{-3/4}\sum_{j=1}^n \varphi_{jk}\right)^2 \\ =\ & n^{-1/2}E\left\{\left(\lambda_k^{-1}\int \phi_k(t)\left[Y_i^* Z_i(t) - E\left\{Y_i^* Z_i(t)\right\}\right]dt + \lambda_k^{-1}\int E\left\{Y_i^* Z_i(t)\right\}\right.\right. \\ &\quad\times\left[\sum_{\ell:\ell\neq k}(\lambda_k-\lambda_\ell)^{-1}\phi_k(t)\iint \{Z_i(u)Z_i(v)-G(u,v)\}\,\phi_k(u)\phi_\ell(v)dudv\right]dt \\ &\left.\left. -\lambda_k^{-2}(\xi_{ik}^2-\lambda_k)\int \phi_k(t)g(t)dt\right)\right\}^2 \end{aligned}$$

$$
\begin{aligned}
&\leqslant 3n^{-1/2}\lambda_k^{-2}E\left\{\left(\int \phi_k(t)\left[Y_i^*Z_i(t)-E\left\{Y_i^*Z_i(t)\right\}\right]dt\right)^2\right\}\\
&\quad +3n^{-1/2}\lambda_k^{-2}E\left\{\left(\int E\left\{Y_i^*Z_i(t)\right\}\left[\sum_{\ell:\ell\neq k}(\lambda_k-\lambda_\ell)^{-1}\phi_k(t)\right.\right.\right.\\
&\quad \left.\left.\left.\times\iint\left\{Z_i(u)Z_i(v)-G(u,v)\right\}\phi_k(u)\phi_\ell(v)dudv\right]dt\right)^2\right\}\\
&\quad +3n^{-1/2}\lambda_k^{-4}E\{(\xi_{ik}^2-\lambda_k)^2\}\left\{\int\phi_k(t)g(t)dt\right\}^2\\
&\leqslant 3n^{-1/2}\lambda_k^{-2}\int E\left[Y_i^*Z_i(t)-E\left\{Y_i^*Z_i(t)\right\}\right]^2dt\left\{\int\phi_k^2(t)dt\right\}\\
&\quad +3n^{-1/2}\lambda_k^{-2}\left\{\int g^2(t)dt\right\}\\
&\quad \times E\left(\int\left[\sum_{\ell:\ell\neq k}(\lambda_k-\lambda_\ell)^{-1}\phi_k(t)\iint\left\{Z_i(u)Z_i(v)-G(u,v)\right\}\phi_k(u)\phi_\ell(v)dudv\right]^2dt\right)\\
&\quad +3(C-1)n^{-1/2}\lambda_k^{-2}\left\{\int g^2(t)dt\right\}\left\{\int\phi_k^2(t)dt\right\}\\
&\leqslant 3n^{-1/2}\lambda_k^{-2}\int \mathrm{var}\left\{Y_i^*Z_i(t)\right\}dt\left\{\int\phi_k^2(t)dt\right\}\\
&\quad +3n^{-1/2}\lambda_k^{-2}\left\{\int g^2(t)dt\right\}\\
&\quad \times E\left[\sum_{\ell:\ell\neq k}(\lambda_k-\lambda_\ell)^{-1}\iint\left\{Z_i(u)Z_i(v)-G(u,v)\right\}\phi_k(u)\phi_\ell(v)dudv\right]^2\int\phi_k^2(t)dt\\
&\quad +3Cn^{-1/2}\lambda_k^{-2}\left\{\int g^2(t)dt\right\}\left\{\int\phi_k^2(t)dt\right\}\\
&\leqslant 3n^{-1/2}\lambda_k^{-2}\int E\left\{Y_i^*Z_i(t)\right\}^2dt+3\widetilde{C}n^{-1/2}\lambda_k^{-2}\|g\|^2(k\log k)^2+3Cn^{-1/2}\lambda_k^{-2}\|g\|^2\\
&\leqslant 3C^2n^{-1/2}k^{2\alpha}E\left(Y_i^{*4}\right)\int E\left\{Z_i^4(t)\right\}dt+3\widetilde{C}C^2n^{-1/2}k^{2\alpha}(\|g\|k\log k)^2+3C^3n^{-1/2}k^{2\alpha}\|g\|^2\\
&= O\{n^{-1/2}k^{2\alpha+2}(\log k)^2\},
\end{aligned}
$$

where the last equality is due to the Condition (C1). Together with $E\varphi_{jk}=0$, this gives

$$
n^{-3/4}\sum_{j=1}^{n}\varphi_{jk}=O_p\left\{n^{-1/4}k^{\alpha+1}(\log k)\right\},
$$

which leads to, by Condition (C5),

$$
\begin{aligned}
& \sum_{k=1}^{K}\left(n^{-3/4}\sum_{j=1}^{n}\varphi_{jk}\right) \\
= \;& O_p\left\{n^{-1/4}\sum_{k=1}^{K}k^{\alpha+1}(\log k)\right\} \\
= \;& O_p\left(n^{-1/4}\int_1^{K+1}x^{\alpha+1}\log x dx\right) \\
= \;& O_p\left(n^{-1/4}\left[\frac{1}{\alpha+2}\log(K+1)(K+1)^{\alpha+2}-\frac{1}{(\alpha+2)^2}\left\{(K+1)^{\alpha+2}-1\right\}\right]\right) \\
= \;& O_p(n^{-1/4}\log K K^{\alpha+2}) \\
= \;& O_p(1).
\end{aligned}
$$

This verifies (S.36). □

S3.12 Proof of (S.43)

For the first part of (S.43), using Lemma S2.7, the Condition (C8) and $\mathrm{cov}(\boldsymbol{\varepsilon})=\sigma^2\mathbf{I}_n$, we note that

$$
\begin{aligned}
& E\left\{\mathscr{A}_k^2(\mathbf{A})\mid\mathbf{A}\right\} \\
= \;& \frac{1}{n^2}E\left\{\boldsymbol{\varepsilon}^{\mathrm{T}}\mathbf{S}^{-\mathrm{T}}(\rho_0)\mathbf{W}^{\mathrm{T}}\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{A}\mathbf{e}_k\mathbf{e}_k^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{W}\mathbf{S}^{-1}(\rho_0)\boldsymbol{\varepsilon}\mid\mathbf{A}\right\} \\
= \;& \frac{\sigma^2}{n^2}\mathrm{trace}\left\{\mathbf{S}^{-\mathrm{T}}(\rho_0)\mathbf{W}^{\mathrm{T}}\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{A}\mathbf{e}_k\mathbf{e}_k^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{W}\mathbf{S}^{-1}(\rho_0)\mid\mathbf{A}\right\} \\
= \;& \frac{\sigma^2}{n^2}\mathrm{trace}\left\{\mathbf{e}_k^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{W}\mathbf{S}^{-1}(\rho_0)\mathbf{S}^{-\mathrm{T}}(\rho_0)\mathbf{W}^{\mathrm{T}}\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{A}\mathbf{e}_k\mid\mathbf{A}\right\} \\
= \;& \frac{\sigma^2}{n^2}\mathbf{e}_k^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{W}\mathbf{S}^{-1}(\rho_0)\mathbf{S}^{-\mathrm{T}}(\rho_0)\mathbf{W}^{\mathrm{T}}\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{A}\mathbf{e}_k \\
\leqslant \;& \sigma^2\lambda_{\max}\left\{\mathbf{S}^{-1}(\rho_0)\mathbf{S}^{-\mathrm{T}}(\rho_0)\right\}\lambda_{\max}\left(\mathbf{W}\mathbf{W}^{\mathrm{T}}\right)\lambda_{\max}\left\{\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{S}(\rho_0)\right\} \\
& \times\lambda_{\max}\left\{\mathbf{D}(\rho_0)^4\right\}\lambda_{\max}\left\{\mathbf{S}(\rho_0)\mathbf{S}(\rho_0)^{\mathrm{T}}\right\}\lambda_{\max}\left(\frac{1}{n}\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)\left(\frac{1}{n}\mathbf{e}_k^{\mathrm{T}}\mathbf{e}_k\right)
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{\sigma^2}{n}\left[\lambda_{\min}\left\{\mathbf{M}(\rho_0)\right\}\right]^{-1}\lambda_{\max}\left(\mathbf{W}\mathbf{W}^{\mathrm{T}}\right)\lambda_{\max}\left\{\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{S}(\rho_0)\right\} \\
&\quad \times\lambda_{\max}\left\{\mathbf{D}(\rho_0)^4\right\}\lambda_{\max}\left\{\mathbf{S}(\rho_0)\mathbf{S}(\rho_0)^{\mathrm{T}}\right\}\lambda_{\max}\left(\frac{1}{n}\mathbf{A}^{\mathrm{T}}\mathbf{A}\right) \\
&\leqslant \frac{1}{n}C,
\end{aligned}
$$

for fixed $\mathbf{A}$ and some constant C, where the last equality is due to Remark 2. Hence, $E\{\mathscr{A}_k^2(\mathbf{A})\} = E[E\{\mathscr{A}_k^2(\mathbf{A}) \mid \mathbf{A}\}] = O(1/n)$. This shows that $\mathscr{A}_k(\mathbf{A}) = O_p(1/\sqrt{n})$.

For the second part of (S.43), write

$$
\mathscr{B}_k(\mathbf{A}) = \mathscr{B}_{k1}(\mathbf{A}) + \mathscr{B}_{k2}(\mathbf{A}),
$$

where

$$
\mathscr{B}_{k1}(\mathbf{A}) = \frac{1}{n}\mathbf{b}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{S}^{-\mathrm{T}}(\rho_0)\mathbf{W}^{\mathrm{T}}\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{A}\mathbf{e}_k,
$$

and

$$
\mathscr{B}_{k2}(\mathbf{A}) = \frac{1}{n}\frac{\partial\mathbf{b}(\rho_0)^{\mathrm{T}}}{\partial\rho}\mathbf{A}^{\mathrm{T}}\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{A}\mathbf{e}_k.
$$

Similarly, by the Condition (C3),

$$
\begin{aligned}
&E\left\{\mathscr{B}_{k1}^2(\mathbf{A}) \mid \mathbf{A}\right\} \\
&= \frac{1}{n^2}E\left\{\mathbf{b}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{S}(\rho_0)^{-\mathrm{T}}\mathbf{W}^{\mathrm{T}}\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{A}\mathbf{e}_k\mathbf{e}_k^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{W}\mathbf{S}(\rho_0)^{-1}\mathbf{A}\mathbf{b} \mid \mathbf{A}\right\} \\
&= \frac{1}{n^2}\mathbf{b}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{S}(\rho_0)^{-\mathrm{T}}\mathbf{W}^{\mathrm{T}}\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{A}\mathbf{e}_k\mathbf{e}_k^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{W}\mathbf{S}(\rho_0)^{-1}\mathbf{A}\mathbf{b} \\
&\leqslant \lambda_{\max}\left(\mathbf{e}_k\mathbf{e}_k^{\mathrm{T}}\right)\lambda_{\max}\left(\frac{1}{n}\mathbf{A}\mathbf{A}^{\mathrm{T}}\right)\lambda_{\max}\left\{\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{S}(\rho_0)\right\}\lambda_{\max}\left\{\mathbf{D}(\rho_0)^4\right\}\lambda_{\max}\left\{\mathbf{S}(\rho_0)\mathbf{S}(\rho_0)^{\mathrm{T}}\right\}
\end{aligned}
$$

$$
\begin{aligned}
& \times\lambda_{\max}\left(\mathbf{W}^{\mathrm{T}}\mathbf{W}\right)\lambda_{\max}\left\{\mathbf{S}(\rho_0)^{-\mathrm{T}}\mathbf{S}(\rho_0)^{-1}\right\}\lambda_{\max}\left(\frac{1}{n}\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)\mathbf{b}^{\mathrm{T}}\mathbf{b} \\
= \; & \lambda_{\max}\left(\frac{1}{n}\mathbf{A}\mathbf{A}^{\mathrm{T}}\right)\lambda_{\max}\left\{\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{S}(\rho_0)\right\}\lambda_{\max}\left\{\mathbf{D}(\rho_0)^4\right\}\lambda_{\max}\left\{\mathbf{S}(\rho_0)\mathbf{S}(\rho_0)^{\mathrm{T}}\right\} \\
& \times\lambda_{\max}\left(\mathbf{W}^{\mathrm{T}}\mathbf{W}\right)\lambda_{\max}\left\{\mathbf{S}(\rho_0)^{-\mathrm{T}}\mathbf{S}(\rho_0)^{-1}\right\}\lambda_{\max}\left(\frac{1}{n}\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)\left(\sum_{k=1}^{K} b_k^2\right) \\
\leqslant \; & C_1\lambda_{\max}\left(\frac{1}{n}\mathbf{A}\mathbf{A}^{\mathrm{T}}\right)\lambda_{\max}\left\{\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{S}(\rho_0)\right\}\lambda_{\max}\left\{\mathbf{D}(\rho_0)^4\right\}\lambda_{\max}\left\{\mathbf{S}(\rho_0)\mathbf{S}(\rho_0)^{\mathrm{T}}\right\} \\
& \times\lambda_{\max}\left(\mathbf{W}^{\mathrm{T}}\mathbf{W}\right)\left[\lambda_{\min}\left\{\mathbf{S}(\rho_0)\mathbf{S}(\rho_0)^{\mathrm{T}}\right\}\right]^{-1}\lambda_{\max}\left(\frac{1}{n}\mathbf{A}^{\mathrm{T}}\mathbf{A}\right) \\
\leqslant \; & C_2.
\end{aligned}
$$

Similar derivation can show that, for some constant C_3, $E\{\mathscr{B}_{k2}^2(\mathbf{A}) \mid \mathbf{A}\} \leqslant C_3$. Since $E\{\mathscr{B}_k^2(\mathbf{A}) \mid \mathbf{A}\} \leqslant 2E\{\mathscr{B}_{k1}^2(\mathbf{A}) \mid \mathbf{A}\} + 2E\{\mathscr{B}_{k2}^2(\mathbf{A}) \mid \mathbf{A}\}$, we have $\mathscr{B}_k(\mathbf{A}) = O_p(1)$. □

S3.13 Proof of $E(\mathbf{F}^{\mathrm{T}}\ddot{\mathbf{F}}) = 0$

Next we show that $E(\mathbf{F}^{\mathrm{T}}\ddot{\mathbf{F}}) = 0$. Note that, by $w_{ii} = 0$ for $i = 1, \ldots, n$, we have

$$
\begin{aligned}
\mathbf{M}(\rho) &= \mathbf{S}(\rho)^{\mathrm{T}}\mathbf{S}(\rho) = \mathbf{I}_n - \rho\mathbf{W} - \rho\mathbf{W}^{\mathrm{T}} + \rho^2\mathbf{W}^{\mathrm{T}}\mathbf{W}, \\
\frac{\partial\mathbf{M}(\rho)}{\partial\rho} &= -\mathbf{W} - \mathbf{W}^{\mathrm{T}} + 2\rho\mathbf{W}^{\mathrm{T}}\mathbf{W}, \\
\frac{\partial^2\mathbf{M}(\rho)}{\partial\rho^2} &= 2\mathbf{W}^{\mathrm{T}}\mathbf{W}, \\
\mathbf{D}(\rho) &= \{\mathrm{diag}\mathbf{M}(\rho)\}^{-1} = \left\{\mathbf{I}_n + \rho^2\mathrm{diag}(\mathbf{W}^{\mathrm{T}}\mathbf{W})\right\}^{-1}, \\
\frac{\partial\mathbf{D}(\rho)}{\partial\rho} &= -\left\{\mathbf{I}_n + \rho^2\mathrm{diag}(\mathbf{W}^{\mathrm{T}}\mathbf{W})\right\}^{-2}\mathrm{diag}(2\rho\mathbf{W}^{\mathrm{T}}\mathbf{W}) \\
&= -\{\mathbf{D}(\rho)\}^2\mathrm{diag}(2\rho\mathbf{W}^{\mathrm{T}}\mathbf{W}), \\
\frac{\partial^2\mathbf{D}(\rho)}{\partial\rho^2} &= 2\left\{\mathbf{I}_n + \rho^2\mathrm{diag}(\mathbf{W}^{\mathrm{T}}\mathbf{W})\right\}^{-3}\{\mathrm{diag}(2\rho\mathbf{W}^{\mathrm{T}}\mathbf{W})\}^2 \\
&\quad - \left\{\mathbf{I}_n + \rho^2\mathrm{diag}(\mathbf{W}^{\mathrm{T}}\mathbf{W})\right\}^{-2}\mathrm{diag}(2\mathbf{W}^{\mathrm{T}}\mathbf{W}) \\
&= 2\{\mathbf{D}(\rho)\}^3\{\mathrm{diag}(2\rho\mathbf{W}^{\mathrm{T}}\mathbf{W})\}^2 - 2\{\mathbf{D}(\rho)\}^2\mathrm{diag}(\mathbf{W}^{\mathrm{T}}\mathbf{W}).
\end{aligned}
$$

By (S.74), we rewrite $\mathbf{F}^{\mathrm{T}}\ddot{\mathbf{F}}$ as follows.

$$\begin{aligned}
\mathbf{F}^{\mathrm{T}}\ddot{\mathbf{F}} &= \boldsymbol{\varepsilon}_n^{\mathrm{T}}\left\{\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)\frac{\partial^2\mathbf{D}(\rho_0)}{\partial\rho^2}\mathbf{S}(\rho_0)^{\mathrm{T}} + 2\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)\frac{\partial\mathbf{D}(\rho_0)}{\partial\rho}\frac{\partial\mathbf{S}(\rho_0)^{\mathrm{T}}}{\partial\rho}\right\}\boldsymbol{\varepsilon}_n \\
&\quad +\boldsymbol{\varepsilon}_n^{\mathrm{T}}\left\{2\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)\frac{\partial\mathbf{D}(\rho_0)}{\partial\rho}\mathbf{S}(\rho_0)^{\mathrm{T}}\frac{\partial\mathbf{S}(\rho_0)}{\partial\rho} + \mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\frac{\partial^2\mathbf{M}(\rho_0)}{\partial\rho^2}\right\}\mathbf{Y} \\
&\quad -2\boldsymbol{\varepsilon}_n^{\mathrm{T}}\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)\frac{\partial\mathbf{D}(\rho_0)}{\partial\rho}\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{A}\frac{\partial\mathbf{b}(\rho_0)}{\partial\rho} - 2\boldsymbol{\varepsilon}_n^{\mathrm{T}}\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\frac{\partial\mathbf{S}(\rho_0)^{\mathrm{T}}}{\partial\rho}\mathbf{A}\frac{\partial\mathbf{b}(\rho_0)}{\partial\rho} \\
&= \boldsymbol{\varepsilon}_n^{\mathrm{T}}\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)\frac{\partial^2\mathbf{D}(\rho_0)}{\partial\rho^2}\mathbf{S}(\rho_0)^{\mathrm{T}}\boldsymbol{\varepsilon}_n \\
&\quad +2\boldsymbol{\varepsilon}_n^{\mathrm{T}}\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)\frac{\partial\mathbf{D}(\rho_0)}{\partial\rho}\left\{\frac{\partial\mathbf{S}(\rho_0)^{\mathrm{T}}}{\partial\rho}\boldsymbol{\varepsilon}_n + \mathbf{S}(\rho_0)^{\mathrm{T}}\frac{\partial\mathbf{S}(\rho_0)}{\partial\rho}\mathbf{Y}\right\} \\
&\quad +\boldsymbol{\varepsilon}_n^{\mathrm{T}}\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\frac{\partial^2\mathbf{M}(\rho_0)}{\partial\rho^2}\mathbf{Y} \\
&\quad +\left\{-2\boldsymbol{\varepsilon}_n^{\mathrm{T}}\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)\frac{\partial\mathbf{D}(\rho_0)}{\partial\rho}\mathbf{S}(\rho_0)^{\mathrm{T}}\mathbf{A}\frac{\partial\mathbf{b}(\rho_0)}{\partial\rho} - 2\boldsymbol{\varepsilon}_n^{\mathrm{T}}\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\frac{\partial\mathbf{S}(\rho_0)^{\mathrm{T}}}{\partial\rho}\mathbf{A}\frac{\partial\mathbf{b}(\rho_0)}{\partial\rho}\right\} \\
&\equiv \mathbb{S}_1 + \mathbb{S}_2 + \mathbb{S}_3 + \mathbb{S}_4.
\end{aligned}$$

Note that $E(\mathbb{S}_4 \mid \mathbf{A}) = 0$ and

$$\begin{aligned}
E(\mathbb{S}_1 \mid \mathbf{A}) &= E\left\{\boldsymbol{\varepsilon}_n^{\mathrm{T}}\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)\frac{\partial^2\mathbf{D}(\rho_0)}{\partial\rho^2}\mathbf{S}(\rho_0)^{\mathrm{T}}\boldsymbol{\varepsilon}_n \mid \mathbf{A}\right\} \\
&= \sigma^2\mathrm{trace}\left\{\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)\frac{\partial^2\mathbf{D}(\rho_0)}{\partial\rho^2}\mathbf{S}(\rho_0)^{\mathrm{T}}\right\} \\
&= \sigma^2\mathrm{trace}\left\{\frac{\partial^2\mathbf{D}(\rho_0)}{\partial\rho^2}\mathbf{M}(\rho_0)\mathbf{D}(\rho_0)\right\} \\
&= \sigma^2\mathrm{trace}\left\{\frac{\partial^2\mathbf{D}(\rho_0)}{\partial\rho^2}\right\},
\end{aligned}$$

where the last equality is due to $\mathbf{D}(\rho) = \{\mathrm{diag}\mathbf{M}(\rho)\}^{-1}$. By (2.6), we have

$$\begin{aligned}
E(\mathbb{S}_2 \mid \mathbf{A}) &= E\left[2\boldsymbol{\varepsilon}_n^{\mathrm{T}}\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)\frac{\partial\mathbf{D}(\rho_0)}{\partial\rho}\left\{\frac{\partial\mathbf{S}(\rho_0)^{\mathrm{T}}}{\partial\rho}\boldsymbol{\varepsilon}_n + \mathbf{S}(\rho_0)^{\mathrm{T}}\frac{\partial\mathbf{S}(\rho_0)}{\partial\rho}\mathbf{Y}\right\} \mid \mathbf{A}\right] \\
&= 2\sigma^2\mathrm{trace}\left\{\mathbf{D}(\rho_0)\frac{\partial\mathbf{D}(\rho_0)}{\partial\rho}\frac{\partial\mathbf{S}(\rho_0)^{\mathrm{T}}}{\partial\rho}\mathbf{S}(\rho_0)\right\}
\end{aligned}$$

$$
\begin{aligned}
& +2E\left\{\boldsymbol{\varepsilon}_n^{\mathrm{T}}\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)\frac{\partial\mathbf{D}(\rho_0)}{\partial\rho}\mathbf{S}(\rho_0)^{\mathrm{T}}\frac{\partial\mathbf{S}(\rho_0)}{\partial\rho}\mathbf{S}^{-1}(\rho_0)\boldsymbol{\varepsilon}_n \mid \mathbf{A}\right\} \\
= \; & 2\sigma^2\mathrm{trace}\left\{\mathbf{D}(\rho_0)\frac{\partial\mathbf{D}(\rho_0)}{\partial\rho}\frac{\partial\mathbf{S}(\rho_0)^{\mathrm{T}}}{\partial\rho}\mathbf{S}(\rho_0)\right\} \\
& +2\sigma^2\mathrm{trace}\left\{\mathbf{D}(\rho_0)\frac{\partial\mathbf{D}(\rho_0)}{\partial\rho}\mathbf{S}(\rho_0)^{\mathrm{T}}\frac{\partial\mathbf{S}(\rho_0)}{\partial\rho}\right\},
\end{aligned}
$$

and

$$
\begin{aligned}
E(\mathbb{S}_3 \mid \mathbf{A}) &= E\left\{\boldsymbol{\varepsilon}_n^{\mathrm{T}}\mathbf{S}(\rho_0)\mathbf{D}(\rho_0)^2\frac{\partial^2\mathbf{M}(\rho_0)}{\partial\rho^2}\mathbf{S}^{-1}(\rho_0)\boldsymbol{\varepsilon}_n \mid \mathbf{A}\right\} \\
&= \sigma^2\mathrm{trace}\left\{\mathbf{D}(\rho_0)^2\frac{\partial^2\mathbf{M}(\rho_0)}{\partial\rho^2}\right\}.
\end{aligned}
$$

Now

$$
\begin{aligned}
E(\mathbb{S}_2 \mid \mathbf{A}) &= 2\sigma^2\mathrm{trace}\left\{\mathbf{D}(\rho_0)\frac{\partial\mathbf{D}(\rho_0)}{\partial\rho}\frac{\partial\mathbf{S}(\rho_0)^{\mathrm{T}}}{\partial\rho}\mathbf{S}(\rho_0)\right\} \\
&\quad +2\sigma^2\mathrm{trace}\left\{\mathbf{D}(\rho_0)\frac{\partial\mathbf{D}(\rho_0)}{\partial\rho}\mathbf{S}(\rho_0)^{\mathrm{T}}\frac{\partial\mathbf{S}(\rho_0)}{\partial\rho}\right\} \\
&= 2\sigma^2\mathrm{trace}\left\{\mathbf{D}(\rho_0)\frac{\partial\mathbf{D}(\rho_0)}{\partial\rho}\frac{\partial\mathbf{S}(\rho_0)^{\mathrm{T}}}{\partial\rho}\right\} \\
&\quad -2\rho\sigma^2\mathrm{trace}\left\{\mathbf{D}(\rho_0)\frac{\partial\mathbf{D}(\rho_0)}{\partial\rho}\frac{\partial\mathbf{S}(\rho_0)^{\mathrm{T}}}{\partial\rho}\mathbf{W}\right\} \\
&\quad +2\sigma^2\mathrm{trace}\left\{\mathbf{D}(\rho_0)\frac{\partial\mathbf{D}(\rho_0)}{\partial\rho}\frac{\partial\mathbf{S}(\rho_0)}{\partial\rho}\right\} \\
&\quad -2\rho\sigma^2\mathrm{trace}\left\{\mathbf{D}(\rho_0)\frac{\partial\mathbf{D}(\rho_0)}{\partial\rho}\mathbf{W}^{\mathrm{T}}\frac{\partial\mathbf{S}(\rho_0)}{\partial\rho}\right\} \\
&= -2\sigma^2\mathrm{trace}\left\{\mathbf{D}(\rho_0)\frac{\partial\mathbf{D}(\rho_0)}{\partial\rho}\mathbf{W}^{\mathrm{T}}\right\} \\
&\quad +2\rho\sigma^2\mathrm{trace}\left\{\mathbf{D}(\rho_0)\frac{\partial\mathbf{D}(\rho_0)}{\partial\rho}\mathbf{W}^{\mathrm{T}}\mathbf{W}\right\} \\
&\quad -2\sigma^2\mathrm{trace}\left\{\mathbf{D}(\rho_0)\frac{\partial\mathbf{D}(\rho_0)}{\partial\rho}\mathbf{W}\right\} \\
&\quad +2\rho\sigma^2\mathrm{trace}\left\{\mathbf{D}(\rho_0)\frac{\partial\mathbf{D}(\rho_0)}{\partial\rho}\mathbf{W}^{\mathrm{T}}\mathbf{W}\right\} \\
&= 4\rho\sigma^2\mathrm{trace}\left\{\mathbf{D}(\rho_0)\frac{\partial\mathbf{D}(\rho_0)}{\partial\rho}\mathbf{W}^{\mathrm{T}}\mathbf{W}\right\}
\end{aligned}
$$

$$
\begin{aligned}
&= -4\rho\sigma^2\text{trace}\left[\mathbf{D}(\rho_0)\left\{\mathbf{D}(\rho)^2\text{diag}(2\rho\mathbf{W}^{\mathrm{T}}\mathbf{W})\right\}\mathbf{W}^{\mathrm{T}}\mathbf{W}\right] \\
&= -2\sigma^2\text{trace}\left[\{\mathbf{D}(\rho_0)\}^3\left\{\text{diag}(2\rho\mathbf{W}^{\mathrm{T}}\mathbf{W})\right\}^2\right],
\end{aligned}
$$

where the last third equality is due to $w_{ii} = 0$. It is also observed that

$$
\begin{aligned}
E(\mathbb{S}_3 \mid \mathbf{A}) &= \sigma^2\text{trace}\left\{\mathbf{D}(\rho_0)^2\frac{\partial^2\mathbf{M}(\rho_0)}{\partial\rho^2}\right\} \\
&= 2\sigma^2\text{trace}\left[\{\mathbf{D}(\rho_0)\}^2\mathbf{W}^{\mathrm{T}}\mathbf{W}\right],
\end{aligned}
$$

and

$$
\begin{aligned}
E(\mathbb{S}_1 \mid \mathbf{A}) &= \sigma^2\text{trace}\left\{\frac{\partial^2\mathbf{D}(\rho_0)}{\partial\rho^2}\right\} \\
&= 2\sigma^2\text{trace}\left[\{\mathbf{D}(\rho)\}^3\{\text{diag}(2\rho\mathbf{W}^{\mathrm{T}}\mathbf{W})\}^2\right] \\
&\quad -2\sigma^2\text{trace}\left[\{\mathbf{D}(\rho)\}^2\,\text{diag}(\mathbf{W}^{\mathrm{T}}\mathbf{W})\right].
\end{aligned}
$$

Consequently, we have

$$
\begin{aligned}
E(\mathbf{F}^{\mathrm{T}}\ddot{\mathbf{F}} \mid \mathbf{A}) &= E(\mathbb{S}_1 + \mathbb{S}_2 + \mathbb{S}_3 + \mathbb{S}_4 \mid \mathbf{A}) \\
&= 0.
\end{aligned}
$$

This completes the proof. □

S3.14 Proof of Lemma S2.10

Part (a) follows from Lemma 3 of Kong et al. (2016). For part (i) in (b), let $T_{is} = (\widehat{\xi}_{is} - \xi_{is})\xi_{ik}$ and define $r_n = k^{1-\alpha/2}n^{-1/2}$ for notational convenience, where $1 \leqslant s \leqslant K_0$ and $K_0 + 1 \leqslant k \leqslant$

K^*. Then, by Condition (C7) and Lemma S2.4(c), it follows that $T_{is} = O_p(r_n)$. Note that

$$\left\| \frac{1}{\sqrt{n}} \left(\widehat{\mathbf{A}}_0^{\mathrm{T}} - \mathbf{A}_0^{\mathrm{T}} \right) \mathbb{A}_k \right\| = \left\{ \sum_{s=1}^{K} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} T_{is} \right)^2 \right\}^{1/2},$$

and

$$\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} T_{is} \right)^2 = \frac{1}{n} \sum_{i=1}^{n} T_{is}^2 + \frac{n(n-1)}{n} U_{ns}, \tag{S.79}$$

where $U_{ns} = \frac{1}{n(n-1)} \sum_{i \neq j} T_{is} T_{js}$. It is obvious that $\frac{1}{n} \sum_{i=1}^{n} T_{is}^2 = O_p(r_n^2)$. Since $k \neq s$, and by Lemma S2.4(c), we have

$$\begin{aligned}
|E(T_{is})| &= |E\{(\widehat{\xi}_{is} - \xi_{is})\xi_{ik}\}| \\
&\leqslant |E\{(\widehat{\xi}_{is} - \widehat{\xi}_{is}^{[-i]})\xi_{ik}\}| + |E\{(\widehat{\xi}_{is}^{[-i]} - \xi_{is})\xi_{ik}\}| \\
&= |E\{(\widehat{\xi}_{is} - \widehat{\xi}_{is}^{[-i]})\xi_{ik}\}| \\
&= |E[\{(\widehat{\xi}_{is} - \xi_{is}) - (\widehat{\xi}_{is}^{[-i]} - \xi_{is})\xi_{ik}\}]| \\
&= a|k^{1-\alpha/2} n^{-1/2} - k^{1-\alpha/2}(n-1)^{-1/2}| \\
&= O(k^{1-\alpha/2} n^{-3/2}),
\end{aligned}$$

where the second equality holds because $\widehat{\xi}_{is}^{[-i]}$ and ξ_{ik} are uncorrelated, and a is a positive constant. This leads to

$$\begin{aligned}
E\left(\frac{n(n-1)}{n} U_{ns} \right) &= \frac{1}{n} \sum_{i \neq j} E(T_{is}) E(T_{js}) \\
&= \frac{1}{n} n(n-1) O(k^{2-\alpha} n^{-3}) \\
&= O(k^{2-\alpha} n^{-2}).
\end{aligned}$$

Further, using Lemma 5.2.1A in Serfling (1980) and some calculations, we obtain

$$\begin{aligned}
\mathrm{var}(U_{ns}) &= \frac{2}{n(n-1)}[2(n-2)(ET_{is})^2 var(T_{is}) + \{\mathrm{var}(T_{is})\}^2 + 2(ET_{is})^2\mathrm{var}(T_{is}^2)] \\
&= O\left(\frac{r_n^4}{n}\right),
\end{aligned}$$

and

$$\mathrm{var}\left(\frac{n(n-1)}{n}U_{ns}\right) = O(nr_n^4).$$

Hence, by (S.79), we have

$$\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} T_{is}\right)^2 = O_p(r_n^2) + O(k^{2-\alpha}n^{-2}) + O_p(n^{1/2}r_n^2) = O_p(k^{2-\alpha}n^{-1/2}),$$

from which we have

$$\left\{\sum_{s=1}^{K}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} T_{is}\right)^2\right\}^{1/2} = O_p\left(\frac{K^{3/2-\alpha/2}}{n^{1/4}}\right).$$

For part (ii) in (b), it follows that

$$\begin{aligned}
&\left\|\widehat{\mathbb{A}}_k^{\mathrm{proj}} - \mathbb{A}_k^{\mathrm{proj}}\right\| \\
=& \left\|\widehat{\mathbf{Q}}_0\widehat{\mathbb{A}}_k - \mathbf{Q}_0\mathbb{A}_k\right\| \\
\leqslant& \left\|\widehat{\mathbb{A}}_k - \mathbb{A}_k\right\| + \left\|\widehat{\mathbf{A}}_0(\widehat{\mathbf{A}}_0^{\mathrm{T}}\widehat{\mathbf{A}}_0)^{-1}\widehat{\mathbf{A}}_0^{\mathrm{T}}\mathbb{A}_k - \mathbf{A}_0(\mathbf{A}_0^{\mathrm{T}}\mathbf{A}_0)^{-1}\mathbf{A}_0^{\mathrm{T}}\mathbb{A}_k\right\|. \qquad (\mathrm{S.80})
\end{aligned}$$

By Lemma S2.4(c), we have $\|\widehat{\mathbb{A}}_k - \mathbb{A}_k\| = O_p(k^{1-\alpha/2}) = O_p(1)$. We next show that the second term of (S.80) is bounded in probability. Note that

$$
\begin{aligned}
& \left\|\widehat{\mathbf{A}}_0(\widehat{\mathbf{A}}_0^{\mathrm{T}}\widehat{\mathbf{A}}_0)^{-1}\widehat{\mathbf{A}}_0^{\mathrm{T}}\mathbb{A}_k - \mathbf{A}_0(\mathbf{A}_0^{\mathrm{T}}\mathbf{A}_0)^{-1}\mathbf{A}_0^{\mathrm{T}}\mathbb{A}_k\right\| \\
\leqslant\ & \left\|\left(\widehat{\mathbf{A}}_0 - \mathbf{A}_0\right)(\widehat{\mathbf{A}}_0^{\mathrm{T}}\widehat{\mathbf{A}}_0)^{-1}\widehat{\mathbf{A}}_0^{\mathrm{T}}\mathbb{A}_k\right\| \\
& + \left\|\mathbf{A}_0\left[(\widehat{\mathbf{A}}_0^{\mathrm{T}}\widehat{\mathbf{A}}_0)^{-1} - (\mathbf{A}_0^{\mathrm{T}}\mathbf{A}_0)^{-1}\right]\widehat{\mathbf{A}}_0^{\mathrm{T}}\mathbb{A}_k\right\| \\
& + \left\|\mathbf{A}_0(\mathbf{A}_0^{\mathrm{T}}\mathbf{A}_0)^{-1}\left(\widehat{\mathbf{A}}_0 - \mathbf{A}_0\right)\mathbb{A}_k\right\| \\
=\ & \text{(I)+(II)+(III)}. \qquad \text{(S.81)}
\end{aligned}
$$

We now analyze each term of (S.81). We first consider the second term (II). For arbitrary matrix $\mathbf{M}$, let $\|\mathbf{M}\|_{\mathrm{F}}$ be its Frobenius norm. Note that for any conformable matrices matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}$, the inequality $\|\mathbf{AB}\|_{\mathrm{F}} \leqslant \|\mathbf{A}\|\|\mathbf{B}\|_{\mathrm{F}}$ and $\|\mathbf{ABC}\|_{\mathrm{F}} \leqslant \|\mathbf{A}\|\|\mathbf{B}\|_{\mathrm{F}}\|\mathbf{C}\|$ hold. Then, by combining Condition (C8), Lemma S2.4(c), Lemma S2.5 and part (a), we obtain that

$$
\begin{aligned}
& \left\|\left(\widehat{\mathbf{A}}_0^{\mathrm{T}}\widehat{\mathbf{A}}_0/n\right)^{-1} - \left(\mathbf{A}_0^{\mathrm{T}}\mathbf{A}_0/n\right)^{-1}\right\|_{\mathrm{F}} \\
=\ & \left\|-\left(\mathbf{A}_0^{\mathrm{T}}\mathbf{A}_0/n\right)^{-1}\left[\frac{1}{\sqrt{n}}\left(\widehat{\mathbf{A}}_0 - \mathbf{A}_0\right)^{\mathrm{T}}\left(\frac{1}{\sqrt{n}}\mathbf{A}_0\right)\right.\right. \\
& + \left(\frac{1}{\sqrt{n}}\mathbf{A}_0^{\mathrm{T}}\right)\frac{1}{\sqrt{n}}\left(\widehat{\mathbf{A}}_0 - \mathbf{A}_0\right) \\
& \left.\left. + \frac{1}{\sqrt{n}}\left(\widehat{\mathbf{A}}_0 - \mathbf{A}_0\right)^{\mathrm{T}}\frac{1}{\sqrt{n}}\left(\widehat{\mathbf{A}}_0 - \mathbf{A}_0\right)\right]\left(\widehat{\mathbf{A}}_0^{\mathrm{T}}\widehat{\mathbf{A}}_0/n\right)^{-1}\right\|_{\mathrm{F}} \\
\leqslant\ & 2\left\|\left(\mathbf{A}_0^{\mathrm{T}}\mathbf{A}_0/n\right)^{-1}\right\|\left\|\frac{1}{\sqrt{n}}\left(\widehat{\mathbf{A}}_0 - \mathbf{A}_0\right)^{\mathrm{T}}\left(\frac{1}{\sqrt{n}}\mathbf{A}_0\right)\right\|_{\mathrm{F}}\left\|\left(\widehat{\mathbf{A}}_0^{\mathrm{T}}\widehat{\mathbf{A}}_0/n\right)^{-1}\right\| \\
& + \left\|\left(\mathbf{A}_0^{\mathrm{T}}\mathbf{A}_0/n\right)^{-1}\right\|\left\|\frac{1}{\sqrt{n}}\left(\widehat{\mathbf{A}}_0 - \mathbf{A}_0\right)^{\mathrm{T}}\frac{1}{\sqrt{n}}\left(\widehat{\mathbf{A}}_0 - \mathbf{A}_0\right)\right\|_{\mathrm{F}}\left\|\left(\widehat{\mathbf{A}}_0^{\mathrm{T}}\widehat{\mathbf{A}}_0/n\right)^{-1}\right\| \\
\leqslant\ & 2\lambda_{\min}^{-1}\left(\mathbf{A}_0^{\mathrm{T}}\mathbf{A}_0/n\right)\left\|\frac{1}{\sqrt{n}}\left(\widehat{\mathbf{A}}_0^{\mathrm{T}} - \mathbf{A}_0\right)^{\mathrm{T}}\right\|_{\mathrm{F}}\lambda_{\max}^{1/2}\left(\mathbf{A}_0\mathbf{A}_0^{\mathrm{T}}/n\right)\lambda_{\min}^{-1}\left(\widehat{\mathbf{A}}_0^{\mathrm{T}}\widehat{\mathbf{A}}_0/n\right) \\
& + \lambda_{\min}^{-1}\left(\mathbf{A}_0^{\mathrm{T}}\mathbf{A}_0/n\right)\left\|\frac{1}{\sqrt{n}}\left(\widehat{\mathbf{A}}_0^{\mathrm{T}} - \mathbf{A}_0\right)^{\mathrm{T}}\right\|_{\mathrm{F}}\lambda_{\max}^{1/2}\left\{\frac{1}{n}\left(\widehat{\mathbf{A}}_0 - \mathbf{A}_0\right)^{\mathrm{T}}\left(\widehat{\mathbf{A}}_0 - \mathbf{A}_0\right)\right\}\lambda_{\min}^{-1}\left(\widehat{\mathbf{A}}_0^{\mathrm{T}}\widehat{\mathbf{A}}_0/n\right) \\
=\ & 2\lambda_{\min}^{-1}\left(\mathbf{A}_0^{\mathrm{T}}\mathbf{A}_0/n\right)\left\|\frac{1}{\sqrt{n}}\left(\widehat{\mathbf{A}}_0^{\mathrm{T}} - \mathbf{A}_0\right)^{\mathrm{T}}\right\|_{\mathrm{F}}\lambda_{\max}^{1/2}\left(\mathbf{A}_0\mathbf{A}_0^{\mathrm{T}}/n\right)\lambda_{\min}^{-1}\left(\mathbf{A}_0^{\mathrm{T}}\mathbf{A}_0/n\right)\{1 + o_p(1)\}
\end{aligned}
$$

$$
\begin{aligned}
& +\lambda_{\min}^{-1}\left(\mathbf{A}_0^{\mathrm{T}}\mathbf{A}_0/n\right)\left\|\frac{1}{\sqrt{n}}\left(\widehat{\mathbf{A}}_0^{\mathrm{T}}-\mathbf{A}_0\right)^{\mathrm{T}}\right\|_{\mathrm{F}}\lambda_{\max}^{1/2}\left\{\frac{1}{n}\left(\widehat{\mathbf{A}}_0-\mathbf{A}_0\right)^{\mathrm{T}}\left(\widehat{\mathbf{A}}_0-\mathbf{A}_0\right)\right\} \\
& \times\lambda_{\min}^{-1}\left(\mathbf{A}_0^{\mathrm{T}}\mathbf{A}_0/n\right)\{1+o_p(1)\} \\
= & \lambda_{\min}^{-2}\left(\mathbf{A}_0^{\mathrm{T}}\mathbf{A}_0/n\right)\left\{\frac{1}{n}\sum_{i=1}^{n}\sum_{s=1}^{K}(\widehat{\xi}_{is}-\xi_{is})^2\right\}^{1/2}\left[2\lambda_{\max}^{1/2}\left(\mathbf{A}_0\mathbf{A}_0^{\mathrm{T}}/n\right)\right. \\
& \left.+\lambda_{\max}^{1/2}\left\{\frac{1}{n}\left(\widehat{\mathbf{A}}_0-\mathbf{A}_0\right)^{\mathrm{T}}\left(\widehat{\mathbf{A}}_0-\mathbf{A}_0\right)\right\}\right]\{1+o_p(1)\} \\
= & O_p\left(n^{-1/2}K^{3/2-\alpha/2}\right)+o_p\left(n^{-1/2}K^{3/2-\alpha/2}\right) \\
= & O_p\left(n^{-1/2}K^{3/2-\alpha/2}\right), \qquad \text{(S.82)}
\end{aligned}
$$

and by Central Limit Theorem, we have

$$
\left\|\frac{1}{\sqrt{n}}\mathbf{A}_0^{\mathrm{T}}\mathbb{A}_k\right\|=\left\{\sum_{s=1}^{K}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_{is}\xi_{ik}\right)^2\right\}^{1/2}=O_p(K^{1/2}). \qquad \text{(S.83)}
$$

Consequently, by part (i) in (b), (S.82) and (S.83), we have

$$
\begin{aligned}
& \text{(II)} \\
= & \left\|\mathbf{A}_0\left[(\widehat{\mathbf{A}}_0^{\mathrm{T}}\widehat{\mathbf{A}}_0)^{-1}-(\mathbf{A}_0^{\mathrm{T}}\mathbf{A}_0)^{-1}\right]\widehat{\mathbf{A}}_0^{\mathrm{T}}\mathbb{A}_k\right\| \\
\leqslant & \left\|\frac{1}{\sqrt{n}}\mathbf{A}_0\right\|\left\|(\widehat{\mathbf{A}}_0^{\mathrm{T}}\widehat{\mathbf{A}}_0/n)^{-1}-(\mathbf{A}_0^{\mathrm{T}}\mathbf{A}_0/n)^{-1}\right\|_{\mathrm{F}}\left|\frac{1}{\sqrt{n}}\widehat{\mathbf{A}}_0^{\mathrm{T}}\mathbb{A}_k\right\| \\
\leqslant & \lambda_{\max}^{1/2}\left(\mathbf{A}_0\mathbf{A}_0^{\mathrm{T}}/n\right)\left\|(\widehat{\mathbf{A}}_0^{\mathrm{T}}\widehat{\mathbf{A}}_0/n)^{-1}-(\mathbf{A}_0^{\mathrm{T}}\mathbf{A}_0/n)^{-1}\right\|_{\mathrm{F}}\left(\left\|\frac{1}{\sqrt{n}}\left(\widehat{\mathbf{A}}_0^{\mathrm{T}}-\mathbf{A}_0^{\mathrm{T}}\right)\mathbb{A}_k\right\|+\left\|\frac{1}{\sqrt{n}}\mathbf{A}_0^{\mathrm{T}}\mathbb{A}_k\right\|\right) \\
= & O_p\left(n^{-1/2}K^{3/2-\alpha/2}\right)\left\{O_p(n^{-1/4}K^{3/2-\alpha/2})+O_p(K^{1/2})\right\} \\
= & O_p(n^{-3/4}K^{3-\alpha})+O_p(n^{-1/2}K^{2-\alpha/2}) \\
= & O_p(n^{-1/2}K^{2-\alpha/2}) \\
= & o_p(1),
\end{aligned}
$$

where the last equality follows from Condition (C5). Under the same condition, for the third term (III) of (S.81), we have

$$
\begin{aligned}
\text{(III)} &= \left\| \mathbf{A}_0(\mathbf{A}_0^{\mathrm{T}}\mathbf{A}_0)^{-1}\left(\widehat{\mathbf{A}}_0-\mathbf{A}_0\right)\mathbb{A}_k \right\| \\
&= \left\| \frac{1}{\sqrt{n}}\mathbf{A}_0\left(\mathbf{A}_0^{\mathrm{T}}\mathbf{A}_0/n\right)^{-1}\frac{1}{\sqrt{n}}\left(\widehat{\mathbf{A}}_0-\mathbf{A}_0\right)\mathbb{A}_k \right\| \\
&\leqslant \left\| \frac{1}{\sqrt{n}}\mathbf{A}_0 \right\| \left\| \left(\mathbf{A}_0^{\mathrm{T}}\mathbf{A}_0/n\right)^{-1} \right\|_{\mathrm{F}} \left\| \frac{1}{\sqrt{n}}\left(\widehat{\mathbf{A}}_0-\mathbf{A}_0\right)\mathbb{A}_k \right\| \\
&= \lambda_{\min}^{-1}\left(\mathbf{A}_0^{\mathrm{T}}\mathbf{A}_0/n\right)\left\{\sum_{i=1}^{n}\sum_{s=1}^{K}\left(\frac{1}{\sqrt{n}}\xi_{is}\right)^2\right\}^{1/2} O_p(n^{-1/4}K^{3/2-\alpha/2}) \\
&= O_p(1)O_p(K^{1/2})O_p(n^{-1/4}K^{3/2-\alpha/2}) \\
&= O_p(n^{-1/4}K^{2-\alpha/2}) \\
&= o_p(1).
\end{aligned}
$$

Similarly, for the first term (I) in (S.81), we have

$$
\begin{aligned}
\text{(I)} &= \left\| \frac{1}{\sqrt{n}}\left(\widehat{\mathbf{A}}_0-\mathbf{A}_0\right)(\widehat{\mathbf{A}}_0^{\mathrm{T}}\widehat{\mathbf{A}}_0/n)^{-1}\frac{1}{\sqrt{n}}\widehat{\mathbf{A}}_0^{\mathrm{T}}\mathbb{A}_k \right\| \\
&\leqslant \left\| \frac{1}{\sqrt{n}}(\widehat{\mathbf{A}}_0-\mathbf{A}_0) \right\|_{\mathrm{F}} \left\| (\widehat{\mathbf{A}}_0^{\mathrm{T}}\widehat{\mathbf{A}}_0/n)^{-1} \right\| \left\| \frac{1}{\sqrt{n}}\widehat{\mathbf{A}}_0^{\mathrm{T}}\mathbb{A}_k \right\| \\
&\leqslant \frac{1}{\sqrt{n}}\left[\sum_{i=1}^{n}\sum_{s=1}^{K}(\widehat{\xi}_{is}-\xi_{is})^2\right]^{1/2}\lambda_{\min}^{-1}\left(\mathbf{A}_0^{\mathrm{T}}\mathbf{A}_0/n\right)\left(\left\| \frac{1}{\sqrt{n}}\left(\widehat{\mathbf{A}}_0^{\mathrm{T}}-\mathbf{A}_0^{\mathrm{T}}\right)\mathbb{A}_k \right\| + \left\| \frac{1}{\sqrt{n}}\mathbf{A}_0^{\mathrm{T}}\mathbb{A}_k \right\|\right) \\
&= O_p(n^{-1/2}K^{3/2-\alpha/2})O_p(1)\left\{O_p(n^{-1/4}K^{3/2-\alpha/2})+O_p(K^{1/2})\right\} \\
&= O_p(n^{-1/2}K^{2-\alpha/2}) \\
&= o_p(1),
\end{aligned}
$$

where the third last equality follows from Lemma S2.4(c), Condition (C8), part (i) of (b) and (S.83). Therefore, based on the above results, along with (S.80) and (S.81), we conclude the proof of part (b), which completes the proof of Lemma S2.10. □

S3.15 Proof of Theorem 3

With a slight abuse of notation, we denote the true regression coefficient under the full model by $\mathbf{b}^0 = (b_1^0, \ldots, b_K^0)^{\mathrm{T}}$ and let $\widetilde{\mathbf{b}}$ denote its estimator. From (S.33), we have $\|\widetilde{\mathbf{b}} - \mathbf{b}^0\|^2 = O_p(K^{2\alpha+3}n^{-1})$. Next, for any arbitrary candidate model, define

$$\widehat{\mathbf{b}}_{K*} = \operatorname{argmin}_{\{\mathbf{b}\in\mathbb{R}^K: b_k=0, \forall k>K*\}}\|\widehat{\mathbf{S}}\mathbf{Y} - \widehat{\mathbf{A}}\widehat{\mathbf{b}}\|^2.$$

We then have

$$\begin{aligned}\min_{K*<K_0}\|\widehat{\mathbf{b}}_{K*} - \widetilde{\mathbf{b}}\|^2 &\geqslant \min_{K*<K_0}\{\|\widehat{\mathbf{b}}_{K*} - \mathbf{b}^0\|^2\} - \|\widetilde{\mathbf{b}} - \mathbf{b}^0\|^2 \\ &\geqslant \min_{1\leqslant k\leqslant K_0}({b_k^0}^2) - O_p(K^{2\alpha+3}n^{-1}). \end{aligned} \tag{S.84}$$

By Condition (C10), we know that the right-hand side of the above inequality is guaranteed to be positive with probability tending to 1. Next, we consider

$$\begin{aligned} &\min_{K*<K_0}(\mathrm{BIC}_{K*} - \mathrm{BIC}_K) \\ = &\min_{K*<K_0}\left\{\log\left(\frac{\widehat{\sigma}_{K*}^2}{\widehat{\sigma}_K^2}\right) + (K^* - K)\frac{\log(n)}{n}C_n^*\right\} \\ \geqslant &\min_{K*<K_0}\left\{\log\left(\frac{\widehat{\sigma}_{K*}^2}{\widehat{\sigma}_K^2}\right)\right\} - \frac{K\log(n)}{n}C_n^*. \end{aligned}$$

Note that $\log(1+x) \geqslant \min\{x/2, \log(2)\}$ for any $x > 0$. Recalling that $n^{-1}\widehat{\mathbf{A}}_{K*}^{\mathrm{T}}\widehat{\mathbf{A}}_{K*} = \widehat{\mathbf{\Lambda}}$ and $n^{-1}\mathbf{A}_{K*}^{\mathrm{T}}\mathbf{A}_{K*} = \mathbf{\Lambda}$, and following Wang et al. (2009), the right-hand side of the above equation can be written as

$$\begin{aligned} &\min_{K*<K_0}\left[\log\left\{1 + \frac{(\widehat{\mathbf{b}}_{K*} - \widetilde{\mathbf{b}})^{\mathrm{T}}\widehat{\mathbf{\Lambda}}(\widehat{\mathbf{b}}_{K*} - \widetilde{\mathbf{b}})}{\widehat{\sigma}_K^2}\right\}\right] - \frac{K\log(n)}{n}C_n^* \\ \geqslant &\min_{K*<K_0}\left\{\log\left(1 + \frac{\widehat{\lambda}_{\min}\|\widehat{\mathbf{b}}_{K*} - \widetilde{\mathbf{b}}\|^2}{\widehat{\sigma}_K^2}\right)\right\} - \frac{K\log(n)}{n}C_n^* \end{aligned}$$

$$\geqslant \min_{K^* < K_0} \min\left\{\log(2), \frac{\widehat{\lambda}_{\min}\|\widehat{\mathbf{b}}_{K^*} - \widetilde{\mathbf{b}}\|^2}{2\widehat{\sigma}_K^2}\right\} - \frac{K\log(n)}{n}C_n^*, \tag{S.85}$$

where $\widehat{\lambda}_{\min} \equiv \lambda_{\min}(\widehat{\mathbf{\Lambda}})$. Because $K\log(n)/n \to 0$ under Condition (C10), thus with probability tending to 1 we must have $\log(2) - C_n^* K^{2\alpha+3}\log(n)/n > 0$. Consequently, as long as we can show that, with probability tending to 1,

$$\min_{K^* < K_0}\left\{\frac{\widehat{\lambda}_{\min}\|\widehat{\mathbf{b}}_{K^*} - \widetilde{\mathbf{b}}\|^2}{2\widehat{\sigma}_K^2}\right\} - \frac{K\log(n)}{n}C_n^* \tag{S.86}$$

is positive, we know that the right-hand side of expression (S.85) is positive asymptotically. Note that $\widehat{\sigma}_K^2 \xrightarrow{\mathrm{P}} \sigma_K^2$ derived from the normality assumption (Wang et al., 2009), where $\xrightarrow{\mathrm{P}}$ denotes the convergence in probability. Furthermore, by Lemma S2.10(a), we have $|\lambda_{\min}(\widehat{\mathbf{\Lambda}}) - \lambda_{\min}(\mathbf{\Lambda})| = o_p(1)$, from which we know that $\widehat{\lambda}_{\min} \xrightarrow{\mathrm{P}} \lambda_{\min} = \lambda_{\min}(\mathbf{\Lambda})$. Applying inequality (S.84) to expression (S.86), we find that (S.86) can be further bounded by

$$\begin{aligned}
&\geqslant \frac{\lambda_{\min}}{2\sigma_K^2}\left\{\min_{1\leqslant k\leqslant K_0}({b_k^0}^2) - O_p(K^{2\alpha+3}n^{-1})\right\}\{1 + o_p(1)\} - \frac{K\log(n)}{n}C_n^* \\
&= \frac{C_n^* K^{2\alpha+3}\log(n)}{n}\frac{\lambda_{\min}}{2\sigma_K^2}\left\{\frac{n}{C_n^* K^{2\alpha+3}\log(n)}\min_{1\leqslant k\leqslant K_0}({b_k^0}^2)\right\}\{1 + o_p(1)\} - \frac{K\log(n)}{n}C_n^* + O_p(K^{2\alpha+3}n^{-1}),
\end{aligned}$$

which is guaranteed to be positive asymptotically under Condition (C10). This proves that, with probability tending to 1, the right hand side of (S.85) must be positive. This completes the proof. □

S3.16 Proof of Theorem 4

Consider an arbitrary overfitted model, i.e., $K^* > K_0$. Note that the residual sum of squares in the overfitted case can be written as

$$\begin{aligned}\mathrm{SSE}_{K*} &= \inf_{\mathbf{b}_{K*}}(\|\widehat{\mathbf{S}}\mathbf{Y} - \widehat{\mathbf{A}}_{K*}\mathbf{b}_{K*}\|^2) \\ &= \inf_{\mathbf{b}_{K_0},\mathbf{b}_{K_0^c}}(\|\widehat{\mathbf{S}}\mathbf{Y} - \widehat{\mathbf{A}}_0\mathbf{b}_{K_0} - \widehat{\mathbf{A}}_0^c\mathbf{b}_{K_0^c}\|^2),\end{aligned}$$

where $\mathbf{b}_{K_0^c} = (b_{K_0+1}, \ldots, b_{K^*})^{\mathrm{T}}$. It is easy to see that

$$\begin{aligned}\mathrm{SSE}_{K_0} &= \|\widehat{\mathbf{S}}\mathbf{Y} - \widehat{\mathbf{A}}_0\widehat{\mathbf{b}}_{K_0}\|^2 \\ &= \|\widehat{\mathbf{S}}\mathbf{Y} - \widehat{\mathbf{A}}_0(\widehat{\mathbf{A}}_0^{\mathrm{T}}\widehat{\mathbf{A}}_0)^{-1}\widehat{\mathbf{A}}_0^{\mathrm{T}}\widehat{\mathbf{S}}\mathbf{Y}\|^2 \\ &= \|(\mathbf{I} - \widehat{\mathbf{A}}_0(\widehat{\mathbf{A}}_0^{\mathrm{T}}\widehat{\mathbf{A}}_0)^{-1}\widehat{\mathbf{A}}_0^{\mathrm{T}})\widehat{\mathbf{S}}\mathbf{Y}\|^2 \\ &= \|\widehat{\mathbf{Q}}_0\widehat{\mathbf{S}}\mathbf{Y}\|^2,\end{aligned}$$

where $\widehat{\mathbf{Q}}_0$ is defined below equation (3.20). For an arbitrary matrix $\mathbf{M}$, we use span($\mathbf{M}$) to denote the linear subspace that is spanned by the column vector of $\mathbf{M}$. According to the results in Wang et al. (2009), we have $\mathrm{span}(\widehat{\mathbf{A}}_0, \widehat{\mathbf{A}}_0^c) = \mathrm{span}(\widehat{\mathbf{A}}_0, \widehat{\mathbf{A}}_{0^c}^{\mathrm{proj}})$, where $\widehat{\mathbf{A}}_{0^c}^{\mathrm{proj}} = \widehat{\mathbf{Q}}_0\widehat{\mathbf{A}}_0^c$ denotes the projection of $\widehat{\mathbf{A}}_0^c$ onto the orthogonal complement of $\mathrm{span}(\widehat{\mathbf{A}}_0)$. Wang et al. (2009) proves that

$$\mathrm{SSE}_{K*} = \inf_{\mathbf{b}_{K_0},\mathbf{b}_{K_0^c}}(\|\widehat{\mathbf{S}}\mathbf{Y} - \widehat{\mathbf{A}}_0\mathbf{b}_{K_0} - \widehat{\mathbf{A}}_{0^c}^{\mathrm{proj}}\mathbf{b}_{K_0^c}\|^2).$$

Further, the minimizer of the above optimization problem is given by $\widehat{\mathbf{b}}_{K_0} = (\widehat{\mathbf{A}}_0^{\mathrm{T}}\widehat{\mathbf{A}}_0)^{-1}\widehat{\mathbf{A}}_0^{\mathrm{T}}\widehat{\mathbf{S}}\mathbf{Y}$ and $\widehat{\mathbf{b}}_{K_0^c} = (\widehat{\mathbf{A}}_{0^c}^{\mathrm{projT}}\widehat{\mathbf{A}}_{0^c}^{\mathrm{proj}})^{-1}\widehat{\mathbf{A}}_{0^c}^{\mathrm{projT}}\widehat{\boldsymbol{\varepsilon}}$, where $\widehat{\boldsymbol{\varepsilon}} = \widehat{\mathbf{S}}\mathbf{Y} - \widehat{\mathbf{A}}_0\widehat{\mathbf{b}}_{K_0} = \widehat{\mathbf{Q}}_0\widehat{\mathbf{S}}\mathbf{Y}$ is an estimator of $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)^{\mathrm{T}} \in \mathbb{R}^n$. Recall that $\widehat{\mathbb{A}}_k$ denotes the kth column of the score matrix $\widehat{\mathbf{A}}_{K*}$ for

$k = K_0 + 1, \ldots, K^*$. Then, we obtain

$$
\begin{aligned}
& \mathrm{SSE}_{K_0} - \mathrm{SSE}_{K*} \\
= \ & \|\widehat{\mathbf{Q}}_0 \widehat{\mathbf{S}} \mathbf{Y}\|^2 - \|\widehat{\mathbf{Q}}_0 \widehat{\mathbf{S}} \mathbf{Y} - \widehat{\mathbf{A}}_{0^c}^{\mathrm{proj}} \mathbf{b}_{K_0^c}\|^2 \\
= \ & \mathbf{Y}^{\mathrm{T}} \widehat{\mathbf{S}}^{\mathrm{T}} \widehat{\mathbf{Q}}_0^{\mathrm{T}} \widehat{\mathbf{A}}_{0^c}^{\mathrm{proj}} (\widehat{\mathbf{A}}_{0^c}^{\mathrm{projT}} \widehat{\mathbf{A}}_{0^c}^{\mathrm{proj}})^{-1} \widehat{\mathbf{A}}_{0^c}^{\mathrm{projT}} \widehat{\boldsymbol{\varepsilon}} + \widehat{\boldsymbol{\varepsilon}}^{\mathrm{T}} \widehat{\mathbf{A}}_{0^c}^{\mathrm{proj}} (\widehat{\mathbf{A}}_{0^c}^{\mathrm{projT}} \widehat{\mathbf{A}}_{0^c}^{\mathrm{proj}})^{-1} \widehat{\mathbf{A}}_{0^c}^{\mathrm{projT}} \widehat{\mathbf{Q}}_0 \widehat{\mathbf{S}} \mathbf{Y} \\
& - \widehat{\boldsymbol{\varepsilon}}^{\mathrm{T}} \widehat{\mathbf{A}}_{0^c}^{\mathrm{proj}} (\widehat{\mathbf{A}}_{0^c}^{\mathrm{projT}} \widehat{\mathbf{A}}_{0^c}^{\mathrm{proj}})^{-1} \widehat{\mathbf{A}}_{0^c}^{\mathrm{projT}} \widehat{\boldsymbol{\varepsilon}} \\
= \ & (n^{-1/2} \widehat{\boldsymbol{\varepsilon}}^{\mathrm{T}} \widehat{\mathbf{A}}_{0^c}^{\mathrm{proj}}) (n^{-1} \widehat{\mathbf{A}}_{0^c}^{\mathrm{projT}} \widehat{\mathbf{A}}_{0^c}^{\mathrm{proj}})^{-1} (n^{-1/2} \widehat{\mathbf{A}}_{0^c}^{\mathrm{proj}} \widehat{\boldsymbol{\varepsilon}}) \\
\leqslant \ & \tau_{\max}^{0^c} \|n^{-1/2} \widehat{\boldsymbol{\varepsilon}}^{\mathrm{T}} \widehat{\mathbf{A}}_{0^c}^{\mathrm{proj}}\|^2 \\
= \ & \tau_{\max}^{0^c} \sum_{k=K_0+1}^{K*} (n^{-1/2} \widehat{\boldsymbol{\varepsilon}}^{\mathrm{T}} \widehat{\mathbb{A}}_k^{\mathrm{proj}})^2 \\
\leqslant \ & \tau_{\max}^{0^c} \max_{K_0+1 \leqslant k \leqslant K*} (n^{-1/2} \widehat{\boldsymbol{\varepsilon}}^{\mathrm{T}} \widehat{\mathbb{A}}_k^{\mathrm{proj}})^2 (K^* - K_0) \\
\leqslant \ & \tau_{\max}^{0^c} \max_{K_0+1 \leqslant k \leqslant K} (n^{-1/2} \widehat{\boldsymbol{\varepsilon}}^{\mathrm{T}} \widehat{\mathbb{A}}_k^{\mathrm{proj}})^2 (K^* - K_0),
\end{aligned}
$$

where $\tau_{\max}^{0^c} = \lambda_{\min}^{-1}(n^{-1} \widehat{\mathbf{A}}_{0^c}^{\mathrm{projT}} \widehat{\mathbf{A}}_{0^c}^{\mathrm{proj}})$ and the last inequality is because $K^* \leqslant K$. Also, we have $\lambda_{\min}(n^{-1} \widehat{\mathbf{A}}_{0^c}^{\mathrm{projT}} \widehat{\mathbf{A}}_{0^c}^{\mathrm{proj}}) \geqslant \lambda_{\min}(n^{-1} \widehat{\mathbf{A}}_{0^c,K}^{\mathrm{projT}} \widehat{\mathbf{A}}_{0^c,K}^{\mathrm{proj}}) \equiv (\tau_{\max}^{0^{c_K}})^{-1}$, where $\widehat{\mathbf{A}}_{0^c,K}^{\mathrm{proj}}$ is the same as $\widehat{\mathbf{A}}_{0^c}^{\mathrm{proj}}$ defined in earlier, except that $\widehat{\mathbf{A}}_0^c$ is replaced by $\widehat{\mathbf{A}}_0^{c_K}$. Based on the above results, we have

$$
\max_{K_0+1 \leqslant k \leqslant K*} \left(\frac{\mathrm{SSE}_{K_0} - \mathrm{SSE}_{K*}}{K^* - K_0} \right) \ \leqslant \ \tau_{\max}^{0^{c_K}} \max_{K_0+1 \leqslant k \leqslant K} \left\{ (n^{-1/2} \widehat{\boldsymbol{\varepsilon}}^{\mathrm{T}} \widehat{\mathbb{A}}_k^{\mathrm{proj}})^2 \right\}.
$$

By Cauchy-Schwarz inequality, we know that

$$
\begin{aligned}
& \left(n^{-1/2} \widehat{\boldsymbol{\varepsilon}}^{\mathrm{T}} \widehat{\mathbb{A}}_k^{\mathrm{proj}} \right)^2 \\
\leqslant \ & 2 \left(n^{-1/2} \widehat{\boldsymbol{\varepsilon}}^{\mathrm{T}} \mathbb{A}_k^{\mathrm{proj}} \right)^2 + 2 \left(n^{-1/2} \widehat{\boldsymbol{\varepsilon}}^{\mathrm{T}} \widehat{\mathbb{A}}_k^{\mathrm{proj}} - n^{-1/2} \widehat{\boldsymbol{\varepsilon}}^{\mathrm{T}} \mathbb{A}_k^{\mathrm{proj}} \right)^2 \\
\leqslant \ & 2 \left(n^{-1/2} \widehat{\boldsymbol{\varepsilon}}^{\mathrm{T}} \mathbb{A}_k^{\mathrm{proj}} \right)^2 + 2 \|n^{-1/2} \widehat{\boldsymbol{\varepsilon}}^{\mathrm{T}}\|^2 \|\widehat{\mathbb{A}}_k^{\mathrm{proj}} - \mathbb{A}_k^{\mathrm{proj}}\|^2
\end{aligned}
$$

$$= \quad 2\left(n^{-1/2}\widehat{\boldsymbol{\varepsilon}}^{\mathrm{T}}\mathbb{A}_k^{\mathrm{proj}}\right)^2 + O_p(1),$$

where the last equality is because $\|n^{-1/2}\widehat{\boldsymbol{\varepsilon}}^{\mathrm{T}}\|^2 = O_p(1)$ following Wang et al. (2009) and $\|\widehat{\mathbb{A}}_k^{\mathrm{proj}} - \mathbb{A}_k^{\mathrm{proj}}\|^2 = O_p(1)$ established in Lemma S2.10(b). Under the normality assumption and following the proof by Wang et al. (2009), we have

$$\max_{K_0+1\leqslant k\leqslant K^*}\left(\frac{\mathrm{SSE}_{K_0} - \mathrm{SSE}_{K^*}}{K^* - K_0}\right) \leqslant 2(1+\varphi)\sigma_{K_0}^2\kappa^{-1}\log(K) + O_p(1), \tag{S.87}$$

where φ is an arbitrary positive but fixed constant, and κ is a constant. By Condition (C5), we know that the number of principal components also satisfies that

$$\limsup_{n\to\infty}(K/n^{\kappa^*}) < 1 \quad \text{for some} \quad \kappa^* < 1. \tag{S.88}$$

By (S.87), (S.88) and the fact that $K^* \leqslant K$, we obtain $\max_{K_0+1\leqslant k\leqslant K^*}(\mathrm{SSE}_{K_0} - \mathrm{SSE}_{K^*}) = O_p(K\log K) = o_p(n)$. Thus, the Taylor series expansion leads to

$$\begin{aligned}
& n\left(\mathrm{BIC}_{K^*} - \mathrm{BIC}_{K_0}\right) \\
= \;& n\left\{\log\left(\frac{\widehat{\sigma}_{K^*}^2}{\widehat{\sigma}_{K_0}^2}\right) + (K^* - K_0)\frac{\log(n)}{n}C_n^*\right\} \\
= \;& n\left\{\log\left(\frac{\mathrm{SSE}_{K_0} + \mathrm{SSE}_{K^*} - \mathrm{SSE}_{K_0}}{\mathrm{SSE}_{K_0}}\right) + (K^* - K_0)\frac{\log(n)}{n}C_n^*\right\} \\
= \;& n\log\left\{1 + \frac{n^{-1}\left(\mathrm{SSE}_{K^*} - \mathrm{SSE}_{K_0}\right)}{\widehat{\sigma}_{K_0}^2}\right\} + (K^* - K_0)\log(n)C_n^* \\
= \;& \frac{1}{\widehat{\sigma}_{K_0}^2}\left(\mathrm{SSE}_{K^*} - \mathrm{SSE}_{K_0}\right) + (K^* - K_0)\log(n)C_n^* + O_p\{(K\log K)^2 n^{-1}\} \\
= \;& \frac{1}{\sigma_{K_0}^2}\left(\mathrm{SSE}_{K^*} - \mathrm{SSE}_{K_0}\right)\{1 + o_p(1)\} + (K^* - K_0)\log(n)C_n^* + o_p(1) \\
\geqslant \;& (K^* - K_0)\left[\left\{-2(1+\varphi)\kappa^{-1}\log(K) + O_p(1)\right\}\{1 + o_p(1)\} + \log(n)C_n^*\right] + o_p(1),
\end{aligned}$$

where the last inequality is due to (S.87). Consequently, by (S.88), we know that

$$\begin{aligned}\max_{K_0+1\leqslant k\leqslant K^*}\left\{\frac{n\left(\mathrm{BIC}_{K^*}-\mathrm{BIC}_{K_0}\right)}{K^*-K_0}\right\} &\geqslant C_n^*\log(n)+\{-2(1+\varphi)\kappa^{-1}\log(K)+O_p(1)\}\{1+o_p(1)\}\\ &\geqslant \log(n)\left\{C_n^*-2(1+\varphi)\kappa^{-1}\kappa^*+O_p(1)\right\}\{1+o_p(1)\}\end{aligned}$$

with probability tending to 1. By Condition (C10), we know that $C_n^*\to\infty$. This implies that, with probability tending to 1, $\max_{K_0+1\leqslant k\leqslant K^*}(\mathrm{BIC}_{K^*}-\mathrm{BIC}_{K_0})$ must be positive. This completes the proof. □

S4 Additional simulation results

To examine the effect of the network autoregression coefficient on our proposed method, we also conduct the simulation studies when $\rho=0.3$ for a more comprehensive comparison. The simulation results are summarized in Tables S.1-S.4 and Figures S.1, respectively. These results are similar to those obtained when $\rho=0.1$.

In addition, as suggested by an anonymous reviewer, (1) we further assess the estimation performance under a near-boundary setting (i.e., $\rho=0.8$), with the corresponding simulation results reported in Tables S.5-S.6 and Figure S.2, respectively; and (2) we evaluate the empirical standard deviation and the asymptotic standard error of the network autoregression coefficient. Moreover, we consider not only the boundary case $\rho=0.8$ but also the moderate dependence case $\rho=0.5$. Specifically, the asymptotic standard error is denoted by $\widehat{\mathrm{SE}}_{\mathrm{asy}}^{(m)}$, defined as the square root of the estimated Σ in Theorem 2 for the mth simulation replication. The overall average asymptotic standard error is then computed as $\widehat{\mathrm{SE}}_{\mathrm{asy}}=M^{-1}\sum_{m=1}^{M}\widehat{\mathrm{SE}}_{\mathrm{asy}}^{(m)}$, where M denotes the number of simulation replications. Meanwhile, the empirical standard deviation is estimated by $\widehat{\mathrm{SE}}_{\mathrm{emp}}=\{M^{-1}\sum_{m=1}^{M}(\widehat{\rho}^{(m)}-\overline{\rho})^2\}^{1/2}$, where $\overline{\rho}=M^{-1}\sum_{m=1}^{M}\widehat{\rho}^{(m)}$.

The simulation setup follows that described in the simulation section, with the true network autoregression coefficient set to $\rho = 0.5$ and 0.8, respectively. The simulation results are presented in Table S.7. As expected, the ratios of $\widehat{\mathrm{SE}}_{\mathrm{emp}}$ to $\widehat{\mathrm{SE}}_{\mathrm{asy}}$ generally fluctuate around one across different network sizes and sparsity levels. This, to some extent, demonstrates the accuracy of our theoretical results.

Table S.1: Simulation results for $n = 200, 500$ with 100 replicates of the dyad independence model, stochastic block model and power-law distribution model respectively. The results are displayed when the error follows normal distribution and t-distribution in each scenario. The IMSE and SUP are both reported. The true number of principal components (# PC) ranges from 2 to 50, with $\rho = 0.3$.

Scenario	n	# PC: 2		# PC: 10		# PC: 50	
		IMSE	SUP	IMSE	SUP	IMSE	SUP
Case 1: Normal distribution							
I	200	5.85×10^{-6}	1.59×10^{-5}	6.82×10^{-4}	3.11×10^{-3}	1.03×10^{-2}	7.33×10^{-2}
	500	1.25×10^{-6}	3.89×10^{-6}	2.64×10^{-4}	1.07×10^{-3}	1.99×10^{-3}	1.34×10^{-2}
II	200	4.61×10^{-6}	1.32×10^{-5}	6.90×10^{-4}	2.96×10^{-4}	9.93×10^{-3}	6.75×10^{-2}
	500	1.53×10^{-6}	4.68×10^{-6}	2.29×10^{-4}	9.43×10^{-4}	1.88×10^{-3}	1.25×10^{-2}
III	200	4.75×10^{-6}	1.36×10^{-5}	7.63×10^{-4}	3.37×10^{-3}	2.32×10^{-3}	8.38×10^{-2}
	500	7.95×10^{-7}	2.85×10^{-6}	2.38×10^{-4}	9.28×10^{-4}	1.92×10^{-3}	1.38×10^{-2}
Case 2: t-distribution							
I	200	2.77×10^{-4}	5.77×10^{-4}	1.49×10^{-2}	6.93×10^{-2}	0.35	2.41
	500	9.87×10^{-5}	2.06×10^{-4}	5.78×10^{-3}	2.60×10^{-2}	0.15	0.99
II	200	2.29×10^{-4}	4.81×10^{-4}	1.67×10^{-2}	7.82×10^{-2}	0.35	2.33
	500	1.18×10^{-4}	2.45×10^{-4}	6.46×10^{-3}	2.89×10^{-2}	0.14	0.95
III	200	3.17×10^{-4}	6.58×10^{-4}	1.59×10^{-2}	7.33×10^{-2}	0.35	2.50
	500	1.00×10^{-4}	2.10×10^{-4}	6.24×10^{-3}	2.86×10^{-2}	0.15	1.07

Table S.2: Simulation results for $n = 200, 500$ with 100 replicates of the dyad independence model, stochastic block model and power-law distribution model respectively. The results are displayed when the error follows normal distribution and t-distribution in each scenario. The Bias and SD are both reported. The true number of principal components (# PC) ranges from 2 to 50, with $\rho = 0.3$.

Scenario	n	# PC: 2		# PC: 10		# PC: 50	
		Bias	SD	Bias	SD	Bias	SD
Case 1: Normal distribution							
I	200	-0.0708	0.0662	-0.0791	0.0695	-0.0698	0.0668
	500	-0.0398	0.0421	-0.0434	0.0409	-0.0450	0.0432
II	200	-0.0092	0.0127	-0.0059	0.0099	-0.0079	0.0105
	500	-0.0028	0.0055	-0.0026	0.0044	-0.0031	0.0057
III	200	-0.0130	0.0224	-0.0078	0.0192	-0.0070	0.0242
	500	-0.0041	0.0092	-0.0043	0.0108	-0.0042	0.0117
Case 2: t-distribution							
I	200	-0.0553	0.1595	-0.0425	0.1505	-0.0486	0.1253
	500	-0.0303	0.1023	-0.0144	0.0914	-0.0381	0.0824
II	200	-0.0049	0.0388	-0.0096	0.0351	-0.0061	0.0331
	500	0.0015	0.0228	0.0014	0.0220	-0.0048	0.0242
III	200	-0.0117	0.0514	-0.0129	0.0522	-0.0034	0.0534
	500	-0.0084	0.0320	-0.0078	0.0360	-0.0027	0.0323

Table S.3: The results are displayed as the error follows normal distribution in three scenarios and the sample size n is set as 200 and 500, respectively. Average numbers of K selected by four criteria as well as the corresponding IMSE in case of $\rho = 0.3$ are shown. We report the selected $\widehat{K}$ in the first line. We report the corresponding IMSE values in the second line (The original IMSE values multiplied by 10^3).

n	Scenario		AIC	BIC	BIC*	BIC**
200	I	$\widehat{K}$	8.48	6.96	**8.73**	8.94
		IMSE	1.80	2.02	**1.68**	1.92
	II	$\widehat{K}$	5.79	4.85	**6.17**	5.89
		IMSE	3.08	4.34	**2.71**	3.03
	III	$\widehat{K}$	5.88	4.72	**6.32**	6.02
		IMSE	2.89	4.30	**2.62**	2.79
500	I	$\widehat{K}$	**10.28**	8.34	10.08	9.29
		IMSE	**0.64**	0.97	0.65	0.75
	II	$\widehat{K}$	6.73	5.57	**6.91**	6.10
		IMSE	1.73	2.75	**1.61**	2.18
	III	$\widehat{K}$	7.09	5.89	**7.21**	6.51
		IMSE	1.55	2.34	**1.46**	1.86

Table S.4: The results are displayed as the error follows t distribution in three scenarios and the sample size n is set as 200 and 500, respectively. Average numbers of K selected by four criteria as well as the corresponding IMSE in case of $\rho = 0.3$ are shown. We report the selected $\widehat{K}$ in the first line. We report the corresponding IMSE values in the second line (The original IMSE values multiplied by 10^3).

n	Scenario		AIC	BIC	BIC*	BIC**
200	I	$\widehat{K}$	5.30	3.73	6.43	**4.30**
		IMSE	2.01	1.01	1.93	**0.96**
	II	$\widehat{K}$	4.92	3.57	5.76	**4.04**
		IMSE	2.31	1.12	1.92	**0.95**
	III	$\widehat{K}$	4.74	**3.58**	5.97	4.19
		IMSE	1.16	**1.05**	1.55	1.07
500	I	$\widehat{K}$	6.41	4.39	6.52	**5.06**
		IMSE	1.12	0.59	0.88	**0.52**
	II	$\widehat{K}$	5.63	4.16	6.14	**4.61**
		IMSE	0.69	0.68	0.66	**0.57**
	III	$\widehat{K}$	5.52	4.26	6.06	**4.87**
		IMSE	0.68	0.64	0.66	**0.63**

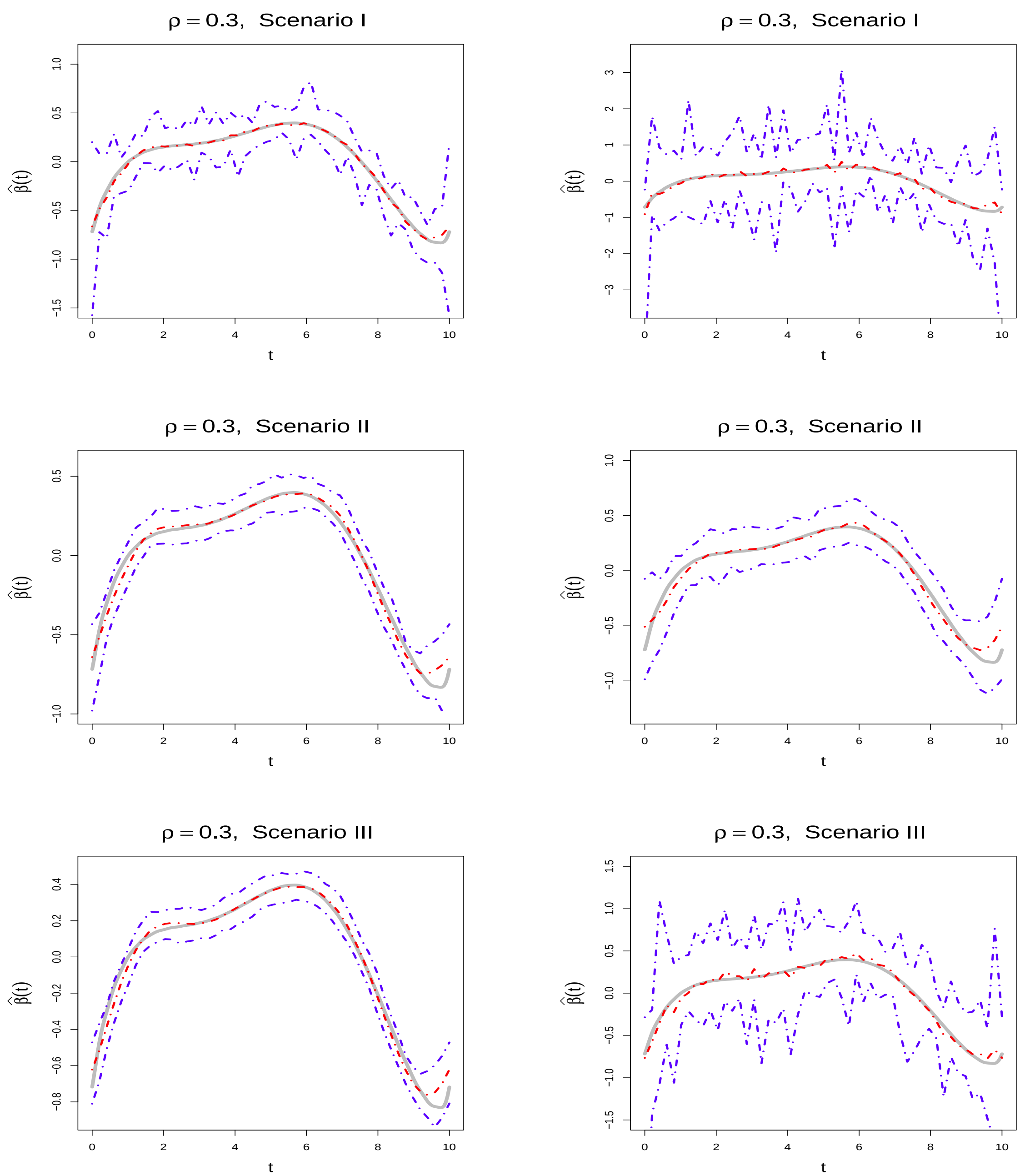

Figure S.1: The functional coefficients estimation $\widehat{\beta}(t)$ under normal distribution and t distribution, respectively, with sample size $n = 200$ and $\rho = 0.3$. The left panels contain the results under normal distribution, while the right panels those under t distribution. In each panel, the solid grey line is the true value, the dashed red line is the average estimated value, and the dashed blue lines are the pointwise 2.5% and 97.5% percentiles of the estimators based on 100 replications.

Table S.5: Simulation results for $n = 200, 500$ with 100 replicates of the dyad independence model, stochastic block model and power-law distribution model respectively. The results are displayed when the error follows normal distribution and t-distribution in each scenario. The IMSE and SUP are both reported. The true number of principal components (# PC) ranges from 2 to 50, with $\rho = 0.8$.

		# PC: 2		# PC: 10		# PC: 50	
Scenario	n	IMSE	SUP	IMSE	SUP	IMSE	SUP
Case 1: Normal distribution							
I	200	4.60×10^{-6}	1.32×10^{-5}	6.58×10^{-4}	3.00×10^{-3}	1.14×10^{-2}	8.07×10^{-2}
	500	9.71×10^{-7}	3.27×10^{-6}	2.47×10^{-4}	9.41×10^{-4}	1.89×10^{-3}	1.33×10^{-2}
II	200	5.21×10^{-6}	1.46×10^{-5}	8.79×10^{-4}	3.92×10^{-3}	1.19×10^{-2}	8.46×10^{-2}
	500	9.95×10^{-7}	3.32×10^{-6}	2.60×10^{-4}	1.05×10^{-3}	1.81×10^{-3}	1.19×10^{-2}
III	200	4.98×10^{-6}	1.41×10^{-6}	6.10×10^{-4}	2.71×10^{-3}	0.01	0.06
	500	9.74×10^{-7}	3.23×10^{-6}	2.63×10^{-4}	1.05×10^{-3}	2.07×10^{-3}	1.38×10^{-2}
Case 2: t-distribution							
I	200	2.49×10^{-4}	5.20×10^{-4}	1.68×10^{-2}	8.07×10^{-2}	0.37	2.5
	500	1.07×10^{-4}	2.23×10^{-4}	5.34×10^{-3}	2.52×10^{-2}	0.15	1.04
II	200	3.02×10^{-4}	6.24×10^{-4}	1.92×10^{-2}	8.66×10^{-2}	0.35	2.44
	500	1.07×10^{-4}	2.25×10^{-4}	6.54×10^{-3}	3.10×10^{-2}	0.15	0.99
III	200	2.43×10^{-4}	5.09×10^{-4}	1.60×10^{-2}	7.05×10^{-2}	0.38	2.72
	500	9.73×10^{-5}	2.02×10^{-4}	6.17×10^{-3}	2.73×10^{-2}	0.16	1.02

Table S.6: Simulation results for $n = 200, 500$ with 100 replicates of the dyad independence model, stochastic block model and power-law distribution model respectively. The results are displayed when the error follows normal distribution and t-distribution in each scenario. The Bias and SD are both reported. The true number of principal components (# PC) ranges from 2 to 50, with $\rho = 0.8$.

		# PC: 2		# PC: 10		# PC: 50	
Scenario	n	Bias	SD	Bias	SD	Bias	SD
Case 1: Normal distribution							
I	200	-0.0217	0.0310	0.0278	0.0301	0.0279	0.0292
	500	-0.0131	0.0176	0.0129	0.0156	0.0131	0.0177
II	200	-0.0037	0.0084	-0.0039	0.0071	-0.0045	0.0081
	500	-0.0013	0.0029	-0.0020	0.0036	-0.0017	0.0028
III	200	-0.0071	0.0204	-0.0103	0.0199	-0.0090	0.0170
	500	-0.0036	0.0111	-0.0041	0.0093	-0.0039	0.0101
Case 2: t-distribution							
I	200	-0.0372	0.1306	-0.0264	0.1354	-0.0279	0.1196
	500	-0.0112	0.0917	-0.0197	0.0958	-0.0140	0.0749
II	200	-0.0041	0.0338	-0.0020	0.0288	-0.0029	0.0277
	500	-0.0034	0.0197	-0.0016	0.0187	-0.0019	0.0187
III	200	-0.0053	0.0403	-0.0032	0.0471	-0.0041	0.0414
	500	-0.0039	0.0304	-0.0012	0.0276	-0.0027	0.0297

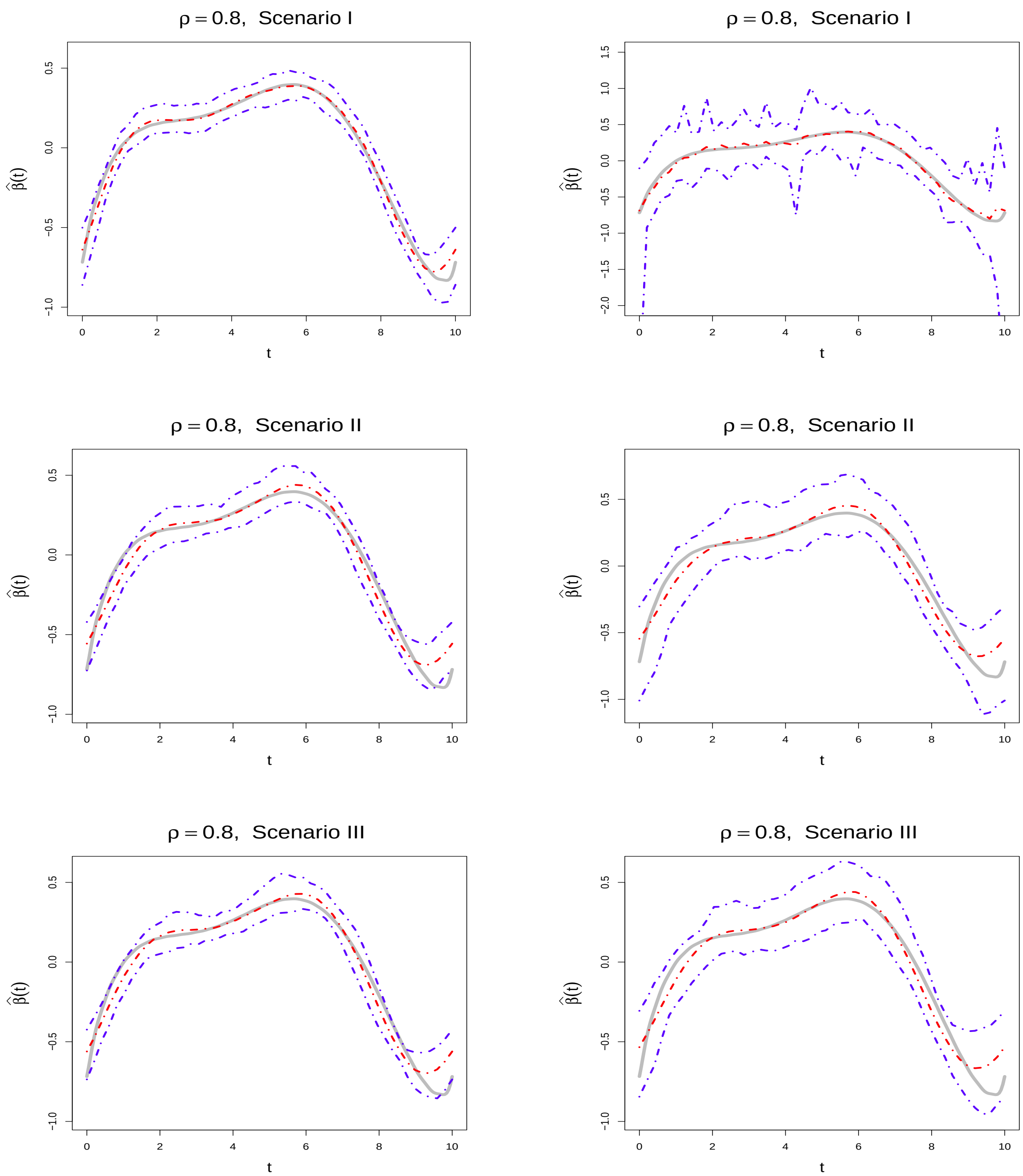

Figure S.2: The functional coefficients estimation $\widehat{\beta}(t)$ under normal distribution and t distribution, respectively, with sample size $n = 200$ and $\rho = 0.8$. The left panels contain the results under normal distribution, while the right panels under t distribution. In each panel, the solid grey line is the true value, the dashed red line is the average estimated value, and the dashed blue lines are the pointwise 2.5% and 97.5% percentiles of the estimators based on 100 replications.

Table S.7: The ratio of $\widehat{\mathrm{SE}}_{\mathrm{emp}}$ to $\widehat{\mathrm{SE}}_{\mathrm{asy}}$ for $\rho = 0.5$ and 0.8 across different sample sizes.

	ρ	$n = 100$	$n = 200$	$n = 300$	$n = 400$	$n = 500$
ratio	0.5	1.2063	1.8947	0.9197	1.4219	1.0810
	0.8	1.1092	0.9133	1.0261	0.8674	1.0069
	ρ	$n = 600$	$n = 700$	$n = 800$	$n = 900$	$n = 1000$
ratio	0.5	0.7842	1.2423	1.1216	0.8462	0.9221
	0.8	0.9234	1.2329	0.8289	1.0455	1.3318

Bibliography

Abadir, K. M. and J. R. Magnus (2005). *Matrix Algebra.* Cambridge University Press.

Bhatia, R., C. Davis, and A. McIntosh (1983). Perturbation of spectral subspaces and solution of linear operator equations. *Linear Algebra and its Applications 52*, 45–67.

Fan, J. Q. and R. Z. Li (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association 96*(456), 1348–1360.

Hall, P. and J. L. Horowitz (2007). Methodology and convergence rates for functional linear regression. *The Annals of Statistics 35*(1), 70–91.

Horn, R. A. and C. R. Johnson (2012). *Matrix Analysis.* Cambridge University Press.

Huang, D., W. Lan, H. H. Zhang, and H. Wang (2019). Least squares estimation of spatial autoregressive models for large-scale social networks. *Electronic Journal of Statistics 13*(1), 1935–7524.

Kong, D. H., K. J. Xue, F. Yao, and H. H. Zhang (2016). Partially functional linear regression in high dimensions. *Biometrika 103*(1), 147–159.

Rencher, A. C. and G. B. Schaalje (2008). *Linear Models in Statistics.* John Wiley & Sons.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* New York: Wiley.

Wang, H., B. Li, and C. Leng (2009). Shrinkage tuning parameter selection with a diverging number of parameters. *Journal of the Royal Statistical Society Series B: Statistical Methodology 71*(3), 671–683.

Wong, R. K., Y. Li, and Z. Zhu (2019). Partially linear functional additive models for multivariate functional data. *Journal of the American Statistical Association 114*(525), 406–418.