Joint Mean and Correlation Regression Models for Multivariate Data

Zhi Yang Tho, Francis K. C. Hui, and Tao Zou

The Australian National University

Supplementary Material

This supplementary material consists of eight parts.

1. Section S1 provides details of the consistent estimator for $\boldsymbol{G}^{(\mathcal{S})}$ in Section 2.2 to conduct inference on the mean regression coefficients and correlation regression parameters.

2. Section S2 presents an algorithm summarizing the full estimation procedure in Section 3 along with the alternating direction method of multipliers (ADMM) algorithm.

3. Section S3 discusses the regularity conditions in the Appendix.

4. Section S4 introduces technical lemmas to facilitate the proofs of the theorems.

5. Section S5 demonstrates the proofs of the lemmas.

6. Section S6 presents the proofs of Proposition 1, Theorems 1 – 2 and Corollary 1.

7. Section S7 presents additional simulation results for negative binomial and Gaussian responses, the results for the empirical coverage probability of the 95% confidence intervals for the mean regression coefficients and correlation regression parameters individually, along with comparisons to existing methods and a numerical investigation of the impact of misspecified similarity measures.

8. Section S8 provides supplementary details of the real data application in Section 5.

Throughout this supplementary material, let $\mathcal{S} = \{s_1, \cdots, s_q\}$ with $q < \infty$ be an arbitrary finite subset of $\{1, \cdots, p\}$ and $\|\cdot\|_t$ be the vector t-norm or matrix t-norm for $1 \leq t \leq \infty$. In other words, for a generic vector $\boldsymbol{a} = (a_1, \cdots, a_q)^\top$, $\|\boldsymbol{a}\|_t = (\sum_{i=1}^q |a_i|^t)^{1/t}$, and, for a generic $m \times q$ matrix $\boldsymbol{H}$,

$$\|\boldsymbol{H}\|_t = \sup\left\{\frac{\|\boldsymbol{H}\boldsymbol{a}\|_t}{\|\boldsymbol{a}\|_t} : \boldsymbol{a} \in \mathbb{R}^q \text{ and } \boldsymbol{a} \neq \boldsymbol{0}_q\right\}.$$

Furthermore, define the element-wise ℓ_∞ norm for a generic matrix $\boldsymbol{H}$ as $|\boldsymbol{H}|_\infty = \|\text{vec}(\boldsymbol{H})\|_\infty$, where $\text{vec}(\boldsymbol{H})$ denotes the vectorization of matrix $\boldsymbol{H}$. Also, let $\lambda_{\min}(\boldsymbol{H})$ and $\lambda_{\max}(\boldsymbol{H})$ denote the smallest and largest eigenvalues, respectively, of a generic square matrix $\boldsymbol{H}$. For the sake of simplicity, we denote $\tilde{\boldsymbol{\Sigma}}(\boldsymbol{\alpha}^{(0)})$ and $\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}^{(0)})$ as $\tilde{\boldsymbol{\Sigma}}_0$ and $\boldsymbol{\varepsilon}$, respectively, in this supplementary material. The source code of this article is provided at `https://github.com/Zy1225/Joint_Mean_Correlation_Regression`.

S1 Consistent Estimator for Asymptotic Covariance Matrix and Inference on Parameters

Recall from Section 2.2 that the expression of $\boldsymbol{G}^{(\mathcal{S})} = \bar{\boldsymbol{Z}}^{-1}\bar{\boldsymbol{\Omega}}(\bar{\boldsymbol{\Xi}}^{(\mathcal{S})})\bar{\boldsymbol{Z}}^{-1}$ involves $\boldsymbol{\vartheta}^{(0)}$, the third-order moment $\mu^{(3)}$ and the fourth-order moment $\mu^{(4)}$ defined in Condition 1. In this section, we demonstrate how to obtain a consistent estimator for $\boldsymbol{G}^{(\mathcal{S})}$ and conduct inference on the mean regression coefficients and correlation regression parameters such as constructing confidence intervals and testing hypotheses. Let $\boldsymbol{\varepsilon}(\hat{\boldsymbol{\vartheta}}) = (\boldsymbol{I}_n \otimes \hat{\boldsymbol{L}})^{-1}\boldsymbol{A}^{-1/2}(\hat{\boldsymbol{\beta}})\{\boldsymbol{Y} - \boldsymbol{\mu}(\hat{\boldsymbol{\beta}})\}$, where $\hat{\boldsymbol{L}}$ is obtained through the Cholesky decomposition of $\boldsymbol{\Sigma}(\hat{\boldsymbol{\alpha}}) = \hat{\boldsymbol{L}}\hat{\boldsymbol{L}}^\top$. We can consistently estimate $\boldsymbol{G}^{(\mathcal{S})}$ with $\hat{\boldsymbol{G}}^{(\mathcal{S})}$ by substituting $\boldsymbol{\vartheta}^{(0)}$, $\mu^{(3)}$ and $\mu^{(4)}$ in $\boldsymbol{G}^{(\mathcal{S})}$ by $\hat{\boldsymbol{\vartheta}}$, the third and fourth order empirical moments of the elements in $\boldsymbol{\varepsilon}(\hat{\boldsymbol{\vartheta}})$, respectively. For the cases where the dispersion parameters are unknown (e.g., the simulation study and real data application in Sections 4 and 5, respectively), a consistent estimator can be obtained by replacing the unknown dispersion parameters in $\boldsymbol{G}^{(\mathcal{S})}$ with $\hat{\boldsymbol{\phi}}$ from Algorithm 1 in Section S2 in addition to the substitutions above.

Let $\hat{\boldsymbol{G}}$ denote the aforementioned $\hat{\boldsymbol{G}}^{(\mathcal{S})}$ estimator with $S = \{1, \cdots, p\}$ and $\Phi^{-1}(\cdot)$ denote the standard normal quantile function. The asymptotic $(1-\alpha)$ confidence interval for the true mean regression coefficient $\beta_{jl}^{(0)}$ and correlation regression parameter $\rho_k^{(0)}$ are given as $\hat{\beta}_{jl} \pm \Phi^{-1}(1-\alpha/2)\hat{g}^{1/2}_{(j-1)d+l,(j-1)d+l}$ and $\hat{\rho}_k \pm \Phi^{-1}(1-\alpha/2)\hat{g}^{1/2}_{pd+k,pd+k}$, respectively, where $j = 1, \cdots, p$, $l = 1, \cdots, d$, $k = 1, \cdots, K$ and $\hat{g}_{l_1,l_2}$ is the (l_1, l_2)-th element of $\hat{\boldsymbol{G}}$ for $l_1, l_2 = 1, \cdots, pd+K$. We can also test the following hypotheses on $\beta_{jl}^{(0)}$ and $\rho_k^{(0)}$ for $j = 1, \cdots, p$, $l = 1, \cdots, d$ and $k = 1, \cdots, K$:

$$H_0 : \beta_{jl}^{(0)} = c_\beta \text{ vs. } H_1 : \beta_{jl}^{(0)} \neq c_\beta,$$

and

$$H_0 : \rho_k^{(0)} = c_\rho \text{ vs. } H_1 : \rho_k^{(0)} \neq c_\rho,$$

with some constants c_β and c_ρ, using the test statistics $T_\beta = (\hat{\beta}_{jl} - c_\beta)/\hat{g}^{1/2}_{(j-1)d+l,(j-1)d+l}$ and $T_\rho = (\hat{\rho}_k - c_\rho)/\hat{g}^{1/2}_{pd+k,pd+k}$, respectively. The test statistics T_β and T_ρ are both asymptotically standard normal under their corresponding null hypothesis. These tests allow researchers to assess the significance of various covariates on the means of different responses and determine important drivers of the correlations between responses, while accounting for the uncertainty associated with the joint estimation of the mean regression coefficients and correlation regression parameters.

S2 Algorithms

Algorithm 1 summarizes the full estimation procedure for the proposed model, where we slightly abuse notation by denoting the resulting joint estimates of the mean regression coefficients and correlation regression parameters as $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}^\top, \hat{\boldsymbol{\rho}}^\top)^\top$ and dispersion parameter estimates as $\hat{\boldsymbol{\phi}} = (\hat{\phi}_1, \cdots, \hat{\phi}_p)^\top$. For the initial values, we set $\hat{\boldsymbol{\beta}}^{[0]}$ as the estimated mean regression coefficients based on fitting a model with $\boldsymbol{\rho} = \mathbf{0}$ i.e., assuming the responses are independent of each other, and set $\hat{\boldsymbol{\alpha}}^{[0]} = (1, 0, \cdots, 0)^\top$. If required, we set $\hat{\boldsymbol{\phi}}^{[0]}$ to be a vector of ones. Additionally, we set $\xi = 10^{-8}$ and $\gamma = 1$ for the stepsize in the simulations in Section 4, while based on preliminary testing we reduced the step size to $\gamma = 0.5$ in the real data application in Section 5 to better improve

Algorithm 1: Estimation procedure for the parameters of the joint mean and correlation regression model along with the dispersion parameters as appropriate.

Input: Initial values $\hat{\boldsymbol{\beta}}^{[0]}$, $\hat{\boldsymbol{\phi}}^{[0]}$, $\hat{\boldsymbol{\alpha}}^{[0]}$ along with a tuning parameter for the stepsize $\gamma > 0$ and a sufficiently small convergence tolerance parameter $\xi > 0$.

Set $t = 0$;

while $\|(\hat{\boldsymbol{\beta}}^{[t+1]\top}, \hat{\boldsymbol{\rho}}^{[t+1]\top})^\top - (\hat{\boldsymbol{\beta}}^{[t]\top}, \hat{\boldsymbol{\rho}}^{[t]\top})^\top\| > \xi$ *or* $t = 0$ **do**

(i) Update all mean regression coefficients as

$$\hat{\boldsymbol{\beta}}^{[t+1]} = \hat{\boldsymbol{\beta}}^{[t]} - \gamma \boldsymbol{J}^{-1}(\hat{\boldsymbol{\beta}}^{[t]}, \hat{\boldsymbol{\alpha}}^{[t]}) \boldsymbol{\psi}_{\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}}^{[t]}, \hat{\boldsymbol{\alpha}}^{[t]}).$$

(ii) If required, update the heterogeneous dispersion parameters based on solving the equation

$$\frac{1}{n-d} \sum_{i=1}^{n} \frac{\{Y_{ij} - \mu_{ij}(\hat{\boldsymbol{\beta}}_j^{[t+1]})\}^2}{h\{\mu_{ij}(\hat{\boldsymbol{\beta}}_j^{[t+1]}); \phi_j\}} = 1$$

for $j = 1, \cdots, p$.

(iii) Calculate the unconstrained estimator

$$\hat{\boldsymbol{\alpha}}^{[t+1]} = \big(\mathrm{tr}(\tilde{\boldsymbol{W}}_{k_1} \tilde{\boldsymbol{W}}_{k_2})\big)^{-1}_{(K+1)\times(K+1)} \big(\boldsymbol{\epsilon}^\top(\hat{\boldsymbol{\beta}}^{[t+1]}) \tilde{\boldsymbol{W}}_k \boldsymbol{\epsilon}(\hat{\boldsymbol{\beta}}^{[t+1]})\big)_{(K+1)\times 1}.$$

If $\boldsymbol{\Sigma}(\hat{\boldsymbol{\alpha}}^{[t+1]})$ is not positive definite, then solve $\hat{\boldsymbol{\alpha}}^{[t+1]} = \arg\min_{\boldsymbol{\alpha} \in \mathscr{A}^+} Q(\hat{\boldsymbol{\beta}}^{[t+1]}, \boldsymbol{\alpha})$ via the ADMM algorithm as detailed in Algorithm 2.

(iv) Set $\hat{\boldsymbol{\rho}}^{[t+1]} = (\hat{\alpha}_1^{[t+1]}/\hat{\alpha}_0^{[t+1]}, \cdots, \hat{\alpha}_K^{[t+1]}/\hat{\alpha}_0^{[t+1]})^\top$.

Output: Joint estimator $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}^\top, \hat{\boldsymbol{\rho}}^\top)^\top$ of the mean regression coefficients and correlation regression parameters based on their converged values, along with $\hat{\boldsymbol{\phi}}$ as appropriate.

convergence of the algorithm.

Algorithm 2 provides a modification of the ADMM algorithm from Zou et al. (2017) for obtaining the constrained estimator $\hat{\boldsymbol{\alpha}}^{[t+1]} = \arg\min_{\boldsymbol{\alpha} \in \mathscr{A}^+} Q(\hat{\boldsymbol{\beta}}^{[t+1]}, \boldsymbol{\alpha})$ in step (iii) of Algorithm 1. To simplify the notation used in Algorithm 2, we denote the constrained estimator as $\hat{\boldsymbol{\alpha}}_+ = \arg\min_{\boldsymbol{\alpha} \in \mathscr{A}^+} Q(\boldsymbol{\beta}, \boldsymbol{\alpha})$, and the unconstrained estimator as $\hat{\boldsymbol{\alpha}} = (\mathrm{tr}(\tilde{\boldsymbol{W}}_{k_1} \tilde{\boldsymbol{W}}_{k_2}))^{-1}_{(K+1)\times(K+1)} (\boldsymbol{\epsilon}^\top(\boldsymbol{\beta}) \tilde{\boldsymbol{W}}_k \boldsymbol{\epsilon}(\boldsymbol{\beta}))_{(K+1)\times 1}$,

Algorithm 2: ADMM algorithm for obtaining the constrained estimator $\hat{\boldsymbol{\alpha}}_+ = \arg\min_{\boldsymbol{\alpha}\in\mathscr{A}^+} Q(\boldsymbol{\beta}, \boldsymbol{\alpha})$.

Input: Initial values $\boldsymbol{\Lambda}_{\{0\}}$ and $\hat{\boldsymbol{\alpha}}_{\{0\}}$, along with $\kappa > 0$, $\nu > 0$, a sufficiently small convergence tolerance parameter $\xi_{\boldsymbol{\alpha}} > 0$, and the unconstrained estimator $\hat{\boldsymbol{\alpha}}$.

Set $s = 0$;

while $\|\hat{\boldsymbol{\alpha}}_{\{s+1\}} - \hat{\boldsymbol{\alpha}}_{\{s\}}\| > \xi_{\boldsymbol{\alpha}}$ *or* $s = 0$ **do**

(i) Update the augmented parameter matrix as

$$\boldsymbol{\Delta}_{\{s+1\}} = \sum_{j=1}^{p} \max\{\lambda_j\{\boldsymbol{\Sigma}(\hat{\boldsymbol{\alpha}}_{\{s\}}) + \kappa\boldsymbol{\Lambda}_{\{s\}}\}, \nu\}\boldsymbol{v}_j\boldsymbol{v}_j^\top,$$

where $\boldsymbol{v}_j$ is the eigenvector corresponding to the j-th largest eigenvalue $\lambda_j\{\boldsymbol{\Sigma}(\hat{\boldsymbol{\alpha}}_{\{s\}}) + \kappa\boldsymbol{\Lambda}_{\{s\}}\}$.

(ii) Compute

$$\hat{\boldsymbol{\alpha}}_{\{s+1\}} = \frac{2n\kappa}{2n\kappa+1}\hat{\boldsymbol{\alpha}} + \frac{1}{2n\kappa+1}\big(\mathrm{tr}(\boldsymbol{W}_{k_1}\boldsymbol{W}_{k_2})\big)^{-1}_{(K+1)\times(K+1)}\big(\mathrm{tr}\{\boldsymbol{W}_k(\boldsymbol{\Delta}_{\{s+1\}} - \kappa\boldsymbol{\Lambda}_{\{s\}})\}\big)_{(K+1)\times 1}.$$

(iii) Update the Lagrange multiplier as

$$\boldsymbol{\Lambda}_{\{s+1\}} = \boldsymbol{\Lambda}_{\{s\}} - \frac{1}{\kappa}\{\boldsymbol{\Delta}_{\{s+1\}} - \boldsymbol{\Sigma}(\hat{\boldsymbol{\alpha}}_{\{s+1\}})\}.$$

Output: Constrained estimator $\hat{\boldsymbol{\alpha}}_+ = \arg\min_{\boldsymbol{\alpha}\in\mathscr{A}^+} Q(\boldsymbol{\beta}, \boldsymbol{\alpha})$ of the reparameterized correlation regression parameters based on the converged values.

for any given $\boldsymbol{\beta}$. Let $\boldsymbol{\Delta}$ be a $p \times p$ augmented parameter matrix, $\boldsymbol{\Lambda}$ be the $p \times p$ Lagrange multiplier, $\kappa > 0$ be a penalty tuning parameter, ν be an arbitrarily small positive constant, and $\xi_{\boldsymbol{\alpha}} > 0$ be a convergence tolerance parameter. Then the initial values are set as $\boldsymbol{\Lambda}_{\{0\}} = \mathbf{0}_{p\times p}$ and $\hat{\boldsymbol{\alpha}}_{\{0\}} = \hat{\boldsymbol{\alpha}}$. Following Xue et al. (2012) and Zou et al. (2017), we set the positive constant $\nu = 10^{-5}$. Finally, we used $\kappa = 0.05$ and $\xi_{\boldsymbol{\alpha}} = 10^{-8}$ for the penalty and convergence tuning parameters, respectively.

S3 Discussions of Conditions in Appendix

It is worth noting that all of the conditions in the Appendix are mild and sensible. Condition 1 is a moment condition, which is weaker than commonly used distribution assumptions (e.g., the normal distribution assumption in Zou et al., 2017, 2022). This condition implies the $(4+\eta)$-th moments of the responses Y_{ij} are finite. Condition 2 for $\boldsymbol{L}_0$, $\boldsymbol{L}_0^{-1}$ and $\boldsymbol{W}_k$ matrices have been considered in Zou et al. (2022)'s Conditions (C8) and (C9), while the condition imposed on the element-wise ℓ_∞ norm of $\boldsymbol{X}$ matrix in Condition 2 is also commonly used (e.g., McGuire et al., 1968). Conditions 3 and 4 are critical for showing the asymptotic normality of the proposed joint estimator, where all the link and variance functions illustrated below equation (2.1) satisfy Condition 4. Condition 5 implies $\sup_{n\geq 1, p\geq 1} |\boldsymbol{B}^{-1}|_\infty \leq \sup_{n\geq 1, p\geq 1} \|\boldsymbol{B}^{-1}\|_2 \leq C_B < \infty$. Moreover, it also implies $\inf_{n\geq 1, p\geq 1} \lambda_{\min}\big(\boldsymbol{\Omega}(\boldsymbol{\Xi}^{(\mathcal{S})})\big) \geq \inf_{n\geq 1, p\geq 1} \lambda_{\min}(\boldsymbol{\Omega}) \geq C_\Omega > 0$ by inequality 6.65a(i) in Seber (2008), which further leads to $\sup_{n\geq 1, p\geq 1} |\boldsymbol{\Omega}^{-1}(\boldsymbol{\Xi}^{(\mathcal{S})})|_\infty \leq \sup_{n\geq 1, p\geq 1} \|\boldsymbol{\Omega}^{-1}(\boldsymbol{\Xi}^{(\mathcal{S})})\|_2 < \infty$. These results will be used as part of the proofs in Section S5.

S4 Technical Lemmas

To facilitate the proofs of Theorems 1 – 2, we present the following lemmas. The proofs of the lemmas are provided in Section S5. Let

$$\mathcal{B}(\boldsymbol{\alpha}^{(0)}) = \begin{pmatrix} \left(\tilde{\boldsymbol{L}}_0^\top \tilde{\boldsymbol{W}}_0 \tilde{\boldsymbol{L}}_0\right)_{11} & \cdots & \left(\tilde{\boldsymbol{L}}_0^\top \tilde{\boldsymbol{W}}_K \tilde{\boldsymbol{L}}_0\right)_{11} \\ \vdots & \ddots & \vdots \\ \left(\tilde{\boldsymbol{L}}_0^\top \tilde{\boldsymbol{W}}_0 \tilde{\boldsymbol{L}}_0\right)_{np,np} & \cdots & \left(\tilde{\boldsymbol{L}}_0^\top \tilde{\boldsymbol{W}}_K \tilde{\boldsymbol{L}}_0\right)_{np,np} \end{pmatrix},$$

where $\tilde{\boldsymbol{L}}_0$ is defined in Condition 1 and $(\tilde{\boldsymbol{L}}_0^\top \tilde{\boldsymbol{W}}_k \tilde{\boldsymbol{L}}_0)_{l_1 l_2}$ is the (l_1, l_2)-th element of $\tilde{\boldsymbol{L}}_0^\top \tilde{\boldsymbol{W}}_k \tilde{\boldsymbol{L}}_0$ for $k = 0, \cdots, K$ and $l_1, l_2 = 1, \cdots, np$. We further denote $\boldsymbol{M}_{jl} = \{\partial \boldsymbol{A}^{-1/2}(\boldsymbol{\beta}^{(0)}) / \partial \beta_{jl}\} \boldsymbol{A}^{1/2}(\boldsymbol{\beta}^{(0)}) \in \mathbb{R}^{np \times np}$ for $j = 1, \cdots, p$ and $l = 1, \cdots, d$. Specifically, $\boldsymbol{M}_{jl}$ is a diagonal matrix with the $((i-1)p+j, (i-1)p+j)$-th element being

$$-\left[h'\{g^{-1}(\boldsymbol{x}_i^\top \boldsymbol{\beta}_j^{(0)}); \phi_j\} g^{-1\prime}(\boldsymbol{x}_i^\top \boldsymbol{\beta}_j^{(0)}) x_{il}\right] / \left[2h\{g^{-1}(\boldsymbol{x}_i^\top \boldsymbol{\beta}_j^{(0)}); \phi_j\}\right], \text{ for } i = 1, \cdots, n,$$

and other elements being zeros.

Lemma S1. *Denote*

$$E\left\{\frac{\partial \boldsymbol{\psi}(\boldsymbol{\vartheta}^{(0)})}{\partial \boldsymbol{\vartheta}^\top}\right\} = \begin{pmatrix} E\left\{\frac{\partial \boldsymbol{\psi}_{\boldsymbol{\beta}}(\boldsymbol{\vartheta}^{(0)})}{\partial \boldsymbol{\beta}^\top}\right\} & E\left\{\frac{\partial \boldsymbol{\psi}_{\boldsymbol{\beta}}(\boldsymbol{\vartheta}^{(0)})}{\partial \boldsymbol{\alpha}^\top}\right\} \\ E\left\{\frac{\partial \boldsymbol{\psi}_{\boldsymbol{\alpha}}(\boldsymbol{\vartheta}^{(0)})}{\partial \boldsymbol{\beta}^\top}\right\} & E\left\{\frac{\partial \boldsymbol{\psi}_{\boldsymbol{\alpha}}(\boldsymbol{\vartheta}^{(0)})}{\partial \boldsymbol{\alpha}^\top}\right\} \end{pmatrix}, \textit{ and}$$

$$\mathrm{Cov}\left\{\boldsymbol{\psi}(\boldsymbol{\vartheta}^{(0)})\right\} = \begin{pmatrix} \mathrm{Cov}\left\{\boldsymbol{\psi}_{\boldsymbol{\beta}}(\boldsymbol{\vartheta}^{(0)})\right\} & \mathrm{Cov}\left\{\boldsymbol{\psi}_{\boldsymbol{\beta}}(\boldsymbol{\vartheta}^{(0)}), \boldsymbol{\psi}_{\boldsymbol{\alpha}}(\boldsymbol{\vartheta}^{(0)})\right\} \\ \mathrm{Cov}\left\{\boldsymbol{\psi}_{\boldsymbol{\alpha}}(\boldsymbol{\vartheta}^{(0)}), \boldsymbol{\psi}_{\boldsymbol{\beta}}(\boldsymbol{\vartheta}^{(0)})\right\} & \mathrm{Cov}\left\{\boldsymbol{\psi}_{\boldsymbol{\alpha}}(\boldsymbol{\vartheta}^{(0)})\right\} \end{pmatrix}.$$

Under Condition 1 in the Appendix, we obtain

$$E\left\{\frac{\partial \boldsymbol{\psi}_{\boldsymbol{\beta}}(\boldsymbol{\vartheta}^{(0)})}{\partial \boldsymbol{\beta}^\top}\right\} = -\boldsymbol{D}^\top(\boldsymbol{\beta}^{(0)})\boldsymbol{A}^{-1/2}(\boldsymbol{\beta}^{(0)})\tilde{\boldsymbol{\Sigma}}_0^{-1}\boldsymbol{A}^{-1/2}(\boldsymbol{\beta}^{(0)})\boldsymbol{D}(\boldsymbol{\beta}^{(0)}),$$

$$E\left\{\frac{\partial \boldsymbol{\psi}_{\boldsymbol{\beta}}(\boldsymbol{\vartheta}^{(0)})}{\partial \boldsymbol{\alpha}^\top}\right\} = \mathbf{0}_{pd\times(K+1)},$$

$$E\left\{\frac{\partial \boldsymbol{\psi}_{\boldsymbol{\alpha}}(\boldsymbol{\vartheta}^{(0)})}{\partial \boldsymbol{\beta}^\top}\right\} = 2\begin{pmatrix} \mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_0\tilde{\boldsymbol{W}}_0\boldsymbol{M}_{11}) & \cdots & \mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_0\tilde{\boldsymbol{W}}_0\boldsymbol{M}_{1d}) & \cdots & \mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_0\tilde{\boldsymbol{W}}_0\boldsymbol{M}_{p1}) & \cdots & \mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_0\tilde{\boldsymbol{W}}_0\boldsymbol{M}_{pd}) \\ \mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_0\tilde{\boldsymbol{W}}_1\boldsymbol{M}_{11}) & \cdots & \mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_0\tilde{\boldsymbol{W}}_1\boldsymbol{M}_{1d}) & \cdots & \mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_0\tilde{\boldsymbol{W}}_1\boldsymbol{M}_{p1}) & \cdots & \mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_0\tilde{\boldsymbol{W}}_1\boldsymbol{M}_{pd}) \\ \vdots & \cdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ \mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_0\tilde{\boldsymbol{W}}_K\boldsymbol{M}_{11}) & \cdots & \mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_0\tilde{\boldsymbol{W}}_K\boldsymbol{M}_{1d}) & \cdots & \mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_0\tilde{\boldsymbol{W}}_K\boldsymbol{M}_{p1}) & \cdots & \mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_0\tilde{\boldsymbol{W}}_K\boldsymbol{M}_{pd}) \end{pmatrix},$$

$$E\left\{\frac{\partial \boldsymbol{\psi}_{\boldsymbol{\alpha}}(\boldsymbol{\vartheta}^{(0)})}{\partial \boldsymbol{\alpha}^\top}\right\} = -\left(\mathrm{tr}(\tilde{\boldsymbol{W}}_{k_1}\tilde{\boldsymbol{W}}_{k_2})\right)_{(K+1)\times(K+1)},$$

$$\mathrm{Cov}\left\{\boldsymbol{\psi}_{\boldsymbol{\beta}}(\boldsymbol{\vartheta}^{(0)})\right\} = \boldsymbol{D}^\top(\boldsymbol{\beta}^{(0)})\boldsymbol{A}^{-1/2}(\boldsymbol{\beta}^{(0)})\tilde{\boldsymbol{\Sigma}}_0^{-1}\boldsymbol{A}^{-1/2}(\boldsymbol{\beta}^{(0)})\boldsymbol{D}(\boldsymbol{\beta}^{(0)}),$$

$$\mathrm{Cov}\left\{\boldsymbol{\psi}_{\boldsymbol{\beta}}(\boldsymbol{\vartheta}^{(0)}), \boldsymbol{\psi}_{\boldsymbol{\alpha}}(\boldsymbol{\vartheta}^{(0)})\right\} = \mu^{(3)}\boldsymbol{D}^\top(\boldsymbol{\beta}^{(0)})\boldsymbol{A}^{-1/2}(\boldsymbol{\beta}^{(0)})\tilde{\boldsymbol{L}}_0^{-1\top}\boldsymbol{\mathcal{B}}(\boldsymbol{\alpha}^{(0)}),$$

$$\mathrm{Cov}\left\{\boldsymbol{\psi}_{\boldsymbol{\alpha}}(\boldsymbol{\vartheta}^{(0)}), \boldsymbol{\psi}_{\boldsymbol{\beta}}(\boldsymbol{\vartheta}^{(0)})\right\} = \left[\mathrm{Cov}\left\{\boldsymbol{\psi}_{\boldsymbol{\beta}}(\boldsymbol{\vartheta}^{(0)}), \boldsymbol{\psi}_{\boldsymbol{\alpha}}(\boldsymbol{\vartheta}^{(0)})\right\}\right]^\top, \text{ and}$$

$$\mathrm{Cov}\left\{\boldsymbol{\psi}_{\boldsymbol{\alpha}}(\boldsymbol{\vartheta}^{(0)})\right\} = 2\left(\mathrm{tr}(\tilde{\boldsymbol{L}}_0^\top\tilde{\boldsymbol{W}}_{k_1}\tilde{\boldsymbol{\Sigma}}_0\tilde{\boldsymbol{W}}_{k_2}\tilde{\boldsymbol{L}}_0)\right)_{(K+1)\times(K+1)} + \left(\mu^{(4)}-3\right)\boldsymbol{\mathcal{B}}^\top(\boldsymbol{\alpha}^{(0)})\boldsymbol{\mathcal{B}}(\boldsymbol{\alpha}^{(0)}),$$

where $\mu^{(3)}$ and $\mu^{(4)}$ are defined in Condition 1.

Lemma S2. *Let $\boldsymbol{u}_1 = -\boldsymbol{\Omega}^{-1/2}(\Xi^{(\mathcal{S})})\Xi^{(\mathcal{S})}\boldsymbol{B}^{-1}\tilde{\boldsymbol{Z}}^{-1}\psi(\boldsymbol{\vartheta}^{(0)})$. Under Conditions*

1 – 5 in the Appendix, we obtain $\boldsymbol{u}_1 \xrightarrow{d} N(\mathbf{0}_{qd+K+1}, \boldsymbol{I}_{qd+K+1})$ *as* $n \to \infty$ *and* $p = o(n^{1/2})$.

Lemma S3. *Let* $\boldsymbol{V} = \tilde{\boldsymbol{Z}}^{-1}[\partial\boldsymbol{\psi}(\boldsymbol{\vartheta}^{(0)})/\partial\boldsymbol{\vartheta}^\top - E\{\partial\boldsymbol{\psi}(\boldsymbol{\vartheta}^{(0)})/\partial\boldsymbol{\vartheta}^\top\}]\tilde{\boldsymbol{Z}}^{-1}$ *and* $\boldsymbol{u}_2 = \boldsymbol{\Omega}^{-1/2}(\Xi^{(\mathcal{S})})\,\Xi^{(\mathcal{S})}\boldsymbol{B}^{-1}\boldsymbol{V}\left(\boldsymbol{B}^{-1}\boldsymbol{V} + \boldsymbol{I}_{pd+K+1}\right)^{-1}\boldsymbol{B}^{-1}\tilde{\boldsymbol{Z}}^{-1}\boldsymbol{\psi}(\boldsymbol{\vartheta}^{(0)})$ *where the closed-form expression of* $\partial\boldsymbol{\psi}(\boldsymbol{\vartheta}^{(0)})/\partial\boldsymbol{\vartheta}^\top - E\{\partial\boldsymbol{\psi}(\boldsymbol{\vartheta}^{(0)})/\partial\boldsymbol{\vartheta}^\top\}$ *is provided in the proof of this lemma in Section S5. Under Conditions 1 – 5 in the Appendix, we obtain* $\boldsymbol{u}_2 = o_{\mathrm{P}}(1)$ *as* $n \to \infty$ *and* $p = o(n^{1/2})$.

Lemma S4. *Let* $\{(\boldsymbol{z}_{mj}, \mathcal{F}_{mj}) : 1 \le j \le k_m, m \ge 1\}$ *be a martingale difference array where* $\boldsymbol{z}_{mj}$ *are finite* s*-dimensional vectors. Here as* $m \to \infty$, $k_m \uparrow \infty$. *Suppose that for some* $\delta > 0$, $\sum_{j=1}^{k_m} E\|\boldsymbol{z}_{mj}\|_2^{2+\delta} \to 0$, *and* $\sum_{j=1}^{k_m} E(\boldsymbol{z}_{mj}\boldsymbol{z}_{mj}^\top|\mathcal{F}_{m,j-1}) \xrightarrow{\mathrm{P}} \boldsymbol{\mathcal{V}}$ *for some positive definite* $s \times s$ *matrix* $\boldsymbol{\mathcal{V}}$. *Then* $\sum_{j=1}^{k_m} \boldsymbol{z}_{mj} \xrightarrow{d} N(\mathbf{0}_s, \boldsymbol{\mathcal{V}})$.

Lemma S5. *For a generic vector* $\boldsymbol{a} = (a_1, \cdots, a_q)^\top \in \mathbb{R}^q$ *and* $s \ge 1$, *we have that*

$$\left|\sum_{j=1}^{q} a_j\right|^s \le q^{s-1}\sum_{j=1}^{q} |a_j|^s .$$

Lemma S6. *For a generic vector* $\boldsymbol{a} = (a_1, \cdots, a_q)^\top \in \mathbb{R}^q$, *generic matrices* $\boldsymbol{H} \in \mathbb{R}^{m\times q}$, $\boldsymbol{U} \in \mathbb{R}^{m\times m}$ *and* $s \ge 1$, *we have that* $\|\boldsymbol{U}\boldsymbol{H}\boldsymbol{a}\|_s \le m^{1/s}|\boldsymbol{H}|_\infty\|\boldsymbol{U}\|_\infty\|\boldsymbol{a}\|_1$.

Lemma S7. *Let* $\boldsymbol{\varepsilon} = (\varepsilon_1, \cdots, \varepsilon_{np})^\top$*, where* $\varepsilon_1, \cdots, \varepsilon_{np}$ *are independent and identically distributed with mean zero and finite variance* σ^2*. Define*

$$\boldsymbol{\mathcal{Q}}_{np} = \begin{pmatrix} \mathrm{vec}^\top(\boldsymbol{H}_1) \\ \vdots \\ \mathrm{vec}^\top(\boldsymbol{H}_L) \end{pmatrix} \mathrm{vec}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\top - \sigma^2 \boldsymbol{I}_{np}) + \boldsymbol{\Gamma}^\top \boldsymbol{\varepsilon},$$

where $\boldsymbol{H}_\ell = (h_{l_1 l_2}^{(\ell)})_{np\times np} \in \mathbb{R}^{np\times np}$ *for* $\ell = 1, \cdots, L$ *with* $L < \infty$ *and* $\boldsymbol{\Gamma} \in \mathbb{R}^{np\times L}$*. Suppose*

(1) $\boldsymbol{H}_\ell$ *matrices are symmetric for* $\ell = 1, \cdots, L$*; and*

(2) for some $\eta_1 > 0$*,* $E|\varepsilon_i|^{4+\eta_1} < \infty$*.*

Then, we have that $E(\boldsymbol{\mathcal{Q}}_{np}) = \mathbf{0}_{np}$ *and*

$$\mathrm{Cov}(\boldsymbol{\mathcal{Q}}_{np}) = 2\sigma^4 \big(\mathrm{tr}(\boldsymbol{H}_{\ell_1}\boldsymbol{H}_{\ell_2})\big)_{np\times np} + \sigma^2 \boldsymbol{\Gamma}^\top\boldsymbol{\Gamma} + \left(\mu^{(4)} - 3\sigma^4\right) \boldsymbol{\Psi}^\top\boldsymbol{\Psi} + \mu^{(3)}(\boldsymbol{\Psi}^\top\boldsymbol{\Gamma} + \boldsymbol{\Gamma}^\top\boldsymbol{\Psi}),$$

where $\boldsymbol{\Psi} = (\boldsymbol{\Psi}_1, \cdots, \boldsymbol{\Psi}_L)$ *is an* $np \times L$ *matrix with the* l*-th column given by* $\boldsymbol{\Psi}_l = (h_{11}^{(\ell)}, \cdots, h_{np,np}^{(\ell)})^\top$ *for* $\ell = 1, \cdots, L$*, and* $\mu^{(3)}$ *and* $\mu^{(4)}$ *are defined in Condition 1 in the Appendix.*

Lemma S8. *For a non-negative constant array* $\{c_j : j = 1, \cdots, q\}$ *and some* $\tau > 0$*, we have that*

$$\sum_{j=1}^{q} c_j^{1+\tau} \le \left(\sum_{j=1}^{q} c_j\right)^{1+\tau}.$$

Lemma S9. *Let* $\{(z_{mj}, \mathcal{F}_{mj}) : 1 \le j \le k_m, m \ge 1\}$ *be a martingale difference* *array and* $\{c_{mj} : 1 \le j \le k_m, m \ge 1\}$ *be a positive constant array. Here as* $m \to \infty$, $k_m \uparrow \infty$. *Suppose for some* $\zeta > 0$, *(1)* $E|z_{mj}/c_{mj}|^{1+\zeta} \le C$, *where* C *is not dependent on* m *and* j*; (2)* $\limsup_{m\to\infty} \sum_{j=1}^{k_m} c_{mj} < \infty$*; and (3)* $\lim_{m\to\infty} \sum_{j=1}^{k_m} c_{mj}^2 = 0$. *Then* $\sum_{j=1}^{k_m} z_{mj} \xrightarrow{\mathrm{P}} 0$.

Lemma S10. *For a generic matrix* $\boldsymbol{H} \in \mathbb{R}^{a\times b}$ *and its submatrix* $\boldsymbol{H}' \in \mathbb{R}^{a'\times b'}$ *that takes* $\mathcal{V} = \{v_1, \cdots, v_{a'}\}$*-th rows and* $\mathcal{T} = \{\tau_1, \cdots, \tau_{b'}\}$*-th columns of* $\boldsymbol{H}$*, where* $a' \le a$, $b' \le b$, $\mathcal{V} \subseteq \{1, \cdots, a\}$ *and* $\mathcal{T} \subseteq \{1, \cdots, b\}$*, we have that* $\|\boldsymbol{H}'\|_2 \le \|\boldsymbol{H}\|_2$.

S5 Proofs of Lemmas

In this section, we provide the proofs of the lemmas from Section S4. As Lemma S4 is directly generalized from the central limit theorem for martingale difference arrays of Hall and Heyde (2014, Chapter 3) in vector form, and Lemmas S5 – S7 are directly modified from Lemmas 1 – 2 and 4 of Zou et al. (2021), we only prove Lemmas S1 – S3 and Lemmas S8 – S10.

Proof of Lemma S1. From equations (2.4) – (2.5), we have

$$\boldsymbol{\psi}_{\boldsymbol{\beta}}(\boldsymbol{\vartheta}^{(0)}) = \boldsymbol{D}^\top(\boldsymbol{\beta}^{(0)})\boldsymbol{A}^{-1/2}(\boldsymbol{\beta}^{(0)})\tilde{\boldsymbol{L}}_0^{-1\top}\boldsymbol{\varepsilon}, \text{ and}$$

$$\boldsymbol{\psi}_{\boldsymbol{\alpha}}(\boldsymbol{\vartheta}^{(0)}) = \begin{pmatrix} \boldsymbol{\varepsilon}^\top \tilde{\boldsymbol{L}}_0^\top \tilde{\boldsymbol{W}}_0 \tilde{\boldsymbol{L}}_0 \boldsymbol{\varepsilon} - \mathrm{tr}(\tilde{\boldsymbol{L}}_0^\top \tilde{\boldsymbol{W}}_0 \tilde{\boldsymbol{L}}_0) \\ \vdots \\ \boldsymbol{\varepsilon}^\top \tilde{\boldsymbol{L}}_0^\top \tilde{\boldsymbol{W}}_K \tilde{\boldsymbol{L}}_0 \boldsymbol{\varepsilon} - \mathrm{tr}(\tilde{\boldsymbol{L}}_0^\top \tilde{\boldsymbol{W}}_K \tilde{\boldsymbol{L}}_0) \end{pmatrix}.$$

Then, by equation (2.3) and recalling the reparameterization of the correlation regression model in Section 2.1 along with some basic calculus, we can obtain

$$E\left\{\frac{\partial \boldsymbol{\psi}(\boldsymbol{\vartheta}^{(0)})}{\partial \boldsymbol{\vartheta}^\top}\right\} = \begin{pmatrix} E\left\{\frac{\partial \boldsymbol{\psi}_{\boldsymbol{\beta}}(\boldsymbol{\vartheta}^{(0)})}{\partial \boldsymbol{\beta}^\top}\right\} & E\left\{\frac{\partial \boldsymbol{\psi}_{\boldsymbol{\beta}}(\boldsymbol{\vartheta}^{(0)})}{\partial \boldsymbol{\alpha}^\top}\right\} \\ E\left\{\frac{\partial \boldsymbol{\psi}_{\boldsymbol{\alpha}}(\boldsymbol{\vartheta}^{(0)})}{\partial \boldsymbol{\beta}^\top}\right\} & E\left\{\frac{\partial \boldsymbol{\psi}_{\boldsymbol{\alpha}}(\boldsymbol{\vartheta}^{(0)})}{\partial \boldsymbol{\alpha}^\top}\right\} \end{pmatrix},$$

where

$$E\left\{\frac{\partial \boldsymbol{\psi}_{\boldsymbol{\beta}}(\boldsymbol{\vartheta}^{(0)})}{\partial \boldsymbol{\beta}^\top}\right\} = -\boldsymbol{D}^\top(\boldsymbol{\beta}^{(0)})\boldsymbol{A}^{-1/2}(\boldsymbol{\beta}^{(0)})\tilde{\boldsymbol{\Sigma}}_0^{-1}\boldsymbol{A}^{-1/2}(\boldsymbol{\beta}^{(0)})\boldsymbol{D}(\boldsymbol{\beta}^{(0)}),$$

$$E\left\{\frac{\partial \boldsymbol{\psi}_{\boldsymbol{\beta}}(\boldsymbol{\vartheta}^{(0)})}{\partial \boldsymbol{\alpha}^\top}\right\} = \mathbf{0}_{pd\times(K+1)},$$

$$E\left\{\frac{\partial \boldsymbol{\psi}_{\boldsymbol{\alpha}}(\boldsymbol{\vartheta}^{(0)})}{\partial \boldsymbol{\beta}^\top}\right\} = 2\begin{pmatrix} \mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_0\tilde{\boldsymbol{W}}_0\boldsymbol{M}_{11}) & \cdots & \mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_0\tilde{\boldsymbol{W}}_0\boldsymbol{M}_{1d}) & \cdots & \mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_0\tilde{\boldsymbol{W}}_0\boldsymbol{M}_{p1}) & \cdots & \mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_0\tilde{\boldsymbol{W}}_0\boldsymbol{M}_{pd}) \\ \mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_0\tilde{\boldsymbol{W}}_1\boldsymbol{M}_{11}) & \cdots & \mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_0\tilde{\boldsymbol{W}}_1\boldsymbol{M}_{1d}) & \cdots & \mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_0\tilde{\boldsymbol{W}}_1\boldsymbol{M}_{p1}) & \cdots & \mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_0\tilde{\boldsymbol{W}}_1\boldsymbol{M}_{pd}) \\ \vdots & \cdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ \mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_0\tilde{\boldsymbol{W}}_K\boldsymbol{M}_{11}) & \cdots & \mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_0\tilde{\boldsymbol{W}}_K\boldsymbol{M}_{1d}) & \cdots & \mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_0\tilde{\boldsymbol{W}}_K\boldsymbol{M}_{p1}) & \cdots & \mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_0\tilde{\boldsymbol{W}}_K\boldsymbol{M}_{pd}) \end{pmatrix},$$

$$E\left\{\frac{\partial \boldsymbol{\psi}_{\boldsymbol{\alpha}}(\boldsymbol{\vartheta}^{(0)})}{\partial \boldsymbol{\alpha}^\top}\right\} = -\left(\mathrm{tr}(\tilde{\boldsymbol{W}}_{k_1}\tilde{\boldsymbol{W}}_{k_2})\right)_{(K+1)\times(K+1)},$$

and $\boldsymbol{M}_{jl}$ is defined above Lemma S1.

By letting $\boldsymbol{H}_\ell = \mathbf{0}_{np\times np}$ for $\ell = 1,\cdots,pd$, $\boldsymbol{H}_{pd+k+1} = \tilde{\boldsymbol{L}}_0^\top \tilde{\boldsymbol{W}}_k \tilde{\boldsymbol{L}}_0$ for $k = 0,\cdots,K$, $\boldsymbol{\Gamma} = (\tilde{\boldsymbol{L}}_0^{-1}\boldsymbol{A}^{-1/2}(\boldsymbol{\beta}^{(0)})\boldsymbol{D}(\boldsymbol{\beta}^{(0)}), \mathbf{0}_{np\times(K+1)})$ and under Condition

1, we can apply Lemma S7 to obtain

$$\mathrm{Cov}\left\{\boldsymbol{\psi}(\boldsymbol{\vartheta}^{(0)})\right\} = \begin{pmatrix} \mathrm{Cov}\left\{\boldsymbol{\psi}_{\boldsymbol{\beta}}(\boldsymbol{\vartheta}^{(0)})\right\} & \mathrm{Cov}\left\{\boldsymbol{\psi}_{\boldsymbol{\beta}}(\boldsymbol{\vartheta}^{(0)}), \boldsymbol{\psi}_{\boldsymbol{\alpha}}(\boldsymbol{\vartheta}^{(0)})\right\} \\ \mathrm{Cov}\left\{\boldsymbol{\psi}_{\boldsymbol{\alpha}}(\boldsymbol{\vartheta}^{(0)}), \boldsymbol{\psi}_{\boldsymbol{\beta}}(\boldsymbol{\vartheta}^{(0)})\right\} & \mathrm{Cov}\left\{\boldsymbol{\psi}_{\boldsymbol{\alpha}}(\boldsymbol{\vartheta}^{(0)})\right\} \end{pmatrix},$$

where

$$\begin{aligned} \mathrm{Cov}\left\{\boldsymbol{\psi}_{\boldsymbol{\beta}}(\boldsymbol{\vartheta}^{(0)})\right\} &= \boldsymbol{D}^\top(\boldsymbol{\beta}^{(0)})\boldsymbol{A}^{-1/2}(\boldsymbol{\beta}^{(0)})\tilde{\boldsymbol{\Sigma}}_0^{-1}\boldsymbol{A}^{-1/2}(\boldsymbol{\beta}^{(0)})\boldsymbol{D}(\boldsymbol{\beta}^{(0)}), \\ \mathrm{Cov}\left\{\boldsymbol{\psi}_{\boldsymbol{\beta}}(\boldsymbol{\vartheta}^{(0)}), \boldsymbol{\psi}_{\boldsymbol{\alpha}}(\boldsymbol{\vartheta}^{(0)})\right\} &= \mu^{(3)}\boldsymbol{D}^\top(\boldsymbol{\beta}^{(0)})\boldsymbol{A}^{-1/2}(\boldsymbol{\beta}^{(0)})\tilde{\boldsymbol{L}}_0^{-1\top}\boldsymbol{\mathcal{B}}(\boldsymbol{\alpha}^{(0)}), \\ \mathrm{Cov}\left\{\boldsymbol{\psi}_{\boldsymbol{\alpha}}(\boldsymbol{\vartheta}^{(0)}), \boldsymbol{\psi}_{\boldsymbol{\beta}}(\boldsymbol{\vartheta}^{(0)})\right\} &= \left[\mathrm{Cov}\left\{\boldsymbol{\psi}_{\boldsymbol{\beta}}(\boldsymbol{\vartheta}^{(0)}), \boldsymbol{\psi}_{\boldsymbol{\alpha}}(\boldsymbol{\vartheta}^{(0)})\right\}\right]^\top, \\ \mathrm{Cov}\left\{\boldsymbol{\psi}_{\boldsymbol{\alpha}}(\boldsymbol{\vartheta}^{(0)})\right\} &= 2\left(\mathrm{tr}(\tilde{\boldsymbol{L}}_0^\top\tilde{\boldsymbol{W}}_{k_1}\tilde{\boldsymbol{\Sigma}}_0\tilde{\boldsymbol{W}}_{k_2}\tilde{\boldsymbol{L}}_0)\right)_{(K+1)\times(K+1)} \\ &\quad + \left(\mu^{(4)} - 3\right)\boldsymbol{\mathcal{B}}^\top(\boldsymbol{\alpha}^{(0)})\boldsymbol{\mathcal{B}}(\boldsymbol{\alpha}^{(0)}), \end{aligned}$$

and $\boldsymbol{\mathcal{B}}(\boldsymbol{\alpha}^{(0)})$ is defined above Lemma S1. □

Proof of Lemma S2. Before we present the main proof by employing Lemma S4, we first study the properties of $\boldsymbol{U}$ and $\boldsymbol{B}$, which are defined before Theorem 1. From the definition of $\boldsymbol{D}(\boldsymbol{\beta}^{(0)}) \in \mathbb{R}^{np\times pd}$, it can be shown that the $((i-1)p+j, (j-1)d+l)$-th element of $\boldsymbol{D}(\boldsymbol{\beta}^{(0)})$ is given as $g^{-1'}(\boldsymbol{x}_i^\top\boldsymbol{\beta}_j^{(0)})x_{il}$ for $i = 1, \cdots, n$, $j = 1, \cdots, p$, $l = 1, \cdots, d$ and other elements being zeros. By Conditions 2 – 4, we have that for any $i = 1, \cdots, n$ and $j = 1, \cdots, p$,

$$|\boldsymbol{x}_i^\top\boldsymbol{\beta}_j^{(0)}| = \left|\sum_{l=1}^d x_{il}\beta_{jl}^{(0)}\right| \le \sum_{l=1}^d |x_{il}|\,|\beta_{jl}^{(0)}| \le \sum_{l=1}^d \sup_{n\ge1}\{|\boldsymbol{X}|_\infty\}\sup_{p\ge1}\left\{\left\|\boldsymbol{\beta}^{(0)}\right\|_\infty\right\} \le dC_XC_\beta < \infty,$$

which implies $\boldsymbol{x}_i^\top \boldsymbol{\beta}_j^{(0)} \in \mathcal{D} = [-dC_X C_\beta, dC_X C_\beta]$. Since $g^{-1'}(\cdot)$ is continuous by Condition 4, we obtain that the range of $g^{-1'}(\cdot)$ acting on $\mathcal{D}$ is compact. In other words, there exists a finite positive constant C_g such that

$$|g^{-1'}(c)| \leq C_g, \ \forall c \in \mathcal{D},$$

and this leads to

$$|g^{-1'}(\boldsymbol{x}_i^\top \boldsymbol{\beta}_j^{(0)})| \leq C_g, \text{ for } i = 1, \cdots, n, j = 1, \cdots, p. \tag{S5.1}$$

Hence, we obtain

$$\begin{aligned} \left\|\boldsymbol{D}(\boldsymbol{\beta}^{(0)})\right\|_1 &= \max_{1\leq j\leq p;\ 1\leq l\leq d} \left\{\sum_{i=1}^n |g^{-1'}(\boldsymbol{x}_i^\top \boldsymbol{\beta}_j^{(0)}) x_{il}|\right\} \\ &\leq \max_{1\leq j\leq p;\ 1\leq l\leq d} \left\{\sum_{i=1}^n C_g C_X\right\} = nC_g C_X, \text{ and} \\ \left\|\boldsymbol{D}(\boldsymbol{\beta}^{(0)})\right\|_\infty &= \max_{1\leq i\leq n;\ 1\leq j\leq p} \left\{\sum_{l=1}^d |g^{-1'}(\boldsymbol{x}_i^\top \boldsymbol{\beta}_j^{(0)}) x_{il}|\right\} \\ &\leq \max_{1\leq i\leq n;\ 1\leq j\leq p} \left\{\sum_{l=1}^d C_g C_X\right\} = dC_g C_X. \end{aligned} \tag{S5.2}$$

By employing inequality 4.67(e) in Seber (2008), we have

$$\left\|\boldsymbol{D}(\boldsymbol{\beta}^{(0)})\right\|_2 \leq \sqrt{\|\boldsymbol{D}(\boldsymbol{\beta}^{(0)})\|_1 \|\boldsymbol{D}(\boldsymbol{\beta}^{(0)})\|_\infty} \leq \sqrt{nC_g C_X dC_g C_X}$$

$$= \sqrt{n}\left(\sqrt{d}C_gC_X\right) \triangleq \sqrt{n}C_1, \tag{S5.3}$$

for some finite positive constant C_1. Similarly, by Conditions 2 – 4 and inequality 4.67(e) in Seber (2008), we obtain

$$\begin{aligned}
\left\|\boldsymbol{A}^{-1/2}(\boldsymbol{\beta}^{(0)})\right\|_2 &\leq \sqrt{\|\boldsymbol{A}^{-1/2}(\boldsymbol{\beta}^{(0)})\|_1 \|\boldsymbol{A}^{-1/2}(\boldsymbol{\beta}^{(0)})\|_\infty} \\
&= \left\|\boldsymbol{A}^{-1/2}(\boldsymbol{\beta}^{(0)})\right\|_1 \\
&= \max_{1\leq i\leq n;\ 1\leq j\leq p} \left\{[h\{g^{-1}(\boldsymbol{x}_i^\top\boldsymbol{\beta}_j^{(0)});\phi_j\}]^{-1/2}\right\} \\
&\leq (C_h)^{-1/2} \triangleq C_2, \\
\left\|\tilde{\boldsymbol{\Sigma}}_0^{-1}\right\|_2 &= \left\|\tilde{\boldsymbol{L}}_0^{-1\top}\tilde{\boldsymbol{L}}_0^{-1}\right\|_2 \\
&\leq \left\|\tilde{\boldsymbol{L}}_0^{-1}\right\|_2^2 \\
&\leq \left\|\tilde{\boldsymbol{L}}_0^{-1}\right\|_1 \left\|\tilde{\boldsymbol{L}}_0^{-1}\right\|_\infty \\
&\leq \sup_{p\geq1}\left\{\left\|\boldsymbol{L}_0^{-1}\right\|_1\right\} \sup_{p\geq1}\left\{\left\|\boldsymbol{L}_0^{-1}\right\|_\infty\right\} \leq C_L^2 \triangleq C_3, \text{ and} \\
\left\|\tilde{\boldsymbol{L}}_0^{-1\top}\right\|_2 &\leq \sqrt{\left\|\tilde{\boldsymbol{L}}_0^{-1}\right\|_1 \left\|\tilde{\boldsymbol{L}}_0^{-1}\right\|_\infty} \\
&\leq \sqrt{\sup_{p\geq1}\left\{\left\|\boldsymbol{L}_0^{-1}\right\|_1\right\} \sup_{p\geq1}\left\{\left\|\boldsymbol{L}_0^{-1}\right\|_\infty\right\}} \leq C_L,
\end{aligned} \tag{S5.4}$$

for some finite positive constants C_2 and C_3.

By Condition 2, we have for any $k = 0, \cdots, K$,

$$\left|\tilde{\boldsymbol{L}}_0^\top \tilde{\boldsymbol{W}}_k \tilde{\boldsymbol{L}}_0\right|_\infty \leq \left\|\tilde{\boldsymbol{L}}_0^\top \tilde{\boldsymbol{W}}_k \tilde{\boldsymbol{L}}_0\right\|_1 \leq \sup_{p\geq 1}\{\|\boldsymbol{L}_0\|_\infty\} \sup_{p\geq 1}\{\|\boldsymbol{W}_k\|_1\} \sup_{p\geq 1}\{\|\boldsymbol{L}_0\|_1\} \leq C_L^2 C_W,$$

which implies

$$\left\|\boldsymbol{\mathcal{B}}(\boldsymbol{\alpha}^{(0)})\right\|_1 = \max_{0\leq k\leq K}\left\{\sum_{i=1}^n \sum_{j=1}^p \left|(\tilde{\boldsymbol{L}}_0^\top \tilde{\boldsymbol{W}}_k \tilde{\boldsymbol{L}}_0)_{(i-1)p+j,(i-1)p+j}\right|\right\} \leq np\left(C_L^2 C_W\right), \text{ and}$$

$$\left\|\boldsymbol{\mathcal{B}}(\boldsymbol{\alpha}^{(0)})\right\|_\infty = \max_{1\leq i\leq n;\ 1\leq j\leq p}\left\{\sum_{k=0}^K \left|(\tilde{\boldsymbol{L}}_0^\top \tilde{\boldsymbol{W}}_k \tilde{\boldsymbol{L}}_0)_{(i-1)p+j,(i-1)p+j}\right|\right\} \leq (K+1)\left(C_L^2 C_W\right),$$

by recalling $(\tilde{\boldsymbol{L}}_0^\top \tilde{\boldsymbol{W}}_k \tilde{\boldsymbol{L}}_0)_{l_1 l_2}$ is the (l_1, l_2)-th element of $\tilde{\boldsymbol{L}}_0^\top \tilde{\boldsymbol{W}}_k \tilde{\boldsymbol{L}}_0$ for $k = 0, \cdots, K$ and $l_1, l_2 = 1, \cdots, np$. Therefore, by applying inequality 4.67(e) in Seber (2008), we obtain

$$\|\boldsymbol{\mathcal{B}}(\boldsymbol{\alpha}^{(0)})\|_2 \leq \sqrt{\|\boldsymbol{\mathcal{B}}(\boldsymbol{\alpha}^{(0)})\|_1 \|\boldsymbol{\mathcal{B}}(\boldsymbol{\alpha}^{(0)})\|_\infty} \leq \sqrt{np}\{\sqrt{K+1}\left(C_L^2 C_W\right)\} \triangleq \sqrt{np} C_4, \tag{S5.5}$$

for some finite positive constant C_4. By Condition 2, we can use similar technique to show

$$\left|\left(\mathrm{tr}(\tilde{\boldsymbol{L}}_0^\top \tilde{\boldsymbol{W}}_{k_1} \tilde{\boldsymbol{\Sigma}}_0 \tilde{\boldsymbol{W}}_{k_2} \tilde{\boldsymbol{L}}_0)\right)_{(K+1)\times(K+1)}\right|_\infty \leq npC,$$

for some finite positive constant C. Then, by applying inequality 4.67(b) in

Seber (2008), we obtain

$$\left\|\left(\operatorname{tr}(\tilde{\boldsymbol{L}}_0^\top \tilde{\boldsymbol{W}}_{k_1} \tilde{\boldsymbol{\Sigma}}_0 \tilde{\boldsymbol{W}}_{k_2} \tilde{\boldsymbol{L}}_0)\right)_{(K+1)\times(K+1)}\right\|_2 \leq np\,(K+1)\,C \triangleq npC_5,$$

for some finite positive constant C_5. From (S5.3) – (S5.5), we obtain

$$\begin{aligned}
\left\|\operatorname{Cov}\left\{\boldsymbol{\psi}_{\boldsymbol{\beta}}(\boldsymbol{\vartheta}^{(0)})\right\}\right\|_2 &\leq \left\|\boldsymbol{D}^\top(\boldsymbol{\beta}^{(0)})\right\|_2 \left\|\boldsymbol{A}^{-1/2}(\boldsymbol{\beta}^{(0)})\right\|_2 \left\|\tilde{\boldsymbol{\Sigma}}_0^{-1}\right\|_2 \\
&\quad \times \left\|\boldsymbol{A}^{-1/2}(\boldsymbol{\beta}^{(0)})\right\|_2 \left\|\boldsymbol{D}(\boldsymbol{\beta}^{(0)})\right\|_2 \\
&\leq nC_1^2 C_2^2 C_3 = O(n), \\
\left\|\operatorname{Cov}\left\{\boldsymbol{\psi}_{\boldsymbol{\beta}}(\boldsymbol{\vartheta}^{(0)}), \boldsymbol{\psi}_{\boldsymbol{\alpha}}(\boldsymbol{\vartheta}^{(0)})\right\}\right\|_2 &= \left\|\operatorname{Cov}\left\{\boldsymbol{\psi}_{\boldsymbol{\alpha}}(\boldsymbol{\vartheta}^{(0)}), \boldsymbol{\psi}_{\boldsymbol{\beta}}(\boldsymbol{\vartheta}^{(0)})\right\}\right\|_2 \\
&\leq \left|\mu^{(3)}\right| \left\|\boldsymbol{D}^\top(\boldsymbol{\beta}^{(0)})\right\|_2 \left\|\boldsymbol{A}^{-1/2}(\boldsymbol{\beta}^{(0)})\right\|_2 \left\|\tilde{\boldsymbol{L}}_0^{-1\top}\right\|_2 \\
&\quad \times \left\|\boldsymbol{\mathcal{B}}(\boldsymbol{\alpha}^{(0)})\right\|_2 \\
&\leq \left|\mu^{(3)}\right| n\sqrt{p} C_1 C_2 C_L C_4 = O(n\sqrt{p}), \text{ and} \\
\left\|\operatorname{Cov}\left\{\boldsymbol{\psi}_{\boldsymbol{\alpha}}(\boldsymbol{\vartheta}^{(0)})\right\}\right\|_2 &\leq 2\left\|\left(\operatorname{tr}(\tilde{\boldsymbol{L}}_0^\top \tilde{\boldsymbol{W}}_{k_1} \tilde{\boldsymbol{\Sigma}}_0 \tilde{\boldsymbol{W}}_{k_2} \tilde{\boldsymbol{L}}_0)\right)_{(K+1)\times(K+1)}\right\|_2 \\
&\quad + \left|\mu^{(4)} - 3\right| \left\|\boldsymbol{\mathcal{B}}^\top(\boldsymbol{\alpha}^{(0)})\right\|_2 \left\|\boldsymbol{\mathcal{B}}(\boldsymbol{\alpha}^{(0)})\right\|_2 \\
&\leq 2npC_5 + \left|\mu^{(4)} - 3\right| npC_4^2 = O(np). \quad \text{(S5.6)}
\end{aligned}$$

By recalling $\boldsymbol{U} = \tilde{\boldsymbol{Z}}^{-1}\operatorname{Cov}\{\boldsymbol{\psi}(\boldsymbol{\vartheta}^{(0)})\}\tilde{\boldsymbol{Z}}^{-1}$ where $\tilde{\boldsymbol{Z}} = \operatorname{diag}(\sqrt{n}\boldsymbol{I}_{pd}, \sqrt{np}\boldsymbol{I}_{K+1})$, we obtain

$$
\begin{aligned}
\|\boldsymbol{U}\|_2 \leq & \left\|n^{-1}\mathrm{Cov}\left\{\boldsymbol{\psi}_{\boldsymbol{\beta}}(\boldsymbol{\vartheta}^{(0)})\right\}\right\|_2 + \left\|(n\sqrt{p})^{-1}\mathrm{Cov}\left\{\boldsymbol{\psi}_{\boldsymbol{\beta}}(\boldsymbol{\vartheta}^{(0)}), \boldsymbol{\psi}_{\boldsymbol{\alpha}}(\boldsymbol{\vartheta}^{(0)})\right\}\right\|_2 \\
& + \left\|(n\sqrt{p})^{-1}\mathrm{Cov}\left\{\boldsymbol{\psi}_{\boldsymbol{\alpha}}(\boldsymbol{\vartheta}^{(0)}), \boldsymbol{\psi}_{\boldsymbol{\beta}}(\boldsymbol{\vartheta}^{(0)})\right\}\right\|_2 + \left\|(np)^{-1}\mathrm{Cov}\left\{\boldsymbol{\psi}_{\boldsymbol{\alpha}}(\boldsymbol{\vartheta}^{(0)})\right\}\right\|_2 \\
= & \, O(1) + O(1) + O(1) + O(1) = O(1).
\end{aligned}
$$

Next, for the matrix $\boldsymbol{B} = \tilde{\boldsymbol{Z}}^{-1} E\{\partial \boldsymbol{\psi}(\boldsymbol{\vartheta}^{(0)})/\partial \boldsymbol{\beta}^\top\} \tilde{\boldsymbol{Z}}^{-1}$, by Conditions 2 – 4 and employing similar techniques to those used in showing (S5.6), we obtain

$$
\begin{aligned}
& \left\| E\left\{\frac{\partial \boldsymbol{\psi}_{\boldsymbol{\beta}}(\boldsymbol{\vartheta}^{(0)})}{\partial \boldsymbol{\beta}^\top}\right\}\right\|_2 = O(n), \left\| E\left\{\frac{\partial \boldsymbol{\psi}_{\boldsymbol{\beta}}(\boldsymbol{\vartheta}^{(0)})}{\partial \boldsymbol{\alpha}^\top}\right\}\right\|_2 = 0, \\
& \left\| E\left\{\frac{\partial \boldsymbol{\psi}_{\boldsymbol{\alpha}}(\boldsymbol{\vartheta}^{(0)})}{\partial \boldsymbol{\beta}^\top}\right\}\right\|_2 = O(n\sqrt{p}), \text{ and } \left\| E\left\{\frac{\partial \boldsymbol{\psi}_{\boldsymbol{\alpha}}(\boldsymbol{\vartheta}^{(0)})}{\partial \boldsymbol{\alpha}^\top}\right\}\right\|_2 = O(np).
\end{aligned}
$$

This implies

$$
\begin{aligned}
\|\boldsymbol{B}\|_2 \leq & \left\| n^{-1} E\left\{\frac{\partial \boldsymbol{\psi}_{\boldsymbol{\beta}}(\boldsymbol{\vartheta}^{(0)})}{\partial \boldsymbol{\beta}^\top}\right\}\right\|_2 + \left\|(n\sqrt{p})^{-1} E\left\{\frac{\partial \boldsymbol{\psi}_{\boldsymbol{\beta}}(\boldsymbol{\vartheta}^{(0)})}{\partial \boldsymbol{\alpha}^\top}\right\}\right\|_2 \\
& + \left\|(n\sqrt{p})^{-1} E\left\{\frac{\partial \boldsymbol{\psi}_{\boldsymbol{\alpha}}(\boldsymbol{\vartheta}^{(0)})}{\partial \boldsymbol{\beta}^\top}\right\}\right\|_2 + \left\|(np)^{-1} E\left\{\frac{\partial \boldsymbol{\psi}_{\boldsymbol{\alpha}}(\boldsymbol{\vartheta}^{(0)})}{\partial \boldsymbol{\alpha}^\top}\right\}\right\|_2 \\
= & \, O(1) + 0 + O(1) + O(1) = O(1),
\end{aligned}
$$

which is a necessary condition for Condition 5.

We then employ Lemma S4 to show the asymptotic normality of $\boldsymbol{\Omega}^{-1/2}(\boldsymbol{\Xi}^{(\mathcal{S})})\boldsymbol{\Xi}^{(\mathcal{S})}$ $\boldsymbol{B}^{-1}\tilde{\boldsymbol{Z}}^{-1}\boldsymbol{\psi}(\boldsymbol{\vartheta}^{(0)})$ via the following four steps.

STEP I. We represent $\boldsymbol{B}^{-1}$ as a block matrix below,

$$\boldsymbol{B}^{-1} = \begin{pmatrix} \boldsymbol{B}^{(11)} & \boldsymbol{0}_{pd\times(K+1)} \\ \boldsymbol{B}^{(21)} & \boldsymbol{B}^{(22)} \end{pmatrix},$$

where

$$\boldsymbol{B}^{(11)} = n\left[E\left\{\frac{\partial \boldsymbol{\psi}_{\boldsymbol{\beta}}(\boldsymbol{\vartheta}^{(0)})}{\partial \boldsymbol{\beta}^\top}\right\}\right]^{-1},$$
$$\boldsymbol{B}^{(21)} = -n\sqrt{p}\left[E\left\{\frac{\partial \boldsymbol{\psi}_{\boldsymbol{\alpha}}(\boldsymbol{\vartheta}^{(0)})}{\partial \boldsymbol{\alpha}^\top}\right\}\right]^{-1} E\left\{\frac{\partial \boldsymbol{\psi}_{\boldsymbol{\alpha}}(\boldsymbol{\vartheta}^{(0)})}{\partial \boldsymbol{\beta}^\top}\right\}\left[E\left\{\frac{\partial \boldsymbol{\psi}_{\boldsymbol{\beta}}(\boldsymbol{\vartheta}^{(0)})}{\partial \boldsymbol{\beta}^\top}\right\}\right]^{-1}, \text{ and}$$
$$\boldsymbol{B}^{(22)} = np\left[E\left\{\frac{\partial \boldsymbol{\psi}_{\boldsymbol{\alpha}}(\boldsymbol{\vartheta}^{(0)})}{\partial \boldsymbol{\alpha}^\top}\right\}\right]^{-1}.$$

Furthermore, let $\boldsymbol{\Gamma}^{(1)} = (\boldsymbol{T}^{(\mathcal{S})} \otimes \boldsymbol{I}_d)\boldsymbol{B}^{(11)}\boldsymbol{D}^\top(\boldsymbol{\beta}^{(0)})\boldsymbol{A}^{-1/2}(\boldsymbol{\beta}^{(0)})\tilde{\boldsymbol{L}}_0^{-1\top} = (\gamma_{li}^{(1)})_{qd\times np}$, $\boldsymbol{\Gamma}^{(2)} = \boldsymbol{B}^{(21)}\boldsymbol{D}^\top(\boldsymbol{\beta}^{(0)})\boldsymbol{A}^{-1/2}(\boldsymbol{\beta}^{(0)})\tilde{\boldsymbol{L}}_0^{-1\top} = (\gamma_{ki}^{(2)})_{(K+1)\times np}$, and $\boldsymbol{\Upsilon}_k = \sum_{l=0}^{K} b_{k+1,l+1}^{(22)} \tilde{\boldsymbol{L}}_0^\top \tilde{\boldsymbol{W}}_l \tilde{\boldsymbol{L}}_0 = (\upsilon_{ij}^{(k)})_{np\times np}$ where $b_{k+1,l+1}^{(22)}$ is the $(k+1, l+1)$-th element of $\boldsymbol{B}^{(22)}$ for $k, l = 0, \cdots, K$.

Next, we let

$$\boldsymbol{z}_{np,i} = \begin{pmatrix} \frac{1}{\sqrt{n}}\varepsilon_i \boldsymbol{\Gamma}_{\cdot i}^{(1)} \\ \frac{1}{\sqrt{n}}\varepsilon_i \boldsymbol{\Gamma}_{\cdot i}^{(2)} + \frac{1}{\sqrt{np}} \begin{pmatrix} 2\varepsilon_i \sum_{j=1}^{i-1} \upsilon_{ij}^{(0)} \varepsilon_j + \upsilon_{ii}^{(0)}(\varepsilon_i^2 - 1) \\ \vdots \\ 2\varepsilon_i \sum_{j=1}^{i-1} \upsilon_{ij}^{(K)} \varepsilon_j + \upsilon_{ii}^{(K)}(\varepsilon_i^2 - 1) \end{pmatrix} \end{pmatrix}, \qquad \text{(S5.7)}$$

where ε_i is the i-th element of $\boldsymbol{\varepsilon}$, $\boldsymbol{\Gamma}^{(1)}_{\cdot i}$ and $\boldsymbol{\Gamma}^{(2)}_{\cdot i}$ represent the i-th column of $\boldsymbol{\Gamma}^{(1)}$ and $\boldsymbol{\Gamma}^{(2)}$ respectively, for $i = 1, \cdots, np$. Then, we obtain

$$-\boldsymbol{\Omega}^{-1/2}(\boldsymbol{\Xi}^{(\mathcal{S})})\boldsymbol{\Xi}^{(\mathcal{S})}\boldsymbol{B}^{-1}\tilde{\boldsymbol{Z}}^{-1}\boldsymbol{\psi}(\boldsymbol{\vartheta}^{(0)}) = \sum_{i=1}^{np} -\boldsymbol{\Omega}^{-1/2}(\boldsymbol{\Xi}^{(\mathcal{S})})\boldsymbol{z}_{np,i}. \qquad \text{(S5.8)}$$

STEP II. For $i = 1, \cdots, np$, we define

$$\mathcal{F}_{np,i} = \sigma\langle \varepsilon_1, \cdots, \varepsilon_i \rangle,$$

which is the σ-algebra generated by $\varepsilon_1, \cdots, \varepsilon_i$. For the sake of completeness, define $\mathcal{F}_{np,0} = \{\emptyset, \Omega\}$. Using the properties of ε_i given in Condition 1, we have

$$E\left\{-\boldsymbol{\Omega}^{-1/2}(\boldsymbol{\Xi}^{(\mathcal{S})})\boldsymbol{z}_{np,i}|\mathcal{F}_{np,i-1}\right\} = \mathbf{0}_{qd+K+1}.$$

Hence, $\{(-\boldsymbol{\Omega}^{-1/2}(\boldsymbol{\Xi}^{(\mathcal{S})})\boldsymbol{z}_{np,i}, \mathcal{F}_{np,i}) : 1 \le i \le np\}$ forms a martingale difference array.

STEP III. In this step, we show $\sum_{i=1}^{np} E\| - \boldsymbol{\Omega}^{-1/2}(\boldsymbol{\Xi}^{(\mathcal{S})})\boldsymbol{z}_{np,i}\|_2^{2+\delta} = o(1) \to 0$ for some $\delta > 0$. In particular, we let $\delta \in (0, \eta/2]$ where η is defined in Condition 1. By using the expression of $\boldsymbol{z}_{np,i}$ in (S5.7) and representing $\boldsymbol{\Omega}^{-1}(\boldsymbol{\Xi}^{(\mathcal{S})})$ as a block matrix below,

$$\boldsymbol{\Omega}^{-1}(\boldsymbol{\Xi}^{(\mathcal{S})}) \triangleq \begin{pmatrix} \boldsymbol{\Omega}^{(11)} & \boldsymbol{\Omega}^{(21)\top} \\ \boldsymbol{\Omega}^{(21)} & \boldsymbol{\Omega}^{(22)} \end{pmatrix} = \begin{pmatrix} (\omega_{l_1 l_2}^{(11)})_{qd\times qd} & (\omega_{kl}^{(21)})_{(K+1)\times qd}^{\top} \\ (\omega_{kl}^{(21)})_{(K+1)\times qd} & (\omega_{k_1 k_2}^{(22)})_{(K+1)\times(K+1)} \end{pmatrix},$$

we obtain

$$\begin{aligned}\sum_{i=1}^{np} E\left\|-\boldsymbol{\Omega}^{-1/2}(\boldsymbol{\Xi}^{(\mathcal{S})})\boldsymbol{z}_{np,i}\right\|_2^{2+\delta} &= \sum_{i=1}^{np} E\left\{\left(-\boldsymbol{\Omega}^{-1/2}(\boldsymbol{\Xi}^{(\mathcal{S})})\boldsymbol{z}_{np,i}\right)^{\top}\left(-\boldsymbol{\Omega}^{-1/2}(\boldsymbol{\Xi}^{(\mathcal{S})})\boldsymbol{z}_{np,i}\right)\right\}^{\frac{2+\delta}{2}} \\ &= \sum_{i=1}^{np} E\left|\boldsymbol{z}_{np,i}^{\top}\boldsymbol{\Omega}^{-1}(\boldsymbol{\Xi}^{(\mathcal{S})})\boldsymbol{z}_{np,i}\right|^{\frac{2+\delta}{2}} \\ &= \sum_{i=1}^{np} E\left|\sum_{u=1}^{11} b_{iu}\right|^{\frac{2+\delta}{2}}, \qquad \text{(S5.9)}\end{aligned}$$

where

$$b_{i1} = \frac{1}{n}\varepsilon_i^2 \sum_{l_1=1}^{qd}\sum_{l_2=1}^{qd} \omega_{l_1 l_2}^{(11)} \gamma_{l_1 i}^{(1)} \gamma_{l_2 i}^{(1)}, \; b_{i2} = \frac{2}{n}\varepsilon_i^2 \sum_{k=0}^{K}\sum_{l=1}^{qd} \omega_{k+1,l}^{(21)} \gamma_{k+1,i}^{(2)} \gamma_{li}^{(1)},$$

$$b_{i3} = \frac{4}{n\sqrt{p}} \sum_{k=0}^{K}\sum_{l=1}^{qd} \omega_{k+1,l}^{(21)} \gamma_{li}^{(1)} \sum_{j=1}^{i-1} \upsilon_{ij}^{(k)} \varepsilon_i^2 \varepsilon_j, \; b_{i4} = \frac{2}{n\sqrt{p}}\left(\varepsilon_i^3 - \varepsilon_i\right) \sum_{k=0}^{K}\sum_{l=1}^{qd} \omega_{k+1,l}^{(21)} \upsilon_{ii}^{(k)} \gamma_{li}^{(1)},$$

$$b_{i5} = \frac{4}{np} \sum_{k_1=0}^{K}\sum_{k_2=0}^{K} \omega_{k_1+1,k_2+1}^{(22)} \sum_{j_1=1}^{i-1}\sum_{j_2=1}^{i-1} \upsilon_{ij_1}^{(k_1)} \upsilon_{ij_2}^{(k_2)} \varepsilon_i^2 \varepsilon_{j_1} \varepsilon_{j_2},$$

$$b_{i6} = \frac{2}{np} \sum_{k_1=0}^{K}\sum_{k_2=0}^{K} \omega_{k_1+1,k_2+1}^{(22)} \sum_{j_1=1}^{i-1} \upsilon_{ii}^{(k_2)} \upsilon_{ij_1}^{(k_1)} \left(\varepsilon_i^3 - \varepsilon_i\right) \varepsilon_{j_1},$$

$$b_{i7} = \frac{2}{np} \sum_{k_1=0}^{K}\sum_{k_2=0}^{K} \omega_{k_1+1,k_2+1}^{(22)} \sum_{j_2=1}^{i-1} \upsilon_{ii}^{(k_1)} \upsilon_{ij_2}^{(k_2)} \left(\varepsilon_i^3 - \varepsilon_i\right) \varepsilon_{j_1},$$

$$b_{i8} = \frac{1}{np} \sum_{k_1=0}^{K} \sum_{k_2=0}^{K} \omega_{k_1+1,k_2+1}^{(22)} \upsilon_{ii}^{(k_1)} \upsilon_{ii}^{(k_2)} \left(\varepsilon_i^4 - 2\varepsilon_i^2 + 1 \right),$$

$$b_{i9} = \frac{1}{n} \varepsilon_i^2 \sum_{k_1=0}^{K} \sum_{k_2=0}^{K} \omega_{k_1+1,k_2+1}^{(22)} \gamma_{k_1+1,i}^{(2)} \gamma_{k_2+1,i}^{(2)},$$

$$b_{i,10} = \frac{4}{n\sqrt{p}} \sum_{k_1=0}^{K} \sum_{k_2=0}^{K} \omega_{k_1+1,k_2+1}^{(22)} \gamma_{k_1+1,i}^{(2)} \sum_{j=1}^{i-1} \upsilon_{ij}^{(k_2)} \varepsilon_i^2 \varepsilon_j, \text{ and}$$

$$b_{i,11} = \frac{2}{n\sqrt{p}} (\varepsilon_i^3 - \varepsilon_i) \sum_{k_1=0}^{K} \sum_{k_2=0}^{K} \omega_{k_1+1,k_2+1}^{(22)} \upsilon_{ii}^{(k_2)} \gamma_{k_1+1,i}^{(2)}.$$

Next, we show $\sum_{i=1}^{np} E\| - \boldsymbol{\Omega}^{-1/2}(\boldsymbol{\Xi}^{(\mathcal{S})})\boldsymbol{z}_{np,i}\|_2^{2+\delta} = o(1) \to 0$ for $\delta \in (0, \eta/2]$ via the following eleven steps, (IIIa) – (IIIk). Before presenting (IIIa) – (IIIk), we first study the properties of $\boldsymbol{\Gamma}^{(1)}$, $\boldsymbol{\Gamma}^{(2)}$ and $\boldsymbol{\Upsilon}_k$ matrices. By (S5.2), (S5.4), Conditions 2 – 5 and Lemma S6, the spectral norm of the s-th column of $\boldsymbol{\Gamma}^{(1)}$ for any $s = 1, \cdots, np$ is given as

$$\begin{aligned}
\left\|\boldsymbol{\Gamma}_{\cdot s}^{(1)}\right\|_2 &= \sqrt{\sum_{l=1}^{qd} (\gamma_{ls}^{(1)})^2} = \left\|(\boldsymbol{T}^{(\mathcal{S})} \otimes \boldsymbol{I}_d)\boldsymbol{B}^{(11)}\boldsymbol{D}^\top(\boldsymbol{\beta}^{(0)})\boldsymbol{A}^{-1/2}(\boldsymbol{\beta}^{(0)})(\tilde{\boldsymbol{L}}_{0_{s\cdot}}^{-1})^\top\right\|_2 \\
&\le (qd)^{\frac{1}{2}} \left|(\boldsymbol{T}^{(\mathcal{S})} \otimes \boldsymbol{I}_d)\boldsymbol{B}^{(11)}\right|_\infty \|\boldsymbol{I}_{qd}\|_\infty \left\|\boldsymbol{D}^\top(\boldsymbol{\beta}^{(0)})\boldsymbol{A}^{-1/2}(\boldsymbol{\beta}^{(0)})(\tilde{\boldsymbol{L}}_{0_{s\cdot}}^{-1})^\top\right\|_1 \\
&\le (qd)^{\frac{1}{2}} \left|\boldsymbol{B}^{-1}\right|_\infty \|\boldsymbol{I}_{qd}\|_\infty \left\|\boldsymbol{D}(\boldsymbol{\beta}^{(0)})\right\|_\infty \left\|\boldsymbol{A}^{-1/2}(\boldsymbol{\beta}^{(0)})\right\|_1 \left\|\tilde{\boldsymbol{L}}_0^{-1}\right\|_\infty \\
&\le (qd)^{\frac{1}{2}} C_B\,(1)\, dC_g C_X \frac{1}{\sqrt{C_h}} C_L \triangleq C_6,
\end{aligned}$$

for some finite positive constant C_6 and $\tilde{\boldsymbol{L}}_{0_{s\cdot}}^{-1}$ denotes the s-th row of $\tilde{\boldsymbol{L}}_0^{-1}$. This, together with inequality 4.57(b) in Seber (2008), implies $\|\boldsymbol{\Gamma}_{\cdot s}^{(1)}\|_\infty \le C_6$ for any

$s = 1, \cdots, np$, which leads to

$$\left|\mathbf{\Gamma}^{(1)}\right|_\infty = O(1).$$

Furthermore, by (S5.3) – (S5.4), Condition 5 and Lemma S10, we obtain

$$\begin{aligned}
\sup_{n\geq 1, p\geq 1}\left\|\mathbf{\Gamma}^{(1)}\right\|_2 &\leq \sup_{n\geq 1, p\geq 1}\left\{\left\|\boldsymbol{T}^{(\mathcal{S})}\otimes \boldsymbol{I}_d\right\|_2\left\|\boldsymbol{B}^{(11)}\right\|_2\left\|\boldsymbol{D}^\top(\boldsymbol{\beta}^{(0)})\right\|_2\left\|\boldsymbol{A}^{-1/2}(\boldsymbol{\beta}^{(0)})\right\|_2\left\|\tilde{\boldsymbol{L}}_0^{-1\top}\right\|_2\right\}\\
&\leq \sup_{n\geq 1, p\geq 1}\left\{\left\|\boldsymbol{T}^{(\mathcal{S})}\otimes \boldsymbol{I}_d\right\|_2\left\|\boldsymbol{B}^{-1}\right\|_2\left\|\boldsymbol{D}(\boldsymbol{\beta}^{(0)})\right\|_2\left\|\boldsymbol{A}^{-1/2}(\boldsymbol{\beta}^{(0)})\right\|_2\left\|\tilde{\boldsymbol{L}}_0^{-1\top}\right\|_2\right\}\\
&\leq 1C_B\sqrt{n}C_1C_2C_L = O(\sqrt{n}). \qquad \text{(S5.10)}
\end{aligned}$$

This, together with Lemma S5, implies

$$\begin{aligned}
\frac{1}{n^{\frac{2+\delta}{2}}}\sum_{i=1}^{np}\left(\sum_{l=1}^{qd}\left|\gamma_{li}^{(1)}\right|\right)^{2+\delta} &\leq \frac{1}{n^{\frac{2+\delta}{2}}}\sum_{i=1}^{np}(qd)^{1+\delta}\sum_{l=1}^{qd}\left|\gamma_{li}^{(1)}\right|^{2+\delta}\\
&= (qd)^{1+\delta}\sum_{l=1}^{qd}\frac{1}{n^{\frac{2+\delta}{2}}}\sum_{i=1}^{np}\left|\gamma_{li}^{(1)}\right|^{2+\delta}\\
&= (qd)^{1+\delta}\sum_{l=1}^{qd}\frac{1}{n^{\frac{2+\delta}{2}}}\sum_{i=1}^{np}\left|\gamma_{li}^{(1)}\right|^{2}\left|\gamma_{li}^{(1)}\right|^{\delta}\\
&\leq (qd)^{1+\delta}\sum_{l=1}^{qd}\frac{1}{n^{\frac{2+\delta}{2}}}\sum_{i=1}^{np}\left|\gamma_{li}^{(1)}\right|^{2}\left|\mathbf{\Gamma}^{(1)}\right|_\infty^{\delta}\\
&= (qd)^{1+\delta}\sum_{l=1}^{qd}\frac{1}{n^{\frac{2+\delta}{2}}}\left|\mathbf{\Gamma}^{(1)}\right|_\infty^{\delta}\left\|\mathbf{\Gamma}^{(1)\top}\boldsymbol{d}_l\right\|_2^2\\
&\leq (qd)^{1+\delta}\sum_{l=1}^{qd}\frac{1}{n^{\frac{2+\delta}{2}}}\left|\mathbf{\Gamma}^{(1)}\right|_\infty^{\delta}\left\|\mathbf{\Gamma}^{(1)}\right\|_2^2\left\|\boldsymbol{d}_l\right\|_2^2\\
&= (qd)^{2+\delta}\frac{1}{n^{\frac{2+\delta}{2}}}\left|\mathbf{\Gamma}^{(1)}\right|_\infty^{\delta}\left\|\mathbf{\Gamma}^{(1)}\right\|_2^2
\end{aligned}$$

$$= (qd)^{2+\delta} \frac{1}{n^{\frac{2+\delta}{2}}} \{O(1)\}^{\delta} \{O(\sqrt{n})\}^2$$
$$= O\left(\frac{1}{n^{\frac{\delta}{2}}}\right), \tag{S5.11}$$

for $\delta \in (0, \eta/2]$, where $\boldsymbol{d}_l$ is the l-th column of $\boldsymbol{I}_{qd}$. By using similar techniques as above, we obtain

$$\left|\boldsymbol{\Gamma}^{(2)}\right|_\infty = O(1), \sup_{n\geq 1, p\geq 1} \left\|\boldsymbol{\Gamma}^{(2)}\right\|_2 = O(\sqrt{n}), \text{and } \frac{1}{n^{\frac{2+\delta}{2}}} \sum_{i=1}^{np} \left(\sum_{k=0}^{K} \left|\gamma_{k+1,i}^{(2)}\right|\right)^{2+\delta} = O\left(\frac{1}{n^{\frac{\delta}{2}}}\right), \tag{S5.12}$$

under Conditions 2 – 5. Then, by Conditions 2 and 5, we have for any $k = 0, \cdots, K$,

$$\sup_{n\geq 1, p\geq 1} \|\boldsymbol{\Upsilon}_k\|_1 = \sup_{n\geq 1, p\geq 1} \left\{ \left\| \sum_{l=0}^{K} b_{k+1,l+1}^{(22)} \tilde{\boldsymbol{L}}_0^\top \tilde{\boldsymbol{W}}_l \tilde{\boldsymbol{L}}_0 \right\|_1 \right\}$$
$$\leq \sup_{n\geq 1, p\geq 1} \left\{ \sum_{l=0}^{K} \left| b_{k+1,l+1}^{(22)} \right| \left\| \tilde{\boldsymbol{L}}_0 \right\|_\infty \left\| \tilde{\boldsymbol{W}}_l \right\|_1 \left\| \tilde{\boldsymbol{L}}_0 \right\|_1 \right\}$$
$$\leq (K+1)\, C_B C_L^2 C_W = O(1). \tag{S5.13}$$

It is worth noting that the first part of (S5.12) provides uniform bound for all elements of $\boldsymbol{\Gamma}^{(2)}$, while (S5.13) gives a uniform bound on the matrix norms of $\boldsymbol{\Upsilon}_k$ matrices for all $k = 0, \cdots, K$.

Let $K_\varepsilon = \left(E\,|\varepsilon_i|^{4+\eta}\right)^{1/(4+\eta)}$ where η is defined in Condition 1.

(IIIa) By Hölder's inequality, we have for $\delta \in (0, \eta/2]$,

$$\begin{aligned}
\left(E\left|b_{i1}\right|^{\frac{2+\delta}{2}}\right)^{\frac{2}{2+\delta}} &= \left(E\left|\frac{1}{n}\varepsilon_i^2\sum_{l_1=1}^{qd}\sum_{l_2=1}^{qd}\omega_{l_1l_2}^{(11)}\gamma_{l_1i}^{(1)}\gamma_{l_2i}^{(1)}\right|^{\frac{2+\delta}{2}}\right)^{\frac{2}{2+\delta}} \\
&= \frac{1}{n}\left|\sum_{l_1=1}^{qd}\sum_{l_2=1}^{qd}\omega_{l_1l_2}^{(11)}\gamma_{l_1i}^{(1)}\gamma_{l_2i}^{(1)}\right|\left(E\left|\varepsilon_i\right|^{2+\delta}\right)^{\frac{2}{2+\delta}} \\
&\leq \frac{1}{n}\left|\sum_{l_1=1}^{qd}\sum_{l_2=1}^{qd}\omega_{l_1l_2}^{(11)}\gamma_{l_1i}^{(1)}\gamma_{l_2i}^{(1)}\right|\left(E\left|\varepsilon_i\right|^{4+\eta}\right)^{\frac{2}{4+\eta}} \\
&\leq \frac{1}{n}K_\varepsilon^2\sum_{l_1=1}^{qd}\sum_{l_2=1}^{qd}\left|\omega_{l_1l_2}^{(11)}\right|\left|\gamma_{l_1i}^{(1)}\right|\left|\gamma_{l_2i}^{(1)}\right| \\
&\leq \frac{1}{n}K_\varepsilon^2\left|\boldsymbol{\Omega}^{(11)}\right|_\infty\sum_{l_1=1}^{qd}\sum_{l_2=1}^{qd}\left|\gamma_{l_1i}^{(1)}\right|\left|\gamma_{l_2i}^{(1)}\right| \\
&= \frac{1}{n}K_\varepsilon^2\left|\boldsymbol{\Omega}^{(11)}\right|_\infty\left(\sum_{l=1}^{qd}\left|\gamma_{li}^{(1)}\right|\right)^2 \\
&\triangleq g_{i1}.
\end{aligned}$$

By (S5.11), Conditions 1 and 5, we obtain

$$\begin{aligned}
\sum_{i=1}^{np} g_{i1}^{\frac{2+\delta}{2}} &= \frac{1}{n^{\frac{2+\delta}{2}}}K_\varepsilon^{2+\delta}\left|\boldsymbol{\Omega}^{(11)}\right|^{\frac{2+\delta}{2}}\sum_{i=1}^{np}\left(\sum_{l=1}^{qd}\left|\gamma_{li}^{(1)}\right|\right)^{2+\delta} \\
&= \{O(1)\}^{2+\delta}\{O(1)\}^{\frac{2+\delta}{2}}\frac{1}{n^{\frac{2+\delta}{2}}}\sum_{i=1}^{np}\left\{\sum_{l=1}^{qd}\left|\gamma_{li}^{(1)}\right|\right\}^{2+\delta} \\
&= \{O(1)\}^{2+\delta}\{O(1)\}^{\frac{2+\delta}{2}}O\left(\frac{1}{n^{\frac{\delta}{2}}}\right) \\
&= O\left(\frac{1}{n^{\frac{\delta}{2}}}\right) \to 0,
\end{aligned}$$

as $n \to \infty$ and $p = o(n^{1/2})$.

(IIIb) By Hölder's inequality, we have for $\delta \in (0, \eta/2]$,

$$\begin{aligned}
\left(E\left|b_{i2}\right|^{\frac{2+\delta}{2}}\right)^{\frac{2}{2+\delta}} &= \left(E\left|\frac{2}{n}\varepsilon_i^2\sum_{k=0}^{K}\sum_{l=1}^{qd}\omega_{k+1,l}^{(21)}\gamma_{k+1,i}^{(2)}\gamma_{li}^{(1)}\right|^{\frac{2+\delta}{2}}\right)^{\frac{2}{2+\delta}}\\
&= \frac{2}{n}\left|\sum_{k=0}^{K}\sum_{l=1}^{qd}\omega_{k+1,l}^{(21)}\gamma_{k+1,i}^{(2)}\gamma_{li}^{(1)}\right|\left(E\left|\varepsilon_i\right|^{2+\delta}\right)^{\frac{2}{2+\delta}}\\
&\leq \frac{2}{n}\left|\sum_{k=0}^{K}\sum_{l=1}^{qd}\omega_{k+1,l}^{(21)}\gamma_{k+1,i}^{(2)}\gamma_{li}^{(1)}\right|\left(E\left|\varepsilon_i\right|^{4+\eta}\right)^{\frac{2}{4+\eta}}\\
&\leq \frac{2}{n}K_\varepsilon^2\sum_{k=0}^{K}\sum_{l=1}^{qd}\left|\omega_{k+1,l}^{(21)}\right|\left|\gamma_{k+1,i}^{(2)}\right|\left|\gamma_{li}^{(1)}\right|\\
&\leq \frac{2}{n}K_\varepsilon^2\left|\mathbf{\Omega}^{(21)}\right|_\infty\left(\sum_{k=0}^{K}\left|\gamma_{k+1,i}^{(2)}\right|\right)\left(\sum_{l=1}^{qd}\left|\gamma_{li}^{(1)}\right|\right)\\
&\triangleq g_{i2}.
\end{aligned}$$

By (S5.11) – (S5.12), Conditions 1 and 5, we obtain

$$\begin{aligned}
\sum_{i=1}^{np} g_{i2}^{\frac{2+\delta}{2}} &= 2^{\frac{2+\delta}{2}}K_\varepsilon^{2+\delta}\left|\mathbf{\Omega}^{(21)}\right|_\infty^{\frac{2+\delta}{2}}\frac{1}{n^{\frac{2+\delta}{2}}}\sum_{i=1}^{np}\left(\sum_{k=0}^{K}\left|\gamma_{k+1,i}^{(2)}\right|\right)^{\frac{2+\delta}{2}}\left(\sum_{l=1}^{qd}\left|\gamma_{li}^{(1)}\right|\right)^{\frac{2+\delta}{2}}\\
&\leq 2^{\frac{2+\delta}{2}}K_\varepsilon^{2+\delta}\left|\mathbf{\Omega}^{(21)}\right|_\infty^{\frac{2+\delta}{2}}\left\{\sum_{k=0}^{K}\left|\mathbf{\Gamma}^{(2)}\right|_\infty\right\}^{\frac{2+\delta}{2}}\frac{1}{n^{\frac{2+\delta}{2}}}\sum_{i=1}^{np}\left\{\sum_{l=1}^{qd}\left|\gamma_{li}^{(1)}\right|\right\}^{\frac{2+\delta}{2}}\\
&= 2^{\frac{2+\delta}{2}}\left\{O(1)\right\}^{2+\delta}\left\{O(1)\right\}^{\frac{2+\delta}{2}}\left\{\sum_{k=0}^{K}O(1)\right\}^{\frac{2+\delta}{2}}O\left(\frac{1}{n^{\frac{\delta}{2}}}\right)\\
&= 2^{\frac{2+\delta}{2}}O\left(\frac{1}{n^{\frac{\delta}{2}}}\right) \to 0,
\end{aligned}$$

as $n \to \infty$ and $p = o(n^{1/2})$.

(IIIc) By Minkowski's inequality and Hölder's inequality, we have for $\delta \in (0, \eta/2]$,

$$\begin{aligned}
\left(E\left|b_{i3}\right|^{\frac{2+\delta}{2}}\right)^{\frac{2}{2+\delta}} &= \left(E\left|\frac{4}{n\sqrt{p}}\sum_{k=0}^{K}\sum_{l=1}^{qd}\omega_{k+1,l}^{(21)}\gamma_{li}^{(1)}\sum_{j=1}^{i-1}\upsilon_{ij}^{(k)}\varepsilon_i^2\varepsilon_j\right|^{\frac{2+\delta}{2}}\right)^{\frac{2}{2+\delta}} \\
&\leq \frac{4}{n\sqrt{p}}\sum_{k=0}^{K}\sum_{l=1}^{qd}\left|\omega_{k+1,l}^{(21)}\right|\left|\gamma_{li}^{(1)}\right|\sum_{j=1}^{i-1}\left|\upsilon_{ij}^{(k)}\right|\left(E\left|\varepsilon_i^2\varepsilon_j\right|^{\frac{2+\delta}{2}}\right)^{\frac{2}{2+\delta}} \\
&= \frac{4}{n\sqrt{p}}\sum_{k=0}^{K}\sum_{l=1}^{qd}\left|\omega_{k+1,l}^{(21)}\right|\left|\gamma_{li}^{(1)}\right|\sum_{j=1}^{i-1}\left|\upsilon_{ij}^{(k)}\right|\left(E\left|\varepsilon_i\right|^{2+\delta}\right)^{\frac{2}{2+\delta}}\left(E\left|\varepsilon_j\right|^{\frac{2+\delta}{2}}\right)^{\frac{2}{2+\delta}} \\
&\leq \frac{4}{n\sqrt{p}}\sum_{k=0}^{K}\sum_{l=1}^{qd}\left|\omega_{k+1,l}^{(21)}\right|\left|\gamma_{li}^{(1)}\right|\sum_{j=1}^{i-1}\left|\upsilon_{ij}^{(k)}\right|\left(E\left|\varepsilon_i\right|^{4+\eta}\right)^{\frac{2}{4+\eta}}\left(E\left|\varepsilon_j\right|^{4+\eta}\right)^{\frac{1}{4+\eta}} \\
&\leq \frac{4}{n\sqrt{p}}K_\varepsilon^3\left|\mathbf{\Omega}^{(21)}\right|_\infty\sum_{k=0}^{K}\sum_{l=1}^{qd}\left|\gamma_{li}^{(1)}\right|\sum_{j=1}^{i-1}\left|\upsilon_{ij}^{(k)}\right| \\
&\leq \frac{4}{n\sqrt{p}}K_\varepsilon^3\left|\mathbf{\Omega}^{(21)}\right|_\infty\sum_{l=1}^{qd}\left|\gamma_{li}^{(1)}\right|\sum_{k=0}^{K}\sum_{j=1}^{i-1}\left|\upsilon_{ij}^{(k)}\right| \\
&\leq \frac{4}{n\sqrt{p}}K_\varepsilon^3\left|\mathbf{\Omega}^{(21)}\right|_\infty\left(\sum_{l=1}^{qd}\left|\gamma_{li}^{(1)}\right|\right)\left(\sum_{k=0}^{K}\sum_{j=1}^{i-1}\left|\upsilon_{ij}^{(k)}\right|\right) \\
&\triangleq g_{i3}.
\end{aligned}$$

By (S5.11), (S5.13), Conditions 1 and 5, we obtain

$$\sum_{i=1}^{np} g_{i3}^{\frac{2+\delta}{2}} = 4^{\frac{2+\delta}{2}}K_\varepsilon^{\frac{3(2+\delta)}{2}}\left|\mathbf{\Omega}^{(21)}\right|_\infty^{\frac{2+\delta}{2}}\frac{1}{n^{\frac{2+\delta}{2}}}\frac{1}{p^{\frac{2+\delta}{4}}}\sum_{i=1}^{np}\left(\sum_{l=1}^{qd}\left|\gamma_{li}^{(1)}\right|\right)^{\frac{2+\delta}{2}}\left(\sum_{k=0}^{K}\sum_{j=1}^{i-1}\left|\upsilon_{ij}^{(k)}\right|\right)^{\frac{2+\delta}{2}}$$

$$
\begin{aligned}
&\le 4^{\frac{2+\delta}{2}} K_\varepsilon^{\frac{3(2+\delta)}{2}} \left|\mathbf{\Omega}^{(21)}\right|_\infty^{\frac{2+\delta}{2}} \frac{1}{n^{\frac{2+\delta}{2}}} \frac{1}{p^{\frac{2+\delta}{4}}} \sum_{i=1}^{np} \left(\sum_{l=1}^{qd} \left|\gamma_{li}^{(1)}\right|\right)^{\frac{2+\delta}{2}} \left(\sum_{k=0}^{K} \sum_{j=1}^{np} \left|\upsilon_{ij}^{(k)}\right|\right)^{\frac{2+\delta}{2}} \\
&\le 4^{\frac{2+\delta}{2}} K_\varepsilon^{\frac{3(2+\delta)}{2}} \left|\mathbf{\Omega}^{(21)}\right|_\infty^{\frac{2+\delta}{2}} \left(\sum_{k=0}^{K} \|\mathbf{\Upsilon}_k\|_1\right)^{\frac{2+\delta}{2}} \frac{1}{n^{\frac{2+\delta}{2}}} \frac{1}{p^{\frac{2+\delta}{4}}} \sum_{i=1}^{np} \left(\sum_{l=1}^{qd} \left|\gamma_{li}^{(1)}\right|\right)^{\frac{2+\delta}{2}} \\
&= 4^{\frac{2+\delta}{2}} \left\{O(1)\right\}^{\frac{3(2+\delta)}{2}} \left\{O(1)\right\}^{\frac{2+\delta}{2}} \left\{\sum_{k=0}^{K} O(1)\right\}^{\frac{2+\delta}{2}} \frac{1}{p^{\frac{2+\delta}{4}}} O\left(\frac{1}{n^{\frac{\delta}{2}}}\right) \\
&= 4^{\frac{2+\delta}{2}} O(1) \frac{1}{p^{\frac{2+\delta}{4}}} O\left(\frac{1}{n^{\frac{\delta}{2}}}\right) \\
&= 4^{\frac{2+\delta}{2}} \frac{1}{p^{\frac{2+\delta}{4}}} O\left(\frac{1}{n^{\frac{\delta}{2}}}\right) \to 0,
\end{aligned}
$$

as $n \to \infty$ and $p = o(n^{1/2})$.

(IIId) By Minkowski's inequality and Hölder's inequality, we have for $\delta \in (0, \eta/2]$,

$$
\begin{aligned}
\left(E\left|b_{i4}\right|^{\frac{2+\delta}{2}}\right)^{\frac{2}{2+\delta}} &= \left\{E\left|\frac{2}{n\sqrt{p}}\left(\varepsilon_i^3 - \varepsilon_i\right) \sum_{k=0}^{K} \sum_{l=1}^{qd} \omega_{k+1,l}^{(21)} \upsilon_{ii}^{(k)} \gamma_{li}^{(1)}\right|^{\frac{2+\delta}{2}}\right\}^{\frac{2}{2+\delta}} \\
&= \frac{2}{n\sqrt{p}} \left|\sum_{k=0}^{K} \sum_{l=1}^{qd} \omega_{k+1,l}^{(21)} \upsilon_{ii}^{(k)} \gamma_{li}^{(1)}\right| \left(E\left|\varepsilon_i^3 - \varepsilon_i\right|^{\frac{2+\delta}{2}}\right)^{\frac{2}{2+\delta}} \\
&\le \frac{2}{n\sqrt{p}} \left|\sum_{k=0}^{K} \sum_{l=1}^{qd} \omega_{k+1,l}^{(21)} \upsilon_{ii}^{(k)} \gamma_{li}^{(1)}\right| \left\{\left(E\left|\varepsilon_i^3\right|^{\frac{2+\delta}{2}}\right)^{\frac{2}{2+\delta}} + \left(E\left|-\varepsilon_i\right|^{\frac{2+\delta}{2}}\right)^{\frac{2}{2+\delta}}\right\} \\
&= \frac{2}{n\sqrt{p}} \left|\sum_{k=0}^{K} \sum_{l=1}^{qd} \omega_{k+1,l}^{(21)} \upsilon_{ii}^{(k)} \gamma_{li}^{(1)}\right| \left\{\left(E\left|\varepsilon_i^3\right|^{\frac{2+\delta}{2}}\right)^{\frac{2}{2+\delta}} + \left(E\left|\varepsilon_i\right|^{\frac{2+\delta}{2}}\right)^{\frac{2}{2+\delta}}\right\} \\
&\le \frac{2}{n\sqrt{p}} \sum_{k=0}^{K} \sum_{l=1}^{qd} \left|\omega_{k+1,l}^{(21)}\right| \left|\upsilon_{ii}^{(k)}\right| \left|\gamma_{li}^{(1)}\right| \left\{\left(E\left|\varepsilon_i\right|^{4+\eta}\right)^{\frac{3}{4+\eta}} + \left(E\left|\varepsilon_i\right|^{4+\eta}\right)^{\frac{1}{4+\eta}}\right\} \\
&\le \frac{2}{n\sqrt{p}} \left\{\left(E\left|\varepsilon_i\right|^{4+\eta}\right)^{\frac{3}{4+\eta}} + \left(E\left|\varepsilon_i\right|^{4+\eta}\right)^{\frac{1}{4+\eta}}\right\} \left|\mathbf{\Omega}^{(21)}\right|_\infty \left(\sum_{k=0}^{K} \left|\upsilon_{ii}^{(k)}\right|\right) \left(\sum_{l=1}^{qd} \left|\gamma_{li}^{(1)}\right|\right)
\end{aligned}
$$

$$\triangleq g_{i4}.$$

By (S5.11), (S5.13), Conditions 1 and 5, we obtain

$$\begin{aligned}
\sum_{i=1}^{np} g_{i4}^{\frac{2+\delta}{2}} &= 2^{\frac{2+\delta}{2}} \left(K_\varepsilon^3 + K_\varepsilon\right)^{\frac{2+\delta}{2}} \left|\boldsymbol{\Omega}^{(21)}\right|_\infty^{\frac{2+\delta}{2}} \frac{1}{p^{\frac{2+\delta}{4}}} \frac{1}{n^{\frac{2+\delta}{2}}} \sum_{i=1}^{np} \left(\sum_{k=0}^{K} \left|\upsilon_{ii}^{(k)}\right|\right)^{\frac{2+\delta}{2}} \left(\sum_{l=1}^{qd} \left|\gamma_{li}^{(1)}\right|\right)^{\frac{2+\delta}{2}} \\
&\le 2^{\frac{2+\delta}{2}} \left(K_\varepsilon^3 + K_\varepsilon\right)^{\frac{2+\delta}{2}} \left|\boldsymbol{\Omega}^{(21)}\right|_\infty^{\frac{2+\delta}{2}} \frac{1}{p^{\frac{2+\delta}{4}}} \frac{1}{n^{\frac{2+\delta}{2}}} \sum_{i=1}^{np} \left(\sum_{k=0}^{K} \sum_{j=1}^{np} \left|\upsilon_{ij}^{(k)}\right|\right)^{\frac{2+\delta}{2}} \left(\sum_{l=1}^{qd} \left|\gamma_{li}^{(1)}\right|\right)^{\frac{2+\delta}{2}} \\
&\le 2^{\frac{2+\delta}{2}} \left(K_\varepsilon^3 + K_\varepsilon\right)^{\frac{2+\delta}{2}} \left|\boldsymbol{\Omega}^{(21)}\right|_\infty^{\frac{2+\delta}{2}} \left(\sum_{k=0}^{K} \|\boldsymbol{\Upsilon}_k\|_1\right)^{\frac{2+\delta}{2}} \frac{1}{p^{\frac{2+\delta}{4}}} \frac{1}{n^{\frac{2+\delta}{2}}} \sum_{i=1}^{np} \left(\sum_{l=1}^{qd} \left|\gamma_{li}^{(1)}\right|\right)^{\frac{2+\delta}{2}} \\
&= 2^{\frac{2+\delta}{2}} \left[\{O(1)\}^3 + O(1)\right]^{\frac{2+\delta}{2}} \{O(1)\}^{\frac{2+\delta}{2}} \left\{\sum_{k=0}^{K} O(1)\right\}^{\frac{2+\delta}{2}} \frac{1}{p^{\frac{2+\delta}{4}}} O\left(\frac{1}{n^{\frac{\delta}{2}}}\right) \\
&= 2^{\frac{2+\delta}{2}} O(1) \frac{1}{p^{\frac{2+\delta}{4}}} O\left(\frac{1}{n^{\frac{\delta}{2}}}\right) \\
&= 2^{\frac{2+\delta}{2}} \frac{1}{p^{\frac{2+\delta}{4}}} O\left(\frac{1}{n^{\frac{\delta}{2}}}\right) \to 0,
\end{aligned}$$

as $n \to \infty$ and $p = o(n^{1/2})$.

(IIIe) By Minkowski's inequality and Hölder's inequality, we have for $\delta \in (0, \eta/2]$,

$$\begin{aligned}
&\left(E\left|b_{i5}\right|^{\frac{2+\delta}{2}}\right)^{\frac{2}{2+\delta}} \\
&= \left(E\left|\frac{4}{np} \sum_{k_1=0}^{K} \sum_{k_2=0}^{K} \omega_{k_1+1,k_2+1}^{(22)} \sum_{j_1=1}^{i-1} \sum_{j_2=1}^{i-1} \upsilon_{ij_1}^{(k_1)} \upsilon_{ij_2}^{(k_2)} \varepsilon_i^2 \varepsilon_{j_1} \varepsilon_{j_2}\right|^{\frac{2+\delta}{2}}\right)^{\frac{2}{2+\delta}} \\
&= \left(E\left|\frac{4}{np} \sum_{k_1=0}^{K} \sum_{k_2=0}^{K} \omega_{k_1+1,k_2+1}^{(22)} \sum_{j_1=1}^{i-1} \upsilon_{ij_1}^{(k_1)} \upsilon_{ij_1}^{(k_2)} \varepsilon_i^2 \varepsilon_{j_1}^2\right.\right.
\end{aligned}$$

$$
\begin{aligned}
&\left.\left. + \frac{4}{np} \sum_{k_1=0}^{K} \sum_{k_2=0}^{K} \omega^{(22)}_{k_1+1,k_2+1} \sum_{j_1=1}^{i-1} \sum_{\substack{j_2=1 \\ j_2 \neq j_1}}^{i-1} \upsilon^{(k_1)}_{ij_1} \upsilon^{(k_2)}_{ij_2} \varepsilon_i^2 \varepsilon_{j_1} \varepsilon_{j_2} \right|^{\frac{2+\delta}{2}} \right)^{\frac{2}{2+\delta}} \\
&\leq \left(E \left| \frac{4}{np} \sum_{k_1=0}^{K} \sum_{k_2=0}^{K} \omega^{(22)}_{k_1+1,k_2+1} \sum_{j_1=1}^{i-1} \upsilon^{(k_1)}_{ij_1} \upsilon^{(k_2)}_{ij_1} \varepsilon_i^2 \varepsilon_{j_1}^2 \right|^{\frac{2+\delta}{2}} \right)^{\frac{2}{2+\delta}} \\
&\quad + \left(E \left| \frac{4}{np} \sum_{k_1=0}^{K} \sum_{k_2=0}^{K} \omega^{(22)}_{k_1+1,k_2+1} \sum_{j_1=1}^{i-1} \sum_{\substack{j_2=1 \\ j_2 \neq j_1}}^{i-1} \upsilon^{(k_1)}_{ij_1} \upsilon^{(k_2)}_{ij_2} \varepsilon_i^2 \varepsilon_{j_1} \varepsilon_{j_2} \right|^{\frac{2+\delta}{2}} \right)^{\frac{2}{2+\delta}} \\
&\leq \frac{4}{np} \sum_{k_1=0}^{K} \sum_{k_2=0}^{K} \left| \omega^{(22)}_{k_1+1,k_2+1} \right| \sum_{j_1=1}^{i-1} \left| \upsilon^{(k_1)}_{ij_1} \right| \left| \upsilon^{(k_2)}_{ij_1} \right| \left(E \left| \varepsilon_i^2 \varepsilon_{j_1}^2 \right|^{\frac{2+\delta}{2}} \right)^{\frac{2}{2+\delta}} \\
&\quad + \frac{4}{np} \sum_{k_1=0}^{K} \sum_{k_2=0}^{K} \left| \omega^{(22)}_{k_1+1,k_2+1} \right| \sum_{j_1=1}^{i-1} \sum_{\substack{j_2=1 \\ j_2 \neq j_1}}^{i-1} \left| \upsilon^{(k_1)}_{ij_1} \right| \left| \upsilon^{(k_2)}_{ij_2} \right| \left(E \left| \varepsilon_i^2 \varepsilon_{j_1} \varepsilon_{j_2} \right|^{\frac{2+\delta}{2}} \right)^{\frac{2}{2+\delta}} \\
&= \frac{4}{np} \sum_{k_1=0}^{K} \sum_{k_2=0}^{K} \left| \omega^{(22)}_{k_1+1,k_2+1} \right| \sum_{j_1=1}^{i-1} \left| \upsilon^{(k_1)}_{ij_1} \right| \left| \upsilon^{(k_2)}_{ij_1} \right| \left(E \left| \varepsilon_i \right|^{2+\delta} \right)^{\frac{2}{2+\delta}} \left(E \left| \varepsilon_{j_1} \right|^{2+\delta} \right)^{\frac{2}{2+\delta}} \\
&\quad + \frac{4}{np} \sum_{k_1=0}^{K} \sum_{k_2=0}^{K} \left| \omega^{(22)}_{k_1+1,k_2+1} \right| \sum_{j_1=1}^{i-1} \sum_{\substack{j_2=1 \\ j_2 \neq j_1}}^{i-1} \left| \upsilon^{(k_1)}_{ij_1} \right| \left| \upsilon^{(k_2)}_{ij_2} \right| \left(E \left| \varepsilon_i \right|^{2+\delta} \right)^{\frac{2}{2+\delta}} \left(E \left| \varepsilon_{j_1} \right|^{\frac{2+\delta}{2}} \right)^{\frac{2}{2+\delta}} \left(E \left| \varepsilon_{j_2} \right|^{\frac{2+\delta}{2}} \right)^{\frac{2}{2+\delta}} \\
&\leq \frac{4}{np} \sum_{k_1=0}^{K} \sum_{k_2=0}^{K} \left| \omega^{(22)}_{k_1+1,k_2+1} \right| \sum_{j_1=1}^{i-1} \left| \upsilon^{(k_1)}_{ij_1} \right| \left| \upsilon^{(k_2)}_{ij_1} \right| \left(E \left| \varepsilon_i \right|^{4+\eta} \right)^{\frac{2}{4+\eta}} \left(E \left| \varepsilon_{j_1} \right|^{4+\eta} \right)^{\frac{2}{4+\eta}} \\
&\quad + \frac{4}{np} \sum_{k_1=0}^{K} \sum_{k_2=0}^{K} \left| \omega^{(22)}_{k_1+1,k_2+1} \right| \sum_{j_1=1}^{i-1} \sum_{\substack{j_2=1 \\ j_2 \neq j_1}}^{i-1} \left| \upsilon^{(k_1)}_{ij_1} \right| \left| \upsilon^{(k_2)}_{ij_2} \right| \left(E \left| \varepsilon_{j_1} \right|^{4+\eta} \right)^{\frac{2}{4+\eta}} \left(E \left| \varepsilon_{j_1} \right|^{4+\eta} \right)^{\frac{1}{4+\eta}} \left(E \left| \varepsilon_{j_1} \right|^{4+\eta} \right)^{\frac{1}{4+\eta}} \\
&= \frac{4}{np} K_\varepsilon^4 \sum_{k_1=0}^{K} \sum_{k_2=0}^{K} \left| \omega^{(22)}_{k_1+1,k_2+1} \right| \sum_{j_1=1}^{i-1} \left| \upsilon^{(k_1)}_{ij_1} \right| \left| \upsilon^{(k_2)}_{ij_1} \right| + \frac{4}{np} K_\varepsilon^4 \sum_{k_1=0}^{K} \sum_{k_2=0}^{K} \left| \omega^{(22)}_{k_1+1,k_2+1} \right| \sum_{j_1=1}^{i-1} \sum_{\substack{j_2=1 \\ j_2 \neq j_1}}^{i-1} \left| \upsilon^{(k_1)}_{ij_1} \right| \left| \upsilon^{(k_2)}_{ij_2} \right| \\
&= \frac{4}{np} K_\varepsilon^4 \sum_{k_1=0}^{K} \sum_{k_2=0}^{K} \left| \omega^{(22)}_{k_1+1,k_2+1} \right| \sum_{j_1=1}^{i-1} \sum_{j_2=1}^{i-1} \left| \upsilon^{(k_1)}_{ij_1} \right| \left| \upsilon^{(k_2)}_{ij_2} \right| \\
&\leq \frac{4}{np} K_\varepsilon^4 \left| \mathbf{\Omega}^{(22)} \right|_\infty \left(\sum_{k_1=0}^{K} \sum_{j_1=1}^{i-1} \left| \upsilon^{(k_1)}_{ij_1} \right| \right) \left(\sum_{k_2=0}^{K} \sum_{j_2=1}^{i-1} \left| \upsilon^{(k_2)}_{ij_2} \right| \right) \\
&\triangleq g_{i5}.
\end{aligned}
$$

By (S5.13), Conditions 1 and 5, we obtain

$$
\begin{aligned}
\sum_{i=1}^{np} g_{i5}^{\frac{2+\delta}{2}} &= 4^{\frac{2+\delta}{2}} K_{\varepsilon}^{4+2\delta} \left|\boldsymbol{\Omega}^{(22)}\right|_{\infty}^{\frac{2+\delta}{2}} \frac{1}{(np)^{\frac{2+\delta}{2}}} \sum_{i=1}^{np} \left(\sum_{k_1=0}^{K}\sum_{j_1=1}^{i-1}\left|\upsilon_{ij_1}^{(k_1)}\right|\right)^{\frac{2+\delta}{2}} \left(\sum_{k_2=0}^{K}\sum_{j_2=1}^{i-1}\left|\upsilon_{ij_2}^{(k_2)}\right|\right)^{\frac{2+\delta}{2}} \\
&\le 4^{\frac{2+\delta}{2}} K_{\varepsilon}^{4+2\delta} \left|\boldsymbol{\Omega}^{(22)}\right|_{\infty}^{\frac{2+\delta}{2}} \frac{1}{(np)^{\frac{2+\delta}{2}}} \sum_{i=1}^{np} \left(\sum_{k_1=0}^{K}\sum_{j_1=1}^{np}\left|\upsilon_{ij_1}^{(k_1)}\right|\right)^{\frac{2+\delta}{2}} \left(\sum_{k_2=0}^{K}\sum_{j_2=1}^{np}\left|\upsilon_{ij_2}^{(k_2)}\right|\right)^{\frac{2+\delta}{2}} \\
&\le 4^{\frac{2+\delta}{2}} K_{\varepsilon}^{4+2\delta} \left|\boldsymbol{\Omega}^{(22)}\right|_{\infty}^{\frac{2+\delta}{2}} \frac{1}{(np)^{\frac{2+\delta}{2}}} \sum_{i=1}^{np} \left(\sum_{k_1=0}^{K}\|\boldsymbol{\Upsilon}_{k_1}\|_1\right)^{\frac{2+\delta}{2}} \left(\sum_{k_2=0}^{K}\|\boldsymbol{\Upsilon}_{k_2}\|_1\right)^{\frac{2+\delta}{2}} \\
&= 4^{\frac{2+\delta}{2}} \{O(1)\}^{4+2\delta} \{O(1)\}^{\frac{2+\delta}{2}} \frac{1}{(np)^{\frac{2+\delta}{2}}} \sum_{i=1}^{np} \left\{\sum_{k_1=0}^{K} O(1)\right\}^{\frac{2+\delta}{2}} \left\{\sum_{k_2=0}^{K} O(1)\right\}^{\frac{2+\delta}{2}} \\
&= 4^{\frac{2+\delta}{2}} \frac{1}{(np)^{\frac{2+\delta}{2}}} O(np) \\
&= 4^{\frac{2+\delta}{2}} O\left(\frac{1}{(np)^{\frac{\delta}{2}}}\right) \to 0,
\end{aligned}
$$

as $n \to \infty$ and $p = o(n^{1/2})$.

(IIIf) By Minkowski's inequality and Hölder's inequality, we have for $\delta \in (0, \eta/2]$,

$$
\begin{aligned}
&\left(E\left|b_{i6}\right|^{\frac{2+\delta}{2}}\right)^{\frac{2}{2+\delta}} \\
&= \left\{E\left|\frac{2}{np}\sum_{k_1=0}^{K}\sum_{k_2=0}^{K}\omega_{k_1+1,k_2+1}^{(22)}\sum_{j_1=1}^{i-1}\upsilon_{ii}^{(k_2)}\upsilon_{ij_1}^{(k_1)}\left(\varepsilon_i^3-\varepsilon_i\right)\varepsilon_{j_1}\right|^{\frac{2+\delta}{2}}\right\}^{\frac{2}{2+\delta}} \\
&\le \frac{2}{np}\sum_{k_1=0}^{K}\sum_{k_2=0}^{K}\left|\omega_{k_1+1,k_2+1}^{(22)}\right|\sum_{j_1=1}^{i-1}\left|\upsilon_{ii}^{(k_2)}\right|\left|\upsilon_{ij_1}^{(k_1)}\right|\left\{E\left|\left(\varepsilon_i^3-\varepsilon_i\right)\varepsilon_{j_1}\right|^{\frac{2+\delta}{2}}\right\}^{\frac{2}{2+\delta}} \\
&= \frac{2}{np}\sum_{k_1=0}^{K}\sum_{k_2=0}^{K}\left|\omega_{k_1+1,k_2+1}^{(22)}\right|\sum_{j_1=1}^{i-1}\left|\upsilon_{ii}^{(k_2)}\right|\left|\upsilon_{ij_1}^{(k_1)}\right|\left(E\left|\varepsilon_i^3-\varepsilon_i\right|^{\frac{2+\delta}{2}}\right)^{\frac{2}{2+\delta}}\left(E\left|\varepsilon_{j_1}\right|^{\frac{2+\delta}{2}}\right)^{\frac{2}{2+\delta}}
\end{aligned}
$$

$$
\begin{aligned}
&\leq \frac{2}{np}\sum_{k_1=0}^{K}\sum_{k_2=0}^{K}\left|\omega_{k_1+1,k_2+1}^{(22)}\right|\sum_{j_1=1}^{i-1}\left|\upsilon_{ii}^{(k_2)}\right|\left|\upsilon_{ij_1}^{(k_1)}\right|\left\{\left(E\left|\varepsilon_i\right|^{\frac{3(2+\delta)}{2}}\right)^{\frac{2}{2+\delta}}+\left(E\left|\varepsilon_i\right|^{\frac{2+\delta}{2}}\right)^{\frac{2}{2+\delta}}\right\}\left(E\left|\varepsilon_{j_1}\right|^{\frac{2+\delta}{2}}\right)^{\frac{2}{2+\delta}}\\
&\leq \frac{2}{np}\sum_{k_1=0}^{K}\sum_{k_2=0}^{K}\left|\omega_{k_1+1,k_2+1}^{(22)}\right|\sum_{j_1=1}^{i-1}\left|\upsilon_{ii}^{(k_2)}\right|\left|\upsilon_{ij_1}^{(k_1)}\right|\left\{\left(E\left|\varepsilon_i\right|^{4+\eta}\right)^{\frac{3}{4+\eta}}+\left(E\left|\varepsilon_i\right|^{4+\eta}\right)^{\frac{1}{4+\eta}}\right\}\left(E\left|\varepsilon_{j_1}\right|^{4+\eta}\right)^{\frac{1}{4+\eta}}\\
&\leq \frac{2}{np}\left(K_\varepsilon^4+K_\varepsilon^2\right)\left|\mathbf{\Omega}^{(22)}\right|_\infty\left(\sum_{k_2=0}^{K}\left|\upsilon_{ii}^{(k_2)}\right|\right)\left(\sum_{k_1=0}^{K}\sum_{j_1=1}^{i-1}\left|\upsilon_{ij_1}^{(k_1)}\right|\right)\\
&\triangleq g_{i6}. \qquad\qquad \text{(S5.14)}
\end{aligned}
$$

By (S5.13), Conditions 1 and 5, we obtain

$$
\begin{aligned}
\sum_{i=1}^{np} g_{i6}^{\frac{2+\delta}{2}} &= 2^{\frac{2+\delta}{2}}\left(K_\varepsilon^4+K_\varepsilon^2\right)^{\frac{2+\delta}{2}}\left|\mathbf{\Omega}^{(22)}\right|_\infty^{\frac{2+\delta}{2}}\frac{1}{(np)^{\frac{2+\delta}{2}}}\sum_{i=1}^{np}\left(\sum_{k_2=0}^{K}\left|\upsilon_{ii}^{(k_2)}\right|\right)^{\frac{2+\delta}{2}}\left(\sum_{k_1=0}^{K}\sum_{j_1=1}^{i-1}\left|\upsilon_{ij_1}^{(k_1)}\right|\right)^{\frac{2+\delta}{2}}\\
&\leq 2^{\frac{2+\delta}{2}}\left(K_\varepsilon^4+K_\varepsilon^2\right)^{\frac{2+\delta}{2}}\left|\mathbf{\Omega}^{(22)}\right|_\infty^{\frac{2+\delta}{2}}\frac{1}{(np)^{\frac{2+\delta}{2}}}\sum_{i=1}^{np}\left(\sum_{k_2=0}^{K}\sum_{j=1}^{np}\left|\upsilon_{ij}^{(k_2)}\right|\right)^{\frac{2+\delta}{2}}\left(\sum_{k_1=0}^{K}\sum_{j_1=1}^{np}\left|\upsilon_{ij_1}^{(k_1)}\right|\right)^{\frac{2+\delta}{2}}\\
&\leq 2^{\frac{2+\delta}{2}}\left(K_\varepsilon^4+K_\varepsilon^2\right)^{\frac{2+\delta}{2}}\left|\mathbf{\Omega}^{(22)}\right|_\infty^{\frac{2+\delta}{2}}\frac{1}{(np)^{\frac{2+\delta}{2}}}\sum_{i=1}^{np}\left(\sum_{k_2=0}^{K}\left\|\mathbf{\Upsilon}_{k_2}\right\|_1\right)^{\frac{2+\delta}{2}}\left(\sum_{k_1=0}^{K}\left\|\mathbf{\Upsilon}_{k_1}\right\|_1\right)^{\frac{2+\delta}{2}}\\
&= 2^{\frac{2+\delta}{2}}\left[\{O(1)\}^4+\{O(1)\}^2\right]^{\frac{2+\delta}{2}}\{O(1)\}^{\frac{2+\delta}{2}}\frac{1}{(np)^{\frac{2+\delta}{2}}}\sum_{i=1}^{np}\left\{\sum_{k_2=0}^{K}O(1)\right\}^{\frac{2+\delta}{2}}\left\{\sum_{k_1=0}^{K}O(1)\right\}^{\frac{2+\delta}{2}}\\
&= 2^{\frac{2+\delta}{2}}\frac{1}{(np)^{\frac{2+\delta}{2}}}O(np)\\
&= 2^{\frac{2+\delta}{2}}O\left(\frac{1}{(np)^{\frac{\delta}{2}}}\right)\to 0,
\end{aligned}
$$

as $n\to\infty$ and $p=o(n^{1/2})$.

(IIIg) By Minkowski's inequality and Hölder's inequality and using similar techniques to derive g_{i6} in (S5.14), we have for $\delta \in (0, \eta/2]$,

$$\begin{aligned}\left(E\left|b_{i7}\right|^{\frac{2+\delta}{2}}\right)^{\frac{2}{2+\delta}} &= \left\{E\left|\frac{2}{np}\sum_{k_1=0}^{K}\sum_{k_2=0}^{K}\omega_{k_1+1,k_2+1}^{(22)}\sum_{j_2=1}^{i-1}\upsilon_{ii}^{(k_1)}\upsilon_{ij_2}^{(k_2)}\left(\varepsilon_i^3-\varepsilon_i\right)\varepsilon_{j_2}\right|^{\frac{2+\delta}{2}}\right\}^{\frac{2}{2+\delta}} \\ &\leq \frac{2}{np}\left(K_\varepsilon^4+K_\varepsilon^2\right)\left|\mathbf{\Omega}^{(22)}\right|_\infty\left(\sum_{k_1=0}^{K}\left|\upsilon_{ii}^{(k_1)}\right|\right)\left(\sum_{k_2=0}^{K}\sum_{j_2=1}^{i-1}\left|\upsilon_{ij_2}^{(k_2)}\right|\right) \\ &\triangleq g_{i7}.\end{aligned}$$

By (S5.13), Conditions 1 and 5, we obtain

$$\begin{aligned}\sum_{i=1}^{np} g_{i7}^{\frac{2+\delta}{2}} &= 2^{\frac{2+\delta}{2}}\left(K_\varepsilon^4+K_\varepsilon^2\right)^{\frac{2+\delta}{2}}\left|\mathbf{\Omega}^{(22)}\right|_\infty^{\frac{2+\delta}{2}}\frac{1}{(np)^{\frac{2+\delta}{2}}}\sum_{i=1}^{np}\left(\sum_{k_1=0}^{K}\left|\upsilon_{ii}^{(k_1)}\right|\right)^{\frac{2+\delta}{2}}\left(\sum_{k_2=0}^{K}\sum_{j_2=1}^{i-1}\left|\upsilon_{ij_2}^{(k_2)}\right|\right)^{\frac{2+\delta}{2}} \\ &\leq 2^{\frac{2+\delta}{2}}\left(K_\varepsilon^4+K_\varepsilon^2\right)^{\frac{2+\delta}{2}}\left|\mathbf{\Omega}^{(22)}\right|_\infty^{\frac{2+\delta}{2}}\frac{1}{(np)^{\frac{2+\delta}{2}}}\sum_{i=1}^{np}\left(\sum_{k_1=0}^{K}\sum_{j=1}^{np}\left|\upsilon_{ij}^{(k_1)}\right|\right)^{\frac{2+\delta}{2}}\left(\sum_{k_2=0}^{K}\sum_{j_2=1}^{np}\left|\upsilon_{ij_2}^{(k_2)}\right|\right)^{\frac{2+\delta}{2}} \\ &\leq 2^{\frac{2+\delta}{2}}\left(K_\varepsilon^4+K_\varepsilon^2\right)^{\frac{2+\delta}{2}}\left|\mathbf{\Omega}^{(22)}\right|_\infty^{\frac{2+\delta}{2}}\frac{1}{(np)^{\frac{2+\delta}{2}}}\sum_{i=1}^{np}\left(\sum_{k_1=0}^{K}\left\|\mathbf{\Upsilon}_{k_1}\right\|_1\right)^{\frac{2+\delta}{2}}\left(\sum_{k_2=0}^{K}\left\|\mathbf{\Upsilon}_{k_2}\right\|_1\right)^{\frac{2+\delta}{2}} \\ &= 2^{\frac{2+\delta}{2}}\left[\{O(1)\}^4+\{O(1)\}^2\right]^{\frac{2+\delta}{2}}\{O(1)\}^{\frac{2+\delta}{2}}\frac{1}{(np)^{\frac{2+\delta}{2}}}\sum_{i=1}^{np}\left\{\sum_{k_1=0}^{K}O(1)\right\}^{\frac{2+\delta}{2}}\left\{\sum_{k_2=0}^{K}O(1)\right\}^{\frac{2+\delta}{2}} \\ &= 2^{\frac{2+\delta}{2}}\frac{1}{(np)^{\frac{2+\delta}{2}}}O(np) \\ &= 2^{\frac{2+\delta}{2}}O\left(\frac{1}{(np)^{\frac{\delta}{2}}}\right)\to 0,\end{aligned}$$

as $n \to \infty$ and $p = o(n^{1/2})$.

(IIIh) By Minkowski's inequality and Hölder's inequality, we have for $\delta \in$

$(0, \eta/2]$,

$$\begin{aligned}
\left(E\left|b_{i8}\right|^{\frac{2+\delta}{2}}\right)^{\frac{2}{2+\delta}} &= \left\{E\left|\frac{1}{np}\left(\varepsilon_i^4 - 2\varepsilon_i^2 + 1\right)\sum_{k_1=0}^{K}\sum_{k_2=0}^{K}\omega_{k_1+1,k_2+1}^{(22)}\upsilon_{ii}^{(k_1)}\upsilon_{ii}^{(k_2)}\right|^{\frac{2+\delta}{2}}\right\}^{\frac{2}{2+\delta}} \\
&= \frac{1}{np}\left|\sum_{k_1=0}^{K}\sum_{k_2=0}^{K}\omega_{k_1+1,k_2+1}^{(22)}\upsilon_{ii}^{(k_1)}\upsilon_{ii}^{(k_2)}\right|\left(E\left|\varepsilon_i^4 - 2\varepsilon_i^2 + 1\right|^{\frac{2+\delta}{2}}\right)^{\frac{2}{2+\delta}} \\
&\leq \frac{1}{np}\sum_{k_1=0}^{K}\sum_{k_2=0}^{K}\left|\omega_{k_1+1,k_2+1}^{(22)}\right|\left|\upsilon_{ii}^{(k_1)}\right|\left|\upsilon_{ii}^{(k_2)}\right| \\
&\quad\times\left\{\left(E\left|\varepsilon_i\right|^{4+2\delta}\right)^{\frac{2}{2+\delta}} + 2\left(E\left|\varepsilon_i\right|^{2+\delta}\right)^{\frac{2}{2+\delta}} + \left(E\left|1\right|^{\frac{2+\delta}{2}}\right)^{\frac{2}{2+\delta}}\right\} \\
&\leq \frac{1}{np}\sum_{k_1=0}^{K}\sum_{k_2=0}^{K}\left|\omega_{k_1+1,k_2+1}^{(22)}\right|\left|\upsilon_{ii}^{(k_1)}\right|\left|\upsilon_{ii}^{(k_2)}\right|\left(K_\varepsilon^4 + K_\varepsilon^2 + 1\right) \\
&\leq \frac{1}{np}\left(K_\varepsilon^4 + K_\varepsilon^2 + 1\right)\left|\mathbf{\Omega}^{(22)}\right|_\infty\left(\sum_{k_1=0}^{K}\left|\upsilon_{ii}^{(k_1)}\right|\right)\left(\sum_{k_2=0}^{K}\left|\upsilon_{ii}^{(k_2)}\right|\right) \\
&\triangleq g_{i8}.
\end{aligned}$$

By (S5.13), Conditions 1 and 5, we obtain

$$\begin{aligned}
\sum_{i=1}^{np} g_{i8}^{\frac{2+\delta}{2}} &= \left(K_\varepsilon^4 + K_\varepsilon^2 + 1\right)^{\frac{2+\delta}{2}}\left|\mathbf{\Omega}^{(22)}\right|_\infty^{\frac{2+\delta}{2}}\frac{1}{(np)^{\frac{2+\delta}{2}}}\sum_{i=1}^{np}\left(\sum_{k_1=0}^{K}\left|\upsilon_{ii}^{(k_1)}\right|\right)^{\frac{2+\delta}{2}}\left(\sum_{k_2=0}^{K}\left|\upsilon_{ii}^{(k_2)}\right|\right)^{\frac{2+\delta}{2}} \\
&\leq \left(K_\varepsilon^4 + K_\varepsilon^2 + 1\right)^{\frac{2+\delta}{2}}\left|\mathbf{\Omega}^{(22)}\right|_\infty^{\frac{2+\delta}{2}}\frac{1}{(np)^{\frac{2+\delta}{2}}}\sum_{i=1}^{np}\left(\sum_{k_1=0}^{K}\sum_{j=1}^{np}\left|\upsilon_{ij}^{(k_1)}\right|\right)^{\frac{2+\delta}{2}}\left(\sum_{k_2=0}^{K}\sum_{j=1}^{np}\left|\upsilon_{ij}^{(k_2)}\right|\right)^{\frac{2+\delta}{2}} \\
&\leq \left(K_\varepsilon^4 + K_\varepsilon^2 + 1\right)^{\frac{2+\delta}{2}}\left|\mathbf{\Omega}^{(22)}\right|_\infty^{\frac{2+\delta}{2}}\frac{1}{(np)^{\frac{2+\delta}{2}}}\sum_{i=1}^{np}\left(\sum_{k_1=0}^{K}\left\|\mathbf{\Upsilon}_{k_1}\right\|_1\right)^{\frac{2+\delta}{2}}\left(\sum_{k_2=0}^{K}\left\|\mathbf{\Upsilon}_{k_2}\right\|_1\right)^{\frac{2+\delta}{2}} \\
&= \left[\{O(1)\}^4 + \{O(1)\}^2 + 1\right]^{\frac{2+\delta}{2}}\{O(1)\}^{\frac{2+\delta}{2}}\frac{1}{(np)^{\frac{2+\delta}{2}}}\sum_{i=1}^{np}\left\{\sum_{k_1=0}^{K}O(1)\right\}^{\frac{2+\delta}{2}}\left\{\sum_{k_2=0}^{K}O(1)\right\}^{\frac{2+\delta}{2}}
\end{aligned}$$

$$= \frac{1}{(np)^{\frac{2+\delta}{2}}} O(np)$$
$$= O\left(\frac{1}{(np)^{\frac{\delta}{2}}}\right) \to 0,$$

as $n \to \infty$ and $p = o(n^{1/2})$.

(IIIi) By using similar techniques to derive and study g_{i1}, we have for $\delta \in (0, \eta/2]$,

$$\left(E\left|b_{i9}\right|^{\frac{2+\delta}{2}}\right)^{\frac{2}{2+\delta}} \leq \frac{1}{n} K_{\varepsilon}^{2} \left|\mathbf{\Omega}^{(22)}\right|_{\infty} \left(\sum_{k=0}^{K} \left|\gamma_{k+1,i}^{(2)}\right|\right)^{2} \triangleq g_{i9},$$

and

$$\sum_{i=1}^{np} g_{i9}^{\frac{2+\delta}{2}} \leq O\left(\frac{1}{n^{\frac{\delta}{2}}}\right) \to 0,$$

as $n \to \infty$ and $p = o(n^{1/2})$.

(IIIj) By using similar techniques to derive and study g_{i3}, we have for $\delta \in (0, \eta/2]$,

$$\left(E\left|b_{i,10}\right|^{\frac{2+\delta}{2}}\right)^{\frac{2}{2+\delta}} \leq \frac{4}{n\sqrt{p}} K_{\varepsilon}^{3} \left|\mathbf{\Omega}^{(22)}\right|_{\infty} \left(\sum_{k_1=0}^{K} \left|\gamma_{k_1+1,i}^{(2)}\right|\right) \left(\sum_{k_2=0}^{K} \sum_{j=1}^{i-1} \left|\upsilon_{ij}^{(k_2)}\right|\right) \triangleq g_{i,10},$$

and

$$\sum_{i=1}^{np} g_{i,10}^{\frac{2+\delta}{2}} \leq 4^{\frac{2+\delta}{2}} \frac{1}{p^{\frac{2+\delta}{4}}} O\left(\frac{1}{n^{\frac{\delta}{2}}}\right) \to 0,$$

as $n \to \infty$ and $p = o(n^{1/2})$.

(IIIk) By using similar techniques to derive and study g_{i4}, we have for $\delta \in (0, \eta/2]$,

$$\left(E\left|b_{i,11}\right|^{\frac{2+\delta}{2}}\right)^{\frac{2}{2+\delta}} \leq \frac{2}{n\sqrt{p}}(K_\varepsilon^3 + K_\varepsilon)\left|\mathbf{\Omega}^{(22)}\right|_\infty \left(\sum_{k_1=0}^{K}\left|\gamma_{k_1+1,i}^{(2)}\right|\right)\left(\sum_{k_2=0}^{K}\left|\upsilon_{ii}^{(k_2)}\right|\right) \triangleq g_{i,11},$$

and

$$\sum_{i=1}^{np} g_{i,11}^{\frac{2+\delta}{2}} \leq 2^{\frac{2+\delta}{2}} \frac{1}{p^{\frac{2+\delta}{4}}} O\left(\frac{1}{n^{\frac{\delta}{2}}}\right) \to 0,$$

as $n \to \infty$ and $p = o(n^{1/2})$.

From (IIIa) – (IIIk), we have $(E\left|b_{iu}\right|^{(2+\delta)/2})^{2/(2+\delta)} \leq g_{iu}$ and $\sum_{i=1}^{np} g_{iu}^{(2+\delta)/2} = o(1) \to 0$ where $u = 1, \cdots, 11$ and $\delta \in (0, \eta/2]$. This, together with (S5.9), Minkowski's inequality and Lemma S5, implies

$$\begin{aligned}
0 \leq \sum_{i=1}^{np} E\left\|-\mathbf{\Omega}^{-1/2}(\mathbf{\Xi}^{(\mathcal{S})})\boldsymbol{z}_{np,i}\right\|_2^{2+\delta} &= \sum_{i=1}^{np} E\left|\sum_{u=1}^{11} b_{iu}\right|^{\frac{2+\delta}{2}} \\
&\leq \sum_{i=1}^{np}\left\{\sum_{u=1}^{11}\left(E\left|b_{iu}\right|^{\frac{2+\delta}{2}}\right)^{\frac{2}{2+\delta}}\right\}^{\frac{2+\delta}{2}} \\
&\leq \sum_{i=1}^{np}\left(\sum_{u=1}^{11} g_{iu}\right)^{\frac{2+\delta}{2}} \\
&\leq \sum_{i=1}^{np} 11^{\frac{\delta}{2}} \sum_{u=1}^{11}\left|g_{iu}\right|^{\frac{2+\delta}{2}} \\
&= 11^{\frac{\delta}{2}} \sum_{u=1}^{11}\sum_{i=1}^{np} g_{iu}^{\frac{2+\delta}{2}} \\
&= 11^{\frac{\delta}{2}} \sum_{u=1}^{11} o(1) \to 0,
\end{aligned}$$

as $n \to \infty$ and $p = o(n^{1/2})$. In summary, we showed $\sum_{i=1}^{np} E\|-\boldsymbol{\Omega}^{-1/2}(\boldsymbol{\Xi}^{(\mathcal{S})})\boldsymbol{z}_{np,i}\|_2^{2+\delta}$ $\to 0$ as $n \to \infty$ and $p = o(n^{1/2})$ for $\delta \in (0, \eta/2]$.

Step IV. In this step, we show

$$\sum_{i=1}^{np} E\left[\left\{-\boldsymbol{\Omega}^{-1/2}(\boldsymbol{\Xi}^{(\mathcal{S})})\boldsymbol{z}_{np,i}\right\}\left\{-\boldsymbol{\Omega}^{-1/2}(\boldsymbol{\Xi}^{(\mathcal{S})})\boldsymbol{z}_{np,i}\right\}^\top |\mathcal{F}_{np,i-1}\right] \xrightarrow{\mathrm{P}} \mathrm{Cov}\left\{\sum_{i=1}^{np} -\boldsymbol{\Omega}^{-1/2}(\boldsymbol{\Xi}^{(\mathcal{S})})\boldsymbol{z}_{np,i}\right\},$$

where

$$\begin{aligned}&\sum_{i=1}^{np} E\left[\left\{-\boldsymbol{\Omega}^{-1/2}(\boldsymbol{\Xi}^{(\mathcal{S})})\boldsymbol{z}_{np,i}\right\}\left\{-\boldsymbol{\Omega}^{-1/2}(\boldsymbol{\Xi}^{(\mathcal{S})})\boldsymbol{z}_{np,i}\right\}^\top |\mathcal{F}_{np,i-1}\right]\\ &= \boldsymbol{\Omega}^{-1/2}(\boldsymbol{\Xi}^{(\mathcal{S})})\left\{\sum_{i=1}^{np} E\left(\boldsymbol{z}_{np,i}\boldsymbol{z}_{np,i}^\top|\mathcal{F}_{np,i-1}\right)\right\}\boldsymbol{\Omega}^{-1/2\top}(\boldsymbol{\Xi}^{(\mathcal{S})}). \qquad \text{(S5.15)}\end{aligned}$$

It can also be shown

$$\mathrm{Cov}\left\{\sum_{i=1}^{np} -\boldsymbol{\Omega}^{-1/2}(\boldsymbol{\Xi}^{(\mathcal{S})})\boldsymbol{z}_{np,i}\right\} = \boldsymbol{\Omega}^{-1/2}(\boldsymbol{\Xi}^{(\mathcal{S})})\left\{\sum_{i=1}^{np} E\left(\boldsymbol{z}_{np,i}\boldsymbol{z}_{np,i}^\top\right)\right\}\boldsymbol{\Omega}^{-1/2\top}(\boldsymbol{\Xi}^{(\mathcal{S})}). \tag{S5.16}$$

Therefore, it suffices to show that

$$\boldsymbol{\Omega}^{-1/2}(\boldsymbol{\Xi}^{(\mathcal{S})})\left[\sum_{i=1}^{np}\left\{E\left(\boldsymbol{z}_{np,i}\boldsymbol{z}_{np,i}^\top|\mathcal{F}_{np,i-1}\right) - E\left(\boldsymbol{z}_{np,i}\boldsymbol{z}_{np,i}^\top\right)\right\}\right]\boldsymbol{\Omega}^{-1/2\top}(\boldsymbol{\Xi}^{(\mathcal{S})}) = o_{\mathrm{P}}(1),$$

where the left hand side consists of products of matrices with finite dimensions $(qd + K + 1) \times (qd + K + 1)$. By Condition 5, we obtain $\|\boldsymbol{\Omega}^{-1/2}(\boldsymbol{\Xi}^{(\mathcal{S})})\|_2 =$

$\|\boldsymbol{\Omega}^{-1/2\top}(\boldsymbol{\Xi}^{(\mathcal{S})})\|_2 = O(1)$, so we only have to show

$$\sum_{i=1}^{np} \left\{ E\left(\boldsymbol{z}_{np,i}\boldsymbol{z}_{np,i}^{\top}|\mathcal{F}_{np,i-1}\right) - E\left(\boldsymbol{z}_{np,i}\boldsymbol{z}_{np,i}^{\top}\right)\right\} = o_{\mathrm{P}}(1).$$

From (S5.7), we have

$$\sum_{i=1}^{np} \left\{ E\left(\boldsymbol{z}_{np,i}\boldsymbol{z}_{np,i}^{\top}|\mathcal{F}_{np,i-1}\right) - E\left(\boldsymbol{z}_{np,i}\boldsymbol{z}_{np,i}^{\top}\right)\right\} = \begin{pmatrix} \mathbf{0}_{qd\times qd} & (e_{lk}^{(12)})_{qd\times(K+1)} \\ (e_{kl}^{(21)})_{(K+1)\times qd} & (e_{k_1,k_2}^{(22)})_{(K+1)\times(K+1)} \end{pmatrix},$$

where

$$e_{l,k+1}^{(12)} = e_{k+1,l}^{(21)} = \sum_{i=1}^{np}\sum_{j=1}^{i-1} \frac{1}{n\sqrt{p}} \left(2\upsilon_{ij}^{(k)}\varepsilon_j\right)\gamma_{li}^{(1)}, \text{ for } k = 0, \cdots, K,\ l = 1, \cdots, qd, \tag{S5.17}$$

and

$$\begin{aligned}
e_{k_1+1+1,k_2+1}^{(22)} &= t_{k_1+1,k_2+1}^{(1)} + t_{k_2+1,k_1+1}^{(1)} + t_{k_1+1,k_2+1}^{(2)} + t_{k_1+1,k_2+1}^{(3)} + t_{k_1+1,k_2+1}^{(4)}, \\
t_{k_1+1,k_2+1}^{(1)} &= \sum_{i=1}^{np}\sum_{j=1}^{i-1} \frac{1}{n\sqrt{p}} \left(2\upsilon_{ij}^{(k_1)}\varepsilon_j\right)\gamma_{k_2+1,i}^{(2)}, \\
t_{k_1+1,k_2+1}^{(2)} &= \sum_{i=1}^{np}\left(\sum_{j_1=1}^{i-1}\sum_{j_2=1}^{i-1} \frac{4}{np}\upsilon_{ij_1}^{(k_1)}\upsilon_{ij_2}^{(k_2)}\varepsilon_{j_1}\varepsilon_{j_2} - \frac{4}{np}\sum_{j=1}^{i-1}\upsilon_{ij}^{(k_1)}\upsilon_{ij}^{(k_2)}\right), \\
t_{k_1+1,k_2+1}^{(3)} &= \sum_{i=1}^{np}\sum_{j=1}^{i-1} \frac{2}{np}\mu^{(3)}\upsilon_{ii}^{(k_1)}\upsilon_{ij}^{(k_2)}\varepsilon_j, \text{ and} \\
t_{k_1+1,k_2+1}^{(4)} &= \sum_{i=1}^{np}\sum_{j=1}^{i-1} \frac{2}{np}\mu^{(3)}\upsilon_{ii}^{(k_2)}\upsilon_{ij}^{(k_1)}\varepsilon_j, \text{ for } k_1, k_2 = 0, \cdots, K.
\end{aligned}$$

Based on (S5.10), we obtain for any $l = 1, \cdots, qd$,

$$\sup_{n\geq 1, p\geq 1}\left\{\frac{1}{n}\sum_{i=1}^{np}(\gamma_{li}^{(1)})^2\right\} = \sup_{n\geq 1, p\geq 1}\left\{\frac{1}{n}\left\|\boldsymbol{\Gamma}^{(1)\top}\boldsymbol{d}_l\right\|_2^2\right\} \leq \|\boldsymbol{d}_l\|_2^2 \sup_{n\geq 1, p\geq 1}\left\{\frac{1}{n}\left\|\boldsymbol{\Gamma}^{(1)}\right\|_2^2\right\} = O(1) < \infty, \tag{S5.18}$$

by recalling $\boldsymbol{d}_l$ is the l-th column of $\boldsymbol{I}_{qd}$. By using similar techniques as above, we can also show for any $k = 0, \cdots, K$,

$$\sup_{n\geq 1, p\geq 1}\left\{\frac{1}{n}\sum_{i=1}^{np}(\gamma_{k+1,i}^{(2)})^2\right\} = O(1) < \infty, \tag{S5.19}$$

under Conditions 2 – 5.

We next demonstrate $e_{k+1,l}^{(21)} = o_{\mathrm{P}}(1)$ for any $k = 0, \cdots, K$ and $l = 1, \cdots, qd$. From (S5.17), we have for any $k = 0, \cdots, K$ and $l = 1, \cdots, qd$,

$$e_{k+1,l}^{(21)} = \sum_{i=1}^{np}\sum_{j=1}^{i-1}\frac{1}{n\sqrt{p}}\left(2\upsilon_{ij}^{(k)}\varepsilon_j\right)\gamma_{li}^{(1)} = \sum_{j=1}^{np-1}\sum_{i=j+1}^{np}\frac{2}{n\sqrt{p}}\gamma_{li}^{(1)}\upsilon_{ij}^{(k)}\varepsilon_j \triangleq \sum_{j=1}^{np-1}h_{1,j},$$

where

$$h_{1,j} = \sum_{i=j+1}^{np}\frac{2}{n\sqrt{p}}\gamma_{li}^{(1)}\upsilon_{ij}^{(k)}\varepsilon_j, \text{ for } j = 1, \cdots, np-1.$$

Let $h_{1,np} = 0$. Then, we have $e_{k+1,l}^{(21)} = \sum_{j=1}^{np} h_{1,j}$ with $\{(h_{1,j}, \mathcal{F}_{np,j}) : 1 \leq j \leq np\}$ being a martingale difference array as $E(h_{1,j}|\mathcal{F}_{np,j-1}) = 0$. Let

$$c_{1,j} = \sum_{i=j+1}^{np}\frac{2}{n\sqrt{p}}\left|\gamma_{li}^{(1)}\right|\left|\upsilon_{ij}^{(k)}\right|,$$

for $j = 1, \cdots, np-1$ and $c_{1,np} = 2/(n\sqrt{p})$. Using Minkowski's inequality and Hölder's inequality, we obtain for $\delta \in (0, \eta/2]$,

$$\begin{aligned}
E\left|h_{1,j}\right|^{2+\delta} &= E\left|\sum_{i=j+1}^{np}\left(\frac{2}{n\sqrt{p}}\gamma_{li}^{(1)}\upsilon_{ij}^{(k)}\varepsilon_j\right)\right|^{2+\delta} \\
&\leq \left\{\sum_{i=j+1}^{np}\left(E\left|\frac{2}{n\sqrt{p}}\gamma_{li}^{(1)}\upsilon_{ij}^{(k)}\varepsilon_j\right|^{2+\delta}\right)^{\frac{1}{2+\delta}}\right\}^{2+\delta} \\
&= \left\{\sum_{i=j+1}^{np}\frac{2}{n\sqrt{p}}\left|\gamma_{li}^{(1)}\right|\left|\upsilon_{ij}^{(k)}\right|\left(E\left|\varepsilon_j\right|^{2+\delta}\right)^{\frac{1}{2+\delta}}\right\}^{2+\delta} \\
&= \left\{\left(E\left|\varepsilon_j\right|^{2+\delta}\right)^{\frac{1}{2+\delta}}\right\}^{2+\delta}\left(\sum_{i=j+1}^{np}\frac{2}{n\sqrt{p}}\left|\gamma_{li}^{(1)}\right|\left|\upsilon_{ij}^{(k)}\right|\right)^{2+\delta} \\
&\leq \left\{\left(E\left|\varepsilon_j\right|^{4+\eta}\right)^{\frac{1}{4+\eta}}\right\}^{2+\delta}\left(\sum_{i=j+1}^{np}\frac{2}{n\sqrt{p}}\left|\gamma_{li}^{(1)}\right|\left|\upsilon_{ij}^{(k)}\right|\right)^{2+\delta}. \qquad \text{(S5.20)}
\end{aligned}$$

Recalling $K_\varepsilon = \left(E\left|\varepsilon_i\right|^{4+\eta}\right)^{1/(4+\eta)}$, the above equation implies

$$E\left|\frac{h_{1,j}}{c_{1,j}}\right|^{2+\delta} \leq K_\varepsilon^{2+\delta},$$

which is bounded by Condition 1. Then, we show $\limsup_{np\to\infty}\sum_{j=1}^{np} c_{1,j} < \infty$ and $\lim_{np\to\infty}\sum_{j=1}^{np} c_{1,j}^2 = 0$. Specifically, by Jensen's inequality,

$$\begin{aligned}
\sum_{j=1}^{np} c_{1,j} &= \sum_{j=1}^{np-1}\sum_{i=j+1}^{np}\frac{2}{n\sqrt{p}}\left|\gamma_{li}^{(1)}\right|\left|\upsilon_{ij}^{(k)}\right| + \frac{2}{n\sqrt{p}} \\
&= \frac{2}{n\sqrt{p}}\sum_{i=1}^{np}\left|\gamma_{li}^{(1)}\right|\sum_{j=1}^{i-1}\left|\upsilon_{ij}^{(k)}\right| + \frac{2}{n\sqrt{p}}
\end{aligned}$$

$$
\begin{aligned}
&\leq \|\boldsymbol{\Upsilon}_k\|_1 \frac{2}{n\sqrt{p}} \sum_{i=1}^{np} \left|\gamma_{li}^{(1)}\right| + \frac{2}{n\sqrt{p}} \\
&\leq 2\|\boldsymbol{\Upsilon}_k\|_1 \sqrt{p} \sqrt{\frac{1}{np} \sum_{i=1}^{np} \left|\gamma_{li}^{(1)}\right|^2} + \frac{2}{n\sqrt{p}} \\
&\leq 2\|\boldsymbol{\Upsilon}_k\|_1 \sqrt{\frac{1}{n} \sum_{i=1}^{np} \left|\gamma_{li}^{(1)}\right|^2} + \frac{2}{n\sqrt{p}}, \qquad \text{(S5.21)}
\end{aligned}
$$

which is bounded by (S5.13) and (S5.18). This leads to $\limsup_{np\to\infty} \sum_{j=1}^{np} c_{1,j} < \infty$. By using Hölder's inequality, we obtain

$$
\begin{aligned}
\sum_{i=1}^{np} c_{1,j}^2 &= \frac{4}{n^2p} \sum_{j=1}^{np-1} \left| \sum_{i=j+1}^{np} \left|\gamma_{li}^{(1)}\right| \left|\upsilon_{ij}^{(k)}\right| \right|^2 + \frac{4}{n^2p} \\
&\leq \frac{4}{n^2p} \sum_{j=1}^{np-1} \left| \sum_{i=j+1}^{np} \left|\gamma_{li}^{(1)}\right|^{2+\delta} \right|^{\frac{2}{2+\delta}} \left| \sum_{i=j+1}^{np} \left|\upsilon_{ij}^{(k)}\right|^{\frac{2+\delta}{1+\delta}} \right|^{\frac{2(1+\delta)}{2+\delta}} + \frac{4}{n^2p} \\
&\leq \frac{4}{n^2p} \|\boldsymbol{\Upsilon}_k\|_1^2 \sum_{j=1}^{np-1} \left| \sum_{i=j+1}^{np} \left|\gamma_{li}^{(1)}\right|^{2+\delta} \right|^{\frac{2}{2+\delta}} + \frac{4}{n^2p} \\
&\leq \frac{4}{n} \|\boldsymbol{\Upsilon}_k\|_1^2 \left| \sum_{i=1}^{np} \left|\gamma_{li}^{(1)}\right|^{2+\delta} \right|^{\frac{2}{2+\delta}} + \frac{4}{n^2p} \\
&\leq 4 \|\boldsymbol{\Upsilon}_k\|_1^2 \left| \frac{1}{n^{\frac{2+\delta}{2}}} \sum_{i=1}^{np} \left|\gamma_{li}^{(1)}\right|^{2+\delta} \right|^{\frac{2}{2+\delta}} + \frac{4}{n^2p}. \qquad \text{(S5.22)}
\end{aligned}
$$

Based on similar techniques used in showing (S5.11), we can show $n^{-(2+\delta)/2} \sum_{i=1}^{np} |\gamma_{li}^{(1)}|^{2+\delta} = O(n^{-\delta/2})$, which leads to $\lim_{np\to\infty} n^{-(2+\delta)/2} \sum_{i=1}^{np} |\gamma_{li}^{(1)}|^{2+\delta} = 0$. This, together with $\|\boldsymbol{\Upsilon}_k\|_1^2$ being bounded by (S5.13) and $\lim_{np\to\infty} 4/(n^2p) = 0$, implies $\lim_{np\to\infty} \sum_{j=1}^{np} c_{1,j}^2 = 0$, and hence all the conditions in Lemma S9 are satisfied. Therefore, $e^{(21)}_{k+1,l} = \sum_{j=1}^{np} h_{1,j} = o_{\mathrm{P}}(1) \xrightarrow{\mathrm{P}} 0$ for any $k = 0, \cdots, K$ and $l = 1, \cdots, qd$.

We then take the following four steps (IVa) – (IVd) to show $e^{(22)}_{k_1+1,k_2+1} = o_{\mathrm{P}}(1)$ for any $k_1, k_2 = 0, \cdots, K$.

(IVa) We have

$$t^{(1)}_{k_1+1,k_2+1} = \sum_{i=1}^{np}\sum_{j=1}^{i-1} \frac{1}{n\sqrt{p}} \left(2\upsilon^{(k_1)}_{ij}\varepsilon_j\right)\gamma^{(2)}_{k_2+1,i} = \sum_{j=1}^{np-1}\sum_{i=j+1}^{np} \frac{2}{n\sqrt{p}}\gamma^{(2)}_{k_2+1,i}\upsilon^{(k_1)}_{ij}\varepsilon_j \triangleq \sum_{j=1}^{np-1} h_{2,j},$$

where

$$h_{2,j} = \sum_{i=j+1}^{np} \frac{2}{n\sqrt{p}}\gamma^{(2)}_{k_2+1,i}\upsilon^{(k_1)}_{ij}\varepsilon_j, \text{ for } j = 1, \cdots, np-1.$$

Let $h_{2,np} = 0$, then $t^{(1)}_{k_1+1,k_2+1} = \sum_{j=1}^{np} h_{2,j}$ with $\{(h_{2,j}, \mathcal{F}_{np,j}) : 1 \le j \le np\}$ being a martingale difference array as $E\left(h_{2,j}|\mathcal{F}_{np,j-1}\right) = 0$. Let $c_{2,j} = \sum_{i=j+1}^{np} 2 |\gamma^{(2)}_{k_2+1,i}||\upsilon^{(k_1)}_{ij}|/(n\sqrt{p})$ for $j = 1, \cdots, np-1$ and $c_{2,np} = 2/(n\sqrt{p})$. By similar techniques to those used in showing (S5.20), we obtain for $\delta \in (0, \eta/2]$,

$$E\left|h_{2,j}\right|^{2+\delta} \le \left\{\left(E\left|\varepsilon_j\right|^{4+\eta}\right)^{\frac{1}{4+\eta}}\right\}^{2+\delta} \left(\sum_{i=j+1}^{np} \frac{2}{n\sqrt{p}}\left|\gamma^{(2)}_{k_2+1,i}\right|\left|\upsilon^{(k_1)}_{ij}\right|\right)^{2+\delta}.$$

Hence,

$$E\left|\frac{h_{2,j}}{c_{2,j}}\right|^{2+\delta} \le \left\{\left(E\left|\varepsilon_j\right|^{4+\eta}\right)^{\frac{1}{4+\eta}}\right\}^{2+\delta} = K_\varepsilon^{2+\delta},$$

which is bounded by Condition 1. Then, we show $\limsup_{np\to\infty}\sum_{j=1}^{np} c_{2,j} < \infty$

and $\lim_{np\to\infty}\sum_{j=1}^{np} c_{2,j}^2 = 0$. By similar techniques to those used in showing (S5.21), we obtain

$$\sum_{j=1}^{np} c_{2,j} = \sum_{j=1}^{np-1}\sum_{i=j+1}^{np}\frac{2}{n\sqrt{p}}\left|\gamma_{k_2+1,i}^{(2)}\right|\left|\upsilon_{ij}^{(k_1)}\right| + \frac{2}{n\sqrt{p}} \le 2\left\|\boldsymbol{\Upsilon}_{k_1}\right\|_1\sqrt{\frac{1}{n}\sum_{i=1}^{np}\left|\gamma_{k_2+1,i}^{(2)}\right|^2} + \frac{2}{n\sqrt{p}},$$

which is bounded by (S5.13) and (S5.19). This leads to $\limsup_{np\to\infty}\sum_{j=1}^{np} c_{2,j} < \infty$. By similar techniques to those used in showing (S5.22), we obtain

$$\sum_{j=1}^{np} c_{2,j}^2 = \frac{4}{n^2p}\sum_{j=1}^{np-1}\left|\sum_{i=j+1}^{np}\left|\gamma_{k_2+1,i}^{(2)}\right|\left|\upsilon_{ij}^{(k_1)}\right|\right|^2 + \frac{4}{n^2p} \le 4\left\|\boldsymbol{\Upsilon}_{k_1}\right\|_1^2\left|\frac{1}{n^{\frac{2+\delta}{2}}}\sum_{i=1}^{np}\left|\gamma_{k_2+1,i}^{(2)}\right|^{2+\delta}\right|^{\frac{2}{2+\delta}} + \frac{4}{n^2p}.$$

By similar techniques to those used in showing (S5.11), we obtain

$$\frac{1}{n^{\frac{2+\delta}{2}}}\sum_{i=1}^{np}\left|\gamma_{k_2+1,i}^{(2)}\right|^{2+\delta} = O\left(\frac{1}{n^{\frac{\delta}{2}}}\right),$$

which leads to $\lim_{np\to\infty} n^{-(2+\delta)/2}\sum_{i=1}^{np}|\gamma_{k_2+1,i}^{(2)}|^{2+\delta} = 0$. This, together with $\|\boldsymbol{\Upsilon}_{k_1}\|_1^2$ being bounded by (S5.13) and $\lim_{np\to\infty} 4/(n^2p) = 0$, implies $\lim_{np\to\infty}\sum_{j=1}^{np} c_{2,j}^2$ $= 0$, and hence all the conditions in Lemma S9 are satisfied. Therefore, $t_{k_1+1,k_2+1}^{(1)} =$ $\sum_{j=1}^{np} h_{2,j} = o_{\mathrm{P}}(1) \xrightarrow{\mathrm{P}} 0$ for any $k_1, k_2 = 0, .., K$.

(IVb) After algebraic calculation, we have

$$\begin{aligned} t_{k_1+1,k_2+1}^{(2)} &= \sum_{i=1}^{np}\left(\sum_{j_1=1}^{i-1}\sum_{j_2=1}^{i-1}\frac{4}{np}\upsilon_{ij_1}^{(k_1)}\upsilon_{ij_2}^{(k_2)}\varepsilon_{j_1}\varepsilon_{j_2} - \sum_{j=1}^{i-1}\frac{4}{np}\upsilon_{ij}^{(k_1)}\upsilon_{ij}^{(k_2)}\right) \\ &= t_{k_1+1,k_2+1}^{(21)} + t_{k_1+1,k_2+1}^{(22)}, \end{aligned}$$

where

$$
\begin{aligned}
t^{(21)}_{k_1+1,k_2+1} &= \sum_{i=1}^{np}\left(\sum_{j=1}^{i-1}\frac{4}{np}\upsilon^{(k_1)}_{ij}\upsilon^{(k_2)}_{ij}\varepsilon_j^2 - \sum_{j=1}^{i-1}\frac{4}{np}\upsilon^{(k_1)}_{ij}\upsilon^{(k_2)}_{ij}\right)\\
&= \sum_{j=1}^{np-1}\sum_{i=j+1}^{np}\left(\frac{4}{np}\upsilon^{(k_1)}_{ij}\upsilon^{(k_2)}_{ij}\varepsilon_j^2 - \frac{4}{np}\upsilon^{(k_1)}_{ij}\upsilon^{(k_2)}_{ij}\right) \triangleq \sum_{j=1}^{np-1} h_{31,j},
\end{aligned}
$$

$$
h_{31,j} = \sum_{i=j+1}^{np}\left(\frac{4}{np}\upsilon^{(k_1)}_{ij}\upsilon^{(k_2)}_{ij}\varepsilon_j^2 - \frac{4}{np}\upsilon^{(k_1)}_{ij}\upsilon^{(k_2)}_{ij}\right),\ \text{for } j = 1,\cdots,np-1,
$$

$$
\begin{aligned}
t^{(22)}_{k_1+1,k_2+1} &= \sum_{i=1}^{np}\left(\sum_{j_1=1}^{i-1}\sum_{\substack{j_2=1\\ j_2\neq j_1}}^{i-1}\frac{4}{np}\upsilon^{(k_1)}_{ij_1}\upsilon^{(k_2)}_{ij_2}\varepsilon_{j_1}\varepsilon_{j_2}\right)\\
&= \sum_{j_1=1}^{np-1}\sum_{i=j_1+1}^{np}\sum_{\substack{j_2=1\\ j_2\neq j_1}}^{i-1}\frac{4}{np}\upsilon^{(k_1)}_{ij_1}\upsilon^{(k_2)}_{ij_2}\varepsilon_{j_1}\varepsilon_{j_2} \triangleq \sum_{j_1=1}^{np-1} h_{32,j_1}, \text{ and}
\end{aligned}
$$

$$
h_{32,j_1} = \sum_{i=j_1+1}^{np}\sum_{\substack{j_2=1\\ j_2\neq j_1}}^{i-1}\frac{4}{np}\upsilon^{(k_1)}_{ij_1}\upsilon^{(k_2)}_{ij_2}\varepsilon_{j_1}\varepsilon_{j_2},\ \text{for } j_1 = 1,\cdots,np-1.
$$

Let $h_{31,np} = 0$ and $h_{32,np} = 0$. Then, $t^{(21)}_{k_1+1,k_2+1} = \sum_{j=1}^{np} h_{31,j}$ with $\{(h_{31,j}, \mathcal{F}_{np,j}) : 1 \leq j \leq np\}$ being a martingale difference array and $t^{(22)}_{k_1+1,k_2+1} = \sum_{j=1}^{np} h_{32,j}$ with $\{(h_{32,j}, \mathcal{F}_{np,j}) : 1 \leq j \leq np\}$ being another martingale difference array, as $E\left(h_{31,j}|\mathcal{F}_{np,j-1}\right) = 0$ and $E\left(h_{32,j}|\mathcal{F}_{np,j-1}\right) = 0$. Let $c_{3,j} = 4/(np)$. Using Minkowski's inequality and Hölder's inequality, we obtain for $\delta \in (0, \eta/2]$,

$$
\begin{aligned}
E\left|\frac{h_{31,j}}{c_{3,j}}\right|^{2+\delta} &= E\left|\sum_{i=j+1}^{np} \upsilon_{ij}^{(k_1)}\upsilon_{ij}^{(k_2)}\varepsilon_j^2 - \upsilon_{ij}^{(k_1)}\upsilon_{ij}^{(k_2)}\right|^{2+\delta} \\
&\le \left\{\sum_{i=j+1}^{np}\left(E\left|\upsilon_{ij}^{(k_1)}\upsilon_{ij}^{(k_2)}\varepsilon_j^2 - \upsilon_{ij}^{(k_1)}\upsilon_{ij}^{(k_2)}\right|^{2+\delta}\right)^{\frac{1}{2+\delta}}\right\}^{2+\delta} \\
&\le \left\{\sum_{i=j+1}^{np}\left(E\left|\upsilon_{ij}^{(k_1)}\upsilon_{ij}^{(k_2)}\varepsilon_j^2\right|^{2+\delta}\right)^{\frac{1}{2+\delta}} + \sum_{i=j+1}^{np}\left(\left|\upsilon_{ij}^{(k_1)}\upsilon_{ij}^{(k_2)}\right|^{2+\delta}\right)^{\frac{1}{2+\delta}}\right\}^{2+\delta} \\
&\le \left(\sum_{i=j+1}^{np}\left|\upsilon_{ij}^{(k_1)}\upsilon_{ij}^{(k_2)}\right| K_\varepsilon^2 + \sum_{i=j+1}^{np}\left|\upsilon_{ij}^{(k_1)}\upsilon_{ij}^{(k_2)}\right|\right)^{2+\delta} \\
&= \left(K_\varepsilon^2+1\right)^{2+\delta}\left(\sum_{i=j+1}^{np}\left|\upsilon_{ij}^{(k_1)}\upsilon_{ij}^{(k_2)}\right|\right)^{2+\delta} \\
&\le \left(K_\varepsilon^2+1\right)^{2+\delta}\left(\sum_{i=1}^{np}\left|\upsilon_{ij}^{(k_1)}\upsilon_{ij}^{(k_2)}\right|\right)^{2+\delta} \\
&\le \left(K_\varepsilon^2+1\right)^{2+\delta}\left\{\left(\sum_{i=1}^{np}\left|\upsilon_{ij}^{(k_1)}\right|\right)\left(\sum_{i=1}^{np}\left|\upsilon_{ij}^{(k_2)}\right|\right)\right\}^{2+\delta} \\
&\le \left(K_\varepsilon^2+1\right)^{2+\delta}\left(\sup_{n\ge1,p\ge1}\|\{\boldsymbol{\Upsilon}_{k_1}\|_1\}\right)^{2+\delta}\left(\sup_{n\ge1,p\ge1}\{\|\boldsymbol{\Upsilon}_{k_2}\|_1\}\right)^{2+\delta},
\end{aligned}
$$

which is bounded by (S5.13) and Condition 1. Moreover,

$$
\limsup_{np\to\infty}\sum_{j=1}^{np} c_{3,j} = \limsup_{np\to\infty}\sum_{j=1}^{np}\frac{4}{np} = \limsup_{np\to\infty} 4 < \infty,
$$

and

$$
\lim_{np\to\infty}\sum_{j=1}^{np} c_{3,j}^2 = \lim_{np\to\infty}\sum_{j=1}^{np}\frac{16}{(np)^2} = \lim_{np\to\infty}\frac{16}{np} = 0.
$$

Thus, all conditions in Lemma S9 are satisfied, which implies $t^{(21)}_{k_1+1,k_2+1} = \sum_{j=1}^{np} h_{31,j} = o_{\mathrm{P}}(1) \xrightarrow{\mathrm{P}} 0$ for any $k_1, k_2 = 0, \cdots, K$. Furthermore, by using Minkowski's inequality and Hölder's inequality, we obtain for $\delta \in (0, \eta/2]$,

$$
\begin{aligned}
E\left|\frac{h_{32,j}}{c_{3,j}}\right|^{2+\delta} &= E\left|\sum_{i=j+1}^{np}\sum_{\substack{j_1=1\\ j_1\neq j}}^{i-1}\upsilon_{ij}^{(k_1)}\upsilon_{ij_1}^{(k_2)}\varepsilon_j\varepsilon_{j_1}\right|^{2+\delta} \\
&\le \left\{\sum_{i=j+1}^{np}\sum_{\substack{j_1=1\\ j_1\neq j}}^{i-1}\left|\upsilon_{ij}^{(k_1)}\right|\left|\upsilon_{ij_1}^{(k_2)}\right|\left(E\left|\varepsilon_j\varepsilon_{j_1}\right|^{2+\delta}\right)^{\frac{1}{2+\delta}}\right\}^{2+\delta} \\
&\le K_\varepsilon^{4+2\delta}\left(\sum_{i=j+1}^{np}\sum_{\substack{j_1=1\\ j_1\neq j}}^{i-1}\left|\upsilon_{ij}^{(k_1)}\right|\left|\upsilon_{ij_1}^{(k_2)}\right|\right)^{2+\delta} \\
&\le K_\varepsilon^{4+2\delta}\left\{\left(\sum_{i=1}^{np}\left|\upsilon_{ij}^{(k_1)}\right|\right)\left(\sum_{j_1=1}^{np}\left|\upsilon_{ij_1}^{(k_2)}\right|\right)\right\}^{2+\delta} \\
&\le K_\varepsilon^{4+2\delta}\left(\left\|\boldsymbol{\Upsilon}_{k_1}\right\|_1\left\|\boldsymbol{\Upsilon}_{k_2}\right\|_1\right)^{2+\delta} \\
&\le K_\varepsilon^{4+2\delta}\left(\sup_{n\ge1,p\ge1}\left\{\left\|\boldsymbol{\Upsilon}_{k_1}\right\|_1\right\}\right)^{2+\delta}\left(\sup_{n\ge1,p\ge1}\left\{\left\|\boldsymbol{\Upsilon}_{k_2}\right\|_1\right\}\right)^{2+\delta},
\end{aligned}
$$

which is bounded by (S5.13) and Condition 1. Moreover, we have shown $\limsup_{np\to\infty} \sum_{j=1}^{np} c_{3,j} < \infty$ and $\lim_{np\to\infty}\sum_{j=1}^{np} c_{3,j}^2 = 0$. This implies all conditions in Lemma S9 are satisfied, and thus $t^{(22)}_{k_1+1,k_2+1} = \sum_{j=1}^{np} h_{32,j} = o_{\mathrm{P}}(1) \xrightarrow{\mathrm{P}} 0$ for any $k_1, k_2 = 0, \cdots, K$. In sum, $t^{(2)}_{k_1+1,k_2+1} = t^{(21)}_{k_1+1,k_2+1} + t^{(22)}_{k_1+1,k_2+1} = o_P(1) \xrightarrow{P} 0$ for any $k_1, k_2 = 0, \cdots, K$.

(IVc) We have

$$t^{(3)}_{k_1+1,k_2+1} = \sum_{i=1}^{np}\sum_{j=1}^{i-1}\frac{2}{np}\mu^{(3)}\upsilon_{ii}^{(k_1)}\upsilon_{ij}^{(k_2)}\varepsilon_j = \sum_{j=1}^{np-1}\sum_{i=j+1}^{np}\frac{2}{np}\mu^{(3)}\upsilon_{ii}^{(k_1)}\upsilon_{ij}^{(k_2)}\varepsilon_j = \sum_{j=1}^{np-1} h_{4,j},$$

where

$$h_{4,j} = \sum_{i=j+1}^{np}\frac{2}{np}\mu^{(3)}\upsilon_{ii}^{(k_1)}\upsilon_{ij}^{(k_2)}\varepsilon_j, \text{ for } j = 1,\cdots,np-1.$$

Let $h_{4,np} = 0$. Then, $t^{(3)}_{k_1+1,k_2+1} = \sum_{j=1}^{np} h_{4,j}$ with $\{(h_{4,j}, \mathcal{F}_{np,j}) : 1 \leq j \leq np\}$ being a martingale difference array, as $E\left(h_{4,j}|\mathcal{F}_{np,j-1}\right) = 0$. Let $c_{4,j} = 2/(np)$. By using Minkowski's inequality, we obtain

$$\begin{aligned}
E\left|\frac{h_{4,j}}{c_{4,j}}\right|^{4+\eta} &\leq \left|\mu^{(3)}\right|^{4+\eta}\left\{\sum_{i=j+1}^{np}\left|\upsilon_{ii}^{(k_1)}\right|\left|\upsilon_{ij}^{(k_2)}\right|\left(E\left|\varepsilon_j\right|^{4+\eta}\right)^{\frac{1}{4+\eta}}\right\}^{4+\eta} \\
&\leq \left|\mu^{(3)}\right|^{4+\eta} K_\varepsilon^{4+\eta}\left(\|\boldsymbol{\Upsilon}_{k_1}\|_1\sum_{i=1}^{np}\left|\upsilon_{ij}^{(k_2)}\right|\right)^{4+\eta} \\
&\leq \left|\mu^{(3)}\right|^{4+\eta} K_\varepsilon^{4+\eta}\left(\|\boldsymbol{\Upsilon}_{k_1}\|_1\|\boldsymbol{\Upsilon}_{k_2}\|_1\right)^{4+\eta} \\
&\leq \left|\mu^{(3)}\right|^{4+\eta} K_\varepsilon^{4+\eta}\left(\sup_{n\geq 1,p\geq 1}\{\|\boldsymbol{\Upsilon}_{k_1}\|_1\}\right)^{4+\eta}\left(\sup_{n\geq 1,p\geq 1}\{\|\boldsymbol{\Upsilon}_{k_2}\|_1\}\right)^{4+\eta},
\end{aligned} \tag{S5.23}$$

which is bounded by (S5.13) and Condition 1. Moreover,

$$\limsup_{np\to\infty}\sum_{j=1}^{np} c_{4,j} = \limsup_{np\to\infty}\sum_{j=1}^{np}\frac{2}{np} = \limsup_{np\to\infty} 2 < \infty,$$

and

$$\lim_{np\to\infty}\sum_{j=1}^{np}c_{4,j}^2=\lim_{np\to\infty}\sum_{j=1}^{np}\frac{4}{(np)^2}=\lim_{np\to\infty}\frac{4}{np}=0.$$

Hence, all the conditions in Lemma S9 are satisfied, which leads to $t^{(3)}_{k_1+1,k_2+1}=\sum_{j=1}^{np}h_{4,j}=o_{\mathrm{P}}(1)\xrightarrow{\mathrm{P}}0$ for any $k_1,k_2=0,\cdots,K$.

(IVd) We have

$$\begin{aligned}t^{(4)}_{k_1+1,k_2+1}&=\sum_{i=1}^{np}\sum_{j=1}^{i-1}\frac{2}{np}\mu^{(3)}\upsilon_{ii}^{(k_2)}\upsilon_{ij}^{(k_1)}\varepsilon_j\\&=\sum_{j=1}^{np-1}\sum_{i=j+1}^{np}\frac{2}{np}\mu^{(3)}\upsilon_{ii}^{(k_2)}\upsilon_{ij}^{(k_1)}\varepsilon_j\\&=\sum_{j=1}^{np-1}h_{5,j},\end{aligned}$$

where

$$h_{5,j}=\sum_{i=j+1}^{np}\frac{2}{np}\mu^{(3)}\upsilon_{ii}^{(k_2)}\upsilon_{ij}^{(k_1)}\varepsilon_j,\text{ for } j=1,\cdots,np-1.$$

Let $h_{5,np}=0$. Then, $t^{(4)}_{k_1+1,k_2+1}=\sum_{j=1}^{np}h_{5,j}$ with $\{(h_{5,j},\mathcal{F}_{np,j}):1\le j\le np\}$ being a martingale difference array, as $E\left(h_{5,j}|\mathcal{F}_{np,j-1}\right)=0$. By similar techniques to those used in showing (S5.23), we obtain

$$E\left|\frac{h_{5,j}}{c_{4,j}}\right|^{4+\eta}\le\left|\mu^{(3)}\right|^{4+\eta}K_\varepsilon^{4+\eta}\left(\sup_{n\ge1,p\ge1}\{\|\boldsymbol{\Upsilon}_{k_2}\|_1\}\right)^{4+\eta}\left(\sup_{n\ge1,p\ge1}\{\|\boldsymbol{\Upsilon}_{k_1}\|_1\}\right)^{4+\eta},$$

which is bounded by (S5.13) and Condition 1. We have also shown $\limsup_{np\to\infty}$

$\sum_{j=1}^{np} c_{4,j} < \infty$ and $\lim_{np\to\infty} \sum_{j=1}^{np} c_{4,j}^2 = 0$. Hence, all the conditions in Lemma S9 are satisfied, which implies $t^{(4)}_{k_1+1,k_2+1} = \sum_{j=1}^{np} h_{5,j} = o_{\mathrm{P}}(1) \xrightarrow{\mathrm{P}} 0$ for any $k_1, k_2 = 0, \cdots, K$.

From (IVa) – (IVd), we obtain $t^{(a)}_{k_1+1,k_2+1} = o_{\mathrm{P}}(1)$ for any $a = 1, \cdots, 4$ and $k_1, k_2 = 0, \cdots, K$. This implies $e^{(22)}_{k_1+1,k_2+1} = t^{(1)}_{k_1+1,k_2+1} + t^{(1)}_{k_2+1,k_1+1} + t^{(2)}_{k_1+1,k_2+1} + t^{(3)}_{k_1+1,k_2+1} + t^{(4)}_{k_1+1,k_2+1} = o_{\mathrm{P}}(1)$ for any $k_1, k_2 = 0, \cdots, K$. In summary, we obtain

$$\sum_{i=1}^{np} \left\{ E\left(\boldsymbol{z}_{np,i}\boldsymbol{z}_{np,i}^\top | \mathcal{F}_{np,i-1}\right) - E\left(\boldsymbol{z}_{np,i}\boldsymbol{z}_{np,i}^\top\right) \right\} = o_{\mathrm{P}}(1).$$

This, together with (S5.8), (S5.15) and (S5.16), implies

$$\begin{aligned}
&\sum_{i=1}^{np} E\left[\left\{-\boldsymbol{\Omega}^{-1/2}(\boldsymbol{\Xi}^{(\mathcal{S})})\boldsymbol{z}_{np,i}\right\}\left\{-\boldsymbol{\Omega}^{-1/2}(\boldsymbol{\Xi}^{(\mathcal{S})})\boldsymbol{z}_{np,i}\right\}^\top | \mathcal{F}_{np,i-1}\right] \\
&= \mathrm{Cov}\left\{\sum_{i=1}^{np} -\boldsymbol{\Omega}^{-1/2}(\boldsymbol{\Xi}^{(\mathcal{S})})\boldsymbol{z}_{np,i}\right\} + o_{\mathrm{P}}(1) \\
&= \mathrm{Cov}\left\{-\boldsymbol{\Omega}^{-1/2}(\boldsymbol{\Xi}^{(\mathcal{S})})\boldsymbol{\Xi}^{(\mathcal{S})}\boldsymbol{B}^{-1}\tilde{\boldsymbol{Z}}^{-1}\boldsymbol{\psi}(\boldsymbol{\vartheta}^{(0)})\right\} + o_{\mathrm{P}}(1) \\
&\xrightarrow{\mathrm{P}} \boldsymbol{I}_{qd+K+1}.
\end{aligned}$$

Finally, by combining the results from Step I to Step IV, we can apply Lemma S4 for the martingale difference array $\{(-\boldsymbol{\Omega}^{-1/2}(\boldsymbol{\Xi}^{(\mathcal{S})})\boldsymbol{z}_{np,i}, \mathcal{F}_{np,i}) : 1 \leq$

$i \leq np\}$ and obtain

$$-\boldsymbol{\Omega}^{-1/2}(\boldsymbol{\Xi}^{(\mathcal{S})})\boldsymbol{\Xi}^{(\mathcal{S})}\boldsymbol{B}^{-1}\tilde{\boldsymbol{Z}}^{-1}\boldsymbol{\psi}(\boldsymbol{\vartheta}^{(0)}) = \sum_{i=1}^{np} -\boldsymbol{\Omega}^{-1/2}(\boldsymbol{\Xi}^{(\mathcal{S})})\boldsymbol{z}_{np,i} \xrightarrow{d} N\left(\boldsymbol{0}_{qd+K+1}, \boldsymbol{I}_{qd+K+1}\right).$$

This completes the proof of Lemma S2.

□

Proof of Lemma S3. By Condition 5, we have $\|\boldsymbol{\Omega}^{-1/2}(\boldsymbol{\Xi}^{(\mathcal{S})})\|_2 = O(1)$ and $\|\boldsymbol{B}^{-1}\|_2 = O(1)$. Furthermore, we obtain from (S5.6) that the elements of $n^{-1/2}\boldsymbol{\psi}_{\boldsymbol{\beta}}(\boldsymbol{\vartheta}^{(0)})$ and $(np)^{-1/2}\boldsymbol{\psi}_{\boldsymbol{\alpha}}(\boldsymbol{\vartheta}^{(0)})$ are of order $O_{\mathrm{P}}(1)$ uniformly. This implies

$$\left\|\tilde{\boldsymbol{Z}}^{-1}\boldsymbol{\psi}(\boldsymbol{\vartheta}^{(0)})\right\|_2 = \sqrt{O_{\mathrm{P}}(p)} = O_{\mathrm{P}}\left(\sqrt{p}\right).$$

Next, we study the properties of $\boldsymbol{V}$ by investigating the properties of the elements for different blocks of $\partial\boldsymbol{\psi}(\boldsymbol{\vartheta}^{(0)})/\partial\boldsymbol{\vartheta}^\top - E\{\partial\boldsymbol{\psi}(\boldsymbol{\vartheta}^{(0)})/\partial\boldsymbol{\vartheta}^\top\}$ via the following four steps, (I) – (IV).

(I) Let $\boldsymbol{F}_{jl} = \{\partial\boldsymbol{D}^\top(\boldsymbol{\beta}^{(0)})\boldsymbol{A}^{-\frac{1}{2}}(\boldsymbol{\beta}^{(0)})\}/\partial\beta_{jl} \in \mathbb{R}^{pd\times np}$ with the $((j-1)d + l_1, (i-1)p+j)$-th element being

$$\left[\frac{g^{-1''}(\boldsymbol{x}_i^\top\boldsymbol{\beta}_j^{(0)})}{\sqrt{h\{g^{-1}(\boldsymbol{x}_i^\top\boldsymbol{\beta}_j^{(0)});\phi_j\}}} - \frac{h'\{g^{-1}(\boldsymbol{x}_i^\top\boldsymbol{\beta}_j^{(0)});\phi_j\}[g^{-1'}(\boldsymbol{x}_i^\top\boldsymbol{\beta}_j^{(0)})]^2}{2\sqrt{h\{g^{-1}(\boldsymbol{x}_i^\top\boldsymbol{\beta}_j^{(0)});\phi_j\}}h\{g^{-1}(\boldsymbol{x}_i^\top\boldsymbol{\beta}_j^{(0)});\phi_j\}}\right]x_{il}x_{il_1},$$

for $i = 1, ..., n$, $l_1 = 1, ..., d$ and other elements being zeros. After algebraic simplification, we obtain for $j = 1, ..., p$ and $l = 1, ..., d$,

$$\frac{\partial \boldsymbol{\psi}_{\boldsymbol{\beta}}(\boldsymbol{\vartheta}^{(0)})}{\partial \beta_{jl}} - E\left\{\frac{\partial \boldsymbol{\psi}_{\boldsymbol{\beta}}(\boldsymbol{\vartheta}^{(0)})}{\partial \beta_{jl}}\right\} = \left\{\boldsymbol{F}_{jl}\tilde{\boldsymbol{L}}_0^{-1\top} + \boldsymbol{D}^\top(\boldsymbol{\beta}^{(0)})\boldsymbol{A}^{-1/2}(\boldsymbol{\beta}^{(0)})\tilde{\boldsymbol{\Sigma}}_0^{-1}\boldsymbol{M}_{jl}\tilde{\boldsymbol{L}}_0\right\}\boldsymbol{\varepsilon},$$

where $\boldsymbol{M}_{jl}$ is defined above Lemma S1. Then, the covariance matrix for the random vector is given as

$$\begin{aligned}&\mathrm{Cov}\left[\frac{\partial \boldsymbol{\psi}_{\boldsymbol{\beta}}(\boldsymbol{\vartheta}^{(0)})}{\partial \beta_{jl}} - E\left\{\frac{\partial \boldsymbol{\psi}_{\boldsymbol{\beta}}(\boldsymbol{\vartheta}^{(0)})}{\partial \beta_{jl}}\right\}\right]\\ &= \left\{\boldsymbol{F}_{jl}\tilde{\boldsymbol{L}}_0^{-1\top} + \boldsymbol{D}^\top(\boldsymbol{\beta}^{(0)})\boldsymbol{A}^{-1/2}(\boldsymbol{\beta}^{(0)})\tilde{\boldsymbol{\Sigma}}_0^{-1}\boldsymbol{M}_{jl}\tilde{\boldsymbol{L}}_0\right\}\left\{\boldsymbol{F}_{jl}\tilde{\boldsymbol{L}}_0^{-1\top} + \boldsymbol{D}^\top(\boldsymbol{\beta}^{(0)})\boldsymbol{A}^{-1/2}(\boldsymbol{\beta}^{(0)})\tilde{\boldsymbol{\Sigma}}_0^{-1}\boldsymbol{M}_{jl}\tilde{\boldsymbol{L}}_0\right\}^\top.\end{aligned}$$

By Conditions 3 – 4 and similar techniques to those used in showing (S5.1), we can show there exists finite positive constants C_h' and C_g' such that $|h'\{g^{-1}(\boldsymbol{x}_i^\top\boldsymbol{\beta}_j^{(0)}); \phi_j\}| \leq C_h'$ and $|g^{-1\prime\prime}(\boldsymbol{x}_i^\top\boldsymbol{\beta}_j^{(0)})| \leq C_g'$ for any $i = 1, \cdots, n$ and $j = 1, \cdots, p$. This, together with inequality 4.67(a) in Seber (2008) and Conditions 2 – 4, implies

$$\begin{aligned}&\|\boldsymbol{F}_{jl}\|_2\\ &\leq \|\boldsymbol{F}_{jl}\|_F\\ &= \sqrt{\sum_{i=1}^n\sum_{l_1=1}^d\left|\frac{g^{-1\prime\prime}(\boldsymbol{x}_i^\top\boldsymbol{\beta}_j^{(0)})}{\sqrt{h\{g^{-1}(\boldsymbol{x}_i^\top\boldsymbol{\beta}_j^{(0)}); \phi_j\}}} - \frac{h'\{g^{-1}(\boldsymbol{x}_i^\top\boldsymbol{\beta}_j^{(0)}); \phi_j\}[g^{-1\prime}(\boldsymbol{x}_i^\top\boldsymbol{\beta}_j^{(0)})]^2}{2\sqrt{h\{g^{-1}(\boldsymbol{x}_i^\top\boldsymbol{\beta}_j^{(0)}); \phi_j\}}h\{g^{-1}(\boldsymbol{x}_i^\top\boldsymbol{\beta}_j^{(0)}); \phi_j\}}\right|^2|x_{il}x_{il_1}|^2}\\ &\leq \sqrt{\sum_{i=1}^n\sum_{l_1=1}^d\left(\frac{C_g'}{\sqrt{C_h}} + \frac{C_h'C_g^2}{2\sqrt{C_h}C_h}\right)^2 C_X^4}\\ &= \sqrt{nd\left(\frac{C_g'}{\sqrt{C_h}} + \frac{C_h'C_g^2}{2\sqrt{C_h}C_h}\right)^2 C_X^4}\\ &= \sqrt{n}\sqrt{d}\left(\frac{C_g'}{\sqrt{C_h}} + \frac{C_h'C_g^2}{2\sqrt{C_h}C_h}\right)C_X^2\end{aligned}$$

$$= O\left(\sqrt{n}\right), \text{ for } j = 1, \cdots, p \text{ and } l = 1, \cdots, d. \tag{S5.24}$$

By Conditions 2 – 4 and using similar methods, we can show $\|\tilde{\boldsymbol{L}}_0^\top\|_2 = O(1)$ and $\|\boldsymbol{M}_{jl}\|_2 = O(1)$ for $j = 1, \cdots, p$ and $l = 1, \cdots, d$. Combining this with (S5.3) and (S5.4), we obtain

$$\left\|\text{Cov}\left[\frac{\partial \boldsymbol{\psi}_{\boldsymbol{\beta}}(\boldsymbol{\vartheta}^{(0)})}{\partial \beta_{jl}} - E\left\{\frac{\partial \boldsymbol{\psi}_{\boldsymbol{\beta}}(\boldsymbol{\vartheta}^{(0)})}{\partial \beta_{jl}}\right\}\right]\right\|_2 = \left\{O\left(\sqrt{n}\right)O\left(1\right) + O\left(\sqrt{n}\right)O\left(1\right)O\left(1\right)O\left(1\right)O\left(1\right)\right\}^2$$
$$= O\left(n\right), \text{ for } j = 1, \cdots, p \text{ and } l = 1, \cdots, d.$$

By denoting $\partial \boldsymbol{\psi}_{\boldsymbol{\beta}}(\boldsymbol{\vartheta}^{(0)})/\partial \boldsymbol{\beta}^\top - E\{\partial \boldsymbol{\psi}_{\boldsymbol{\beta}}(\boldsymbol{\vartheta}^{(0)})/\partial \boldsymbol{\beta}^\top\} = (v_{l_1 l_2}^{(\beta)})_{pd\times pd}$, the above implies

$$v_{l_1 l_2}^{(\beta)} = O_{\mathrm{P}}\left(\sqrt{n}\right),$$

uniformly in $l_1, l_2 = 1, \cdots, pd$, which leads to

$$\left|\frac{\partial \boldsymbol{\psi}_{\boldsymbol{\beta}}(\boldsymbol{\vartheta}^{(0)})}{\partial \boldsymbol{\beta}} - E\left\{\frac{\partial \boldsymbol{\psi}_{\boldsymbol{\beta}}(\boldsymbol{\vartheta}^{(0)})}{\partial \boldsymbol{\beta}}\right\}\right|_\infty = O_{\mathrm{P}}\left(\sqrt{n}\right).$$

(II) For $k = 0, \cdots, K$, we have

$$\frac{\partial \boldsymbol{\psi}_{\boldsymbol{\beta}}(\boldsymbol{\vartheta}^{(0)})}{\partial \alpha_k} - E\left\{\frac{\partial \boldsymbol{\psi}_{\boldsymbol{\beta}}(\boldsymbol{\vartheta}^{(0)})}{\partial \alpha_k}\right\} = -\boldsymbol{D}^\top(\boldsymbol{\beta}^{(0)})\boldsymbol{A}^{-1/2}(\boldsymbol{\beta}^{(0)})\tilde{\boldsymbol{\Sigma}}_0^{-1}\tilde{\boldsymbol{W}}_k\tilde{\boldsymbol{L}}_0^{-1\top}\boldsymbol{\varepsilon},$$

and

$$\operatorname{Cov}\left[\frac{\partial \boldsymbol{\psi}_{\boldsymbol{\beta}}(\boldsymbol{\vartheta}^{(0)})}{\partial \alpha_k} - E\left\{\frac{\partial \boldsymbol{\psi}_{\boldsymbol{\beta}}(\boldsymbol{\vartheta}^{(0)})}{\partial \alpha_k}\right\}\right]$$
$$= \boldsymbol{D}^\top(\boldsymbol{\beta}^{(0)})\boldsymbol{A}^{-1/2}(\boldsymbol{\beta}^{(0)})\tilde{\boldsymbol{\Sigma}}_0^{-1}\tilde{\boldsymbol{W}}_k\tilde{\boldsymbol{\Sigma}}_0^{-1}\tilde{\boldsymbol{W}}_k\tilde{\boldsymbol{\Sigma}}_0^{-1}\boldsymbol{A}^{-1/2}(\boldsymbol{\beta}^{(0)})\boldsymbol{D}(\boldsymbol{\beta}^{(0)}).$$

By Conditions 2 – 4, (S5.3) – (S5.4), we have that

$$\left\|\operatorname{Cov}\left[\frac{\partial \boldsymbol{\psi}_{\boldsymbol{\beta}}(\boldsymbol{\vartheta}^{(0)})}{\partial \alpha_k} - E\left\{\frac{\partial \boldsymbol{\psi}_{\boldsymbol{\beta}}(\boldsymbol{\vartheta}^{(0)})}{\partial \alpha_k}\right\}\right]\right\|_2 = O\left(n\right), \text{ for } k = 0, \cdots, K.$$

By denoting $\partial \boldsymbol{\psi}_{\boldsymbol{\beta}}(\boldsymbol{\vartheta}^{(0)})/\partial \boldsymbol{\alpha}^\top = (v_{lk}^{(\beta\alpha)})_{pd\times(K+1)}$, the above implies

$$v_{l,k+1}^{(\beta\alpha)} = O_{\mathrm{P}}\left(\sqrt{n}\right),$$

uniformly in $l = 1, \cdots, pd$ and $k = 0, \cdots, K$, which leads to

$$\left|\frac{\partial \boldsymbol{\psi}_{\boldsymbol{\beta}}(\boldsymbol{\vartheta}^{(0)})}{\partial \boldsymbol{\alpha}} - E\left\{\frac{\partial \boldsymbol{\psi}_{\boldsymbol{\beta}}(\boldsymbol{\vartheta}^{(0)})}{\partial \boldsymbol{\alpha}}\right\}\right|_\infty = O_{\mathrm{P}}\left(\sqrt{n}\right).$$

(III) For $j = 1, \cdots, p$ and $l = 1, \cdots, d$, we have that

$$\frac{\partial \boldsymbol{\psi}_{\boldsymbol{\alpha}}(\boldsymbol{\vartheta}^{(0)})}{\partial \beta_{jl}} - E\left\{\frac{\partial \boldsymbol{\psi}_{\boldsymbol{\alpha}}(\boldsymbol{\vartheta}^{(0)})}{\partial \beta_{jl}}\right\}$$

$$= 2\left\{\begin{pmatrix} \boldsymbol{\varepsilon}^\top \left[\tilde{\boldsymbol{L}}_0^\top \tilde{\boldsymbol{W}}_0 \boldsymbol{M}_{jl} \tilde{\boldsymbol{L}}_0\right]_{\mathfrak{S}} \boldsymbol{\varepsilon} \\ \vdots \\ \boldsymbol{\varepsilon}^\top \left[\tilde{\boldsymbol{L}}_0^\top \tilde{\boldsymbol{W}}_K \boldsymbol{M}_{jl} \tilde{\boldsymbol{L}}_0\right]_{\mathfrak{S}} \boldsymbol{\varepsilon} \end{pmatrix} - \begin{pmatrix} \mathrm{tr}\left\{\left[\tilde{\boldsymbol{L}}_0^\top \tilde{\boldsymbol{W}}_0 \boldsymbol{M}_{jl} \tilde{\boldsymbol{L}}_0\right]_{\mathfrak{S}}\right\} \\ \vdots \\ \mathrm{tr}\left\{\left[\tilde{\boldsymbol{L}}_0^\top \tilde{\boldsymbol{W}}_K \boldsymbol{M}_{jl} \tilde{\boldsymbol{L}}_0\right]_{\mathfrak{S}}\right\} \end{pmatrix}\right\}$$

$$- 2\begin{pmatrix} \boldsymbol{D}^\top_{\cdot((j-1)d+l)}(\boldsymbol{\beta}^{(0)})\boldsymbol{A}^{-1/2}(\boldsymbol{\beta}^{(0)})\tilde{\boldsymbol{W}}_0\tilde{\boldsymbol{L}}_0 \\ \vdots \\ \boldsymbol{D}^\top_{\cdot((j-1)d+l)}(\boldsymbol{\beta}^{(0)})\boldsymbol{A}^{-1/2}(\boldsymbol{\beta}^{(0)})\tilde{\boldsymbol{W}}_K\tilde{\boldsymbol{L}}_0 \end{pmatrix}\boldsymbol{\varepsilon},$$

where $[\boldsymbol{H}]_{\mathfrak{S}} = (\boldsymbol{H} + \boldsymbol{H}^\top)/2$ is the symmetrize operator for a generic square matrix $\boldsymbol{H}$ and $\boldsymbol{D}^\top_{\cdot((j-1)d+l)}(\boldsymbol{\beta}^{(0)})$ is a row vector that consists of the $((j-1)d+l)$-th column of $\boldsymbol{D}(\boldsymbol{\beta}^{(0)})$ matrix. More specifically, it is an np-dimensional row vector whose $((i-1)p+j)$-th element is given as $g^{-1'}(\boldsymbol{x}_i^\top \boldsymbol{\beta}_j^{(0)})x_{il}$ for $i = 1, \cdots, n$ and other elements being zeros. Let

$$\boldsymbol{\Gamma}^{(jl)} = -\left(\tilde{\boldsymbol{L}}_0^\top \tilde{\boldsymbol{W}}_0 \boldsymbol{A}^{-\frac{1}{2}}(\boldsymbol{\beta}^{(0)})\boldsymbol{D}_{\cdot((j-1)d+l)}(\boldsymbol{\beta}^{(0)}), \cdots, \tilde{\boldsymbol{L}}_0^\top \tilde{\boldsymbol{W}}_K \boldsymbol{A}^{-\frac{1}{2}}(\boldsymbol{\beta}^{(0)})\boldsymbol{D}_{\cdot((j-1)d+l)}(\boldsymbol{\beta}^{(0)})\right),$$

which is an $np \times (K+1)$ matrix, and

$$\boldsymbol{\Psi}^{(jl)} = \begin{pmatrix} \upsilon_{11}^{(jl)(0)} & \cdots & \upsilon_{11}^{(jl)(K)} \\ \vdots & \cdots & \vdots \\ \upsilon_{np,np}^{(jl)(0)} & \cdots & \upsilon_{np,np}^{(jl)(K)} \end{pmatrix},$$

where $\upsilon_{ii}^{(jl)(k)}$ denotes the (i,i)-th element of $\boldsymbol{\Upsilon}_k^{(jl)} = [\tilde{\boldsymbol{L}}_0^\top \tilde{\boldsymbol{W}}_k \boldsymbol{M}_{jl} \tilde{\boldsymbol{L}}_0]_{\mathfrak{S}}$ for $i = 1, \cdots, np$, $k = 0, \cdots, K$, $j = 1, \cdots, p$ and $l = 1, \cdots, d$. Then, we employ

Lemma S7 to obtain the covariance matrix as

$$\begin{aligned}
&\mathrm{Cov}\left[\frac{\partial \boldsymbol{\psi}_{\boldsymbol{\alpha}}(\boldsymbol{\vartheta}^{(0)})}{\partial \beta_{jl}} - E\left\{\frac{\partial \boldsymbol{\psi}_{\boldsymbol{\alpha}}(\boldsymbol{\vartheta}^{(0)})}{\partial \boldsymbol{\beta}_{jl}}\right\}\right] \\
&= 8\left(\mathrm{tr}(\boldsymbol{\Upsilon}_{k_1}^{(jl)}\boldsymbol{\Upsilon}_{k_2}^{(jl)})\right)_{(K+1)\times(K+1)} + 4\boldsymbol{\Gamma}^{(jl)\top}\boldsymbol{\Gamma}^{(jl)} \\
&\quad + 4\left(\mu^{(4)}-3\right)\boldsymbol{\Psi}^{(jl)\top}\boldsymbol{\Psi}^{(jl)} + 4\mu^{(3)}\left(\boldsymbol{\Psi}^{(jl)\top}\boldsymbol{\Gamma}^{(jl)} + \boldsymbol{\Gamma}^{(jl)\top}\boldsymbol{\Psi}^{(jl)}\right).
\end{aligned}$$

By Conditions 2 – 4 and similar techniques to those used in showing (S5.24), we obtain

$$\left\|\left(\mathrm{tr}(\boldsymbol{\Upsilon}_{k_1}^{(jl)}\boldsymbol{\Upsilon}_{k_2}^{(jl)})\right)_{(K+1)\times(K+1)}\right\|_2 = O(np),\ \left\|\boldsymbol{\Gamma}^{(jl)}\right\|_2 = O\left(\sqrt{n}\right),\ \text{and}$$
$$\left\|\boldsymbol{\Psi}^{(jl)}\right\|_2 = O\left(\sqrt{np}\right),\ \text{for } j=1,\cdots,p \text{ and } l=1,\cdots,d.$$

This, together with Condition 1, implies

$$\left\|\mathrm{Cov}\left[\frac{\partial \boldsymbol{\psi}_{\boldsymbol{\alpha}}(\boldsymbol{\vartheta}^{(0)})}{\partial \beta_{jl}} - E\left\{\frac{\partial \boldsymbol{\psi}_{\boldsymbol{\alpha}}(\boldsymbol{\vartheta}^{(0)})}{\partial \beta_{jl}}\right\}\right]\right\|_2 = O(np),\ \text{for } j=1,\cdots,p \text{ and } l=1,\cdots,d.$$

By denoting $\partial \boldsymbol{\psi}_{\boldsymbol{\alpha}}(\boldsymbol{\vartheta}^{(0)})/\partial \boldsymbol{\beta} - E\{\partial \boldsymbol{\psi}_{\boldsymbol{\alpha}}(\boldsymbol{\vartheta}^{(0)})/\partial \boldsymbol{\beta}\} = (v_{kl}^{(\alpha\beta)})_{(K+1)\times pd}$, the above implies

$$v_{k+1,l}^{(\alpha\beta)} = O_{\mathrm{P}}\left(\sqrt{np}\right),$$

uniformly in $k = 0, \cdots, K$ and $l = 1, \cdots, pd$, which leads to

$$\left| \frac{\partial \boldsymbol{\psi}_{\boldsymbol{\alpha}}(\boldsymbol{\vartheta}^{(0)})}{\partial \boldsymbol{\beta}} - E\left\{ \frac{\partial \boldsymbol{\psi}_{\boldsymbol{\alpha}}(\boldsymbol{\vartheta}^{(0)})}{\partial \boldsymbol{\beta}} \right\} \right|_{\infty} = O_{\mathrm{P}}\left(\sqrt{np}\right).$$

(IV) We have that $\partial \boldsymbol{\psi}_{\boldsymbol{\alpha}}(\boldsymbol{\vartheta}^{(0)})/\partial \boldsymbol{\alpha} - E\{\partial \boldsymbol{\psi}_{\boldsymbol{\alpha}}(\boldsymbol{\vartheta}^{(0)})/\partial \boldsymbol{\alpha}\} = \mathbf{0}_{(K+1)\times(K+1)}$.

By denoting

$$\boldsymbol{V} = \tilde{\boldsymbol{Z}}^{-1} \left[\frac{\partial \boldsymbol{\psi}(\boldsymbol{\vartheta}^{(0)})}{\partial \boldsymbol{\theta}^{\top}} - E\left\{ \frac{\partial \boldsymbol{\psi}(\boldsymbol{\vartheta}^{(0)})}{\partial \boldsymbol{\theta}^{\top}} \right\} \right] \tilde{\boldsymbol{Z}}^{-1} = \begin{pmatrix} \boldsymbol{V}^{(11)} & \boldsymbol{V}^{(12)} \\ \boldsymbol{V}^{(21)} & \mathbf{0}_{(K+1)\times(K+1)} \end{pmatrix},$$

where $\boldsymbol{V}^{(11)} \in \mathbb{R}^{pd\times pd}, \boldsymbol{V}^{(12)} \in \mathbb{R}^{pd\times(K+1)}$ and $\boldsymbol{V}^{(21)} \in \mathbb{R}^{(K+1)\times pd}$, the results from (I) – (IV) imply

$$\left|\boldsymbol{V}^{11}\right|_{\infty} = O_{\mathrm{P}}\left(\frac{1}{\sqrt{n}}\right), \left|\boldsymbol{V}^{12}\right|_{\infty} = O_{\mathrm{P}}\left(\frac{1}{\sqrt{np}}\right), \text{ and } \left|\boldsymbol{V}^{21}\right|_{\infty} = O_{\mathrm{P}}\left(\frac{1}{\sqrt{n}}\right). \tag{S5.25}$$

Then, we study the properties of $\Xi^{(\mathcal{S})}\boldsymbol{B}^{-1}\boldsymbol{V}$ where

$$\begin{aligned} \Xi^{(\mathcal{S})}\boldsymbol{B}^{-1}\boldsymbol{V} &= \begin{pmatrix} \boldsymbol{T}^{(\mathcal{S})} \otimes \boldsymbol{I}_d & \mathbf{0}_{qd\times(K+1)} \\ \mathbf{0}_{(K+1)\times pd} & \boldsymbol{I}_{K+1} \end{pmatrix} \begin{pmatrix} \boldsymbol{B}^{(11)} & \mathbf{0}_{pd\times(K+1)} \\ \boldsymbol{B}^{(21)} & \boldsymbol{B}^{(22)} \end{pmatrix} \begin{pmatrix} \boldsymbol{V}^{(11)} & \boldsymbol{V}^{(12)} \\ \boldsymbol{V}^{(21)} & \mathbf{0}_{(K+1)\times(K+1)} \end{pmatrix} \\ &= \begin{pmatrix} (\boldsymbol{T}^{(\mathcal{S})} \otimes \boldsymbol{I}_d)\boldsymbol{B}^{(11)}\boldsymbol{V}^{(11)} & (\boldsymbol{T}^{(\mathcal{S})} \otimes \boldsymbol{I}_d)\boldsymbol{B}^{(11)}\boldsymbol{V}^{(12)} \\ \boldsymbol{B}^{(21)}\boldsymbol{V}^{(11)} + \boldsymbol{B}^{(22)}\boldsymbol{V}^{(21)} & \boldsymbol{B}^{(21)}\boldsymbol{V}^{(12)} \end{pmatrix}. \end{aligned}$$

Let $(\boldsymbol{B}^{(11)}\boldsymbol{V}^{(11)})_{l_1 l_2}$ be the (l_1, l_2)-th element of $\boldsymbol{B}^{(11)}\boldsymbol{V}^{(11)}$, $b_{l_1 l}^{(11)}$ and $v_{l l_2}^{(11)}$

be the (l_1, l)-th element of $\boldsymbol{B}^{(11)}$ and the (l, l_2)-th element of $\boldsymbol{V}^{(11)}$, respectively, for $l, l_1, l_2 = 1, \cdots, pd$. By (S5.25) and Condition 5, we obtain

$$
\begin{aligned}
\left(\boldsymbol{B}^{(11)}\boldsymbol{V}^{(11)}\right)_{l_1 l_2} = \sum_{l=1}^{pd} b_{l_1 l}^{(11)} v_{l l_2}^{(11)} &\leq \left|\boldsymbol{V}^{(11)}\right|_\infty \sum_{l=1}^{pd} b_{l_1 l}^{(11)} \\
&\leq \left|\boldsymbol{V}^{(11)}\right|_\infty \left\|\boldsymbol{B}^{-1}\right\|_\infty \\
&= O_{\mathrm{P}}\left(\frac{1}{\sqrt{n}}\right), \text{ uniformly for } l_1, l_2 = 1, \cdots, pd.
\end{aligned}
$$

This implies

$$
\left|\boldsymbol{B}^{(11)}\boldsymbol{V}^{(11)}\right|_\infty = O_{\mathrm{P}}\left(\frac{1}{\sqrt{n}}\right). \tag{S5.26}
$$

By using the fact that $(\boldsymbol{T}^{(\mathcal{S})} \otimes \boldsymbol{I}_d)\boldsymbol{B}^{(11)}\boldsymbol{V}^{(11)}$ is a $qd \times pd$ submatrix of $\boldsymbol{B}^{(11)}\boldsymbol{V}^{(11)}$ and employing inequality 4.67(b) in Seber (2008), we obtain

$$
\begin{aligned}
\left\|(\boldsymbol{T}^{(\mathcal{S})} \otimes \boldsymbol{I}_d)\boldsymbol{B}^{(11)}\boldsymbol{V}^{(11)}\right\|_2 &\leq \sqrt{(qd)(pd)} \left|(\boldsymbol{T}^{(\mathcal{S})} \otimes \boldsymbol{I}_d)\boldsymbol{B}^{(11)}\boldsymbol{V}^{(11)}\right|_\infty \\
&\leq \sqrt{(qd)(pd)} \left|\boldsymbol{B}^{(11)}\boldsymbol{V}^{(11)}\right|_\infty \\
&= \sqrt{(qd)(pd)} O_{\mathrm{P}}\left(\frac{1}{\sqrt{n}}\right) \\
&= O_{\mathrm{P}}\left(\sqrt{\frac{p}{n}}\right).
\end{aligned} \tag{S5.27}
$$

By similar techniques to those used in showing (S5.26) and (S5.27), we obtain

$$
\left|\boldsymbol{B}^{(11)}\boldsymbol{V}^{(12)}\right|_\infty = O_{\mathrm{P}}\left(\frac{1}{\sqrt{np}}\right), \ \left|\boldsymbol{B}^{(21)}\boldsymbol{V}^{(11)}\right|_\infty = O_{\mathrm{P}}\left(\frac{1}{\sqrt{n}}\right),
$$

$$\left|\boldsymbol{B}^{(22)}\boldsymbol{V}^{(21)}\right|_\infty = O_{\mathrm{P}}\left(\frac{1}{\sqrt{n}}\right), \text{ and } \left|\boldsymbol{B}^{(21)}\boldsymbol{V}^{(12)}\right|_\infty = O_{\mathrm{P}}\left(\frac{1}{\sqrt{np}}\right), \quad \text{(S5.28)}$$

which leads to

$$\left\|(\boldsymbol{T}^{(\mathcal{S})}\otimes \boldsymbol{I}_d)\boldsymbol{B}^{(11)}\boldsymbol{V}^{(12)}\right\|_2 = O_{\mathrm{P}}\left(\frac{1}{\sqrt{np}}\right), \ \left\|\boldsymbol{B}^{(21)}\boldsymbol{V}^{(11)} + \boldsymbol{B}^{(22)}\boldsymbol{V}^{(21)}\right\|_2 = O_{\mathrm{P}}\left(\sqrt{\frac{p}{n}}\right), \text{ and}$$
$$\left\|\boldsymbol{B}^{(21)}\boldsymbol{V}^{(12)}\right\|_2 = O_{\mathrm{P}}\left(\frac{1}{\sqrt{np}}\right).$$

The above results imply

$$\begin{aligned}
&\left\|\boldsymbol{\Xi}^{(\mathcal{S})}\boldsymbol{B}^{-1}\boldsymbol{V}\right\|_2 \\
&\le \left\|(\boldsymbol{T}^{(\mathcal{S})}\otimes \boldsymbol{I}_d)\boldsymbol{B}^{(11)}\boldsymbol{V}^{(11)}\right\|_2 + \left\|(\boldsymbol{T}^{(\mathcal{S})}\otimes \boldsymbol{I}_d)\boldsymbol{B}^{(11)}\boldsymbol{V}^{(12)}\right\|_2 \\
&\quad + \left\|\boldsymbol{B}^{(21)}\boldsymbol{V}^{(11)} + \boldsymbol{B}^{(22)}\boldsymbol{V}^{(21)}\right\|_2 + \left\|\boldsymbol{B}^{(21)}\boldsymbol{V}^{(12)}\right\|_2 \\
&= O_{\mathrm{P}}\left(\sqrt{\frac{p}{n}}\right) + O_{\mathrm{P}}\left(\frac{1}{\sqrt{np}}\right) + O_{\mathrm{P}}\left(\sqrt{\frac{p}{n}}\right) + O_{\mathrm{P}}\left(\frac{1}{\sqrt{np}}\right) \\
&= O_{\mathrm{P}}\left(\sqrt{\frac{p}{n}}\right).
\end{aligned}$$

By recalling $\lambda_{\min}(\boldsymbol{H})$ and $\lambda_{\max}(\boldsymbol{H})$ denote the smallest and largest eigenvalues, respestively, of a generic square matrix $\boldsymbol{H}$, we have

$$\begin{aligned}
\left\|\left(\boldsymbol{B}^{-1}\boldsymbol{V} + \boldsymbol{I}_{pd+K+1}\right)^{-1}\right\|_2 &= \sqrt{\lambda_{\max}\left\{\left(\boldsymbol{B}^{-1}\boldsymbol{V} + \boldsymbol{I}_{pd+K+1}\right)^{-1\top}\left(\boldsymbol{B}^{-1}\boldsymbol{V} + \boldsymbol{I}_{pd+K+1}\right)^{-1}\right\}} \\
&= \frac{1}{\sqrt{\lambda_{\min}\left\{\left(\boldsymbol{B}^{-1}\boldsymbol{V} + \boldsymbol{I}_{pd+K+1}\right)\left(\boldsymbol{B}^{-1}\boldsymbol{V} + \boldsymbol{I}_{pd+K+1}\right)^{\top}\right\}}} \\
&= \frac{1}{\sqrt{\lambda_{\min}\left(\boldsymbol{I}_{pd+K+1} + \boldsymbol{B}^{-1}\boldsymbol{V} + \boldsymbol{V}^\top\boldsymbol{B}^{-1\top} + \boldsymbol{B}^{-1}\boldsymbol{V}\boldsymbol{V}^\top\boldsymbol{B}^{-1\top}\right)}}
\end{aligned}$$

$$= \left\{1 + \lambda_{\min}\left(\boldsymbol{B}^{-1}\boldsymbol{V} + \boldsymbol{V}^\top\boldsymbol{B}^{-1\top} + \boldsymbol{B}^{-1}\boldsymbol{V}\boldsymbol{V}^\top\boldsymbol{B}^{-1\top}\right)\right\}^{-\frac{1}{2}}.$$

By (S5.28) and inequality 4.67(b) in Seber (2008), we obtain

$$\begin{aligned}
\left\|\boldsymbol{B}^{-1}\boldsymbol{V}\right\|_2 &\le \left\|\boldsymbol{B}^{(11)}\boldsymbol{V}^{(11)}\right\|_2 + \left\|\boldsymbol{B}^{(11)}\boldsymbol{V}^{(12)}\right\|_2 + \left\|\boldsymbol{B}^{(21)}\boldsymbol{V}^{(11)} + \boldsymbol{B}^{(22)}\boldsymbol{V}^{(21)}\right\|_2 + \left\|\boldsymbol{B}^{(21)}\boldsymbol{V}^{(12)}\right\|_2 \\
&\le \sqrt{(pd)(pd)}\left|\boldsymbol{B}^{(11)}\boldsymbol{V}^{(11)}\right|_\infty + \sqrt{(pd)(K+1)}\left|\boldsymbol{B}^{(11)}\boldsymbol{V}^{(12)}\right|_\infty + \sqrt{(K+1)(pd)}\left|\boldsymbol{B}^{(21)}\boldsymbol{V}^{(11)}\right|_\infty \\
&\quad + \sqrt{(K+1)(pd)}\left|\boldsymbol{B}^{(22)}\boldsymbol{V}^{(21)}\right|_\infty + \sqrt{(K+1)(K+1)}\left|\boldsymbol{B}^{(21)}\boldsymbol{V}^{(12)}\right|_\infty \\
&= O_{\mathrm{P}}\left(\frac{p}{\sqrt{n}}\right) + O_{\mathrm{P}}\left(\frac{1}{\sqrt{n}}\right) + O_{\mathrm{P}}\left(\sqrt{\frac{p}{n}}\right) + O_{\mathrm{P}}\left(\sqrt{\frac{p}{n}}\right) + O_{\mathrm{P}}\left(\frac{1}{\sqrt{np}}\right) \\
&= O_{\mathrm{P}}\left(\frac{p}{\sqrt{n}}\right),
\end{aligned}$$

which leads to

$$\begin{aligned}
\left|\lambda_{\min}\left(\boldsymbol{B}^{-1}\boldsymbol{V} + \boldsymbol{V}^\top\boldsymbol{B}^{-1\top} + \boldsymbol{B}^{-1}\boldsymbol{V}\boldsymbol{V}^\top\boldsymbol{B}^{-1\top}\right)\right| &\le \left\|\boldsymbol{B}^{-1}\boldsymbol{V} + \boldsymbol{V}^\top\boldsymbol{B}^{-1\top} + \boldsymbol{B}^{-1}\boldsymbol{V}\boldsymbol{V}^\top\boldsymbol{B}^{-1\top}\right\|_2 \\
&\le 2\left\|\boldsymbol{B}^{-1}\boldsymbol{V}\right\|_2 + \left\|\boldsymbol{B}^{-1}\boldsymbol{V}\right\|_2^2 \\
&\le 2O_{\mathrm{P}}\left(\frac{p}{\sqrt{n}}\right) + O_{\mathrm{P}}\left(\frac{p^2}{n}\right) \\
&\xrightarrow{\mathrm{P}} 0,
\end{aligned}$$

as $n \to \infty$ and $p = o(n^{1/2})$. This implies

$$1 + \lambda_{\min}\left(\boldsymbol{B}^{-1}\boldsymbol{V} + \boldsymbol{V}^\top\boldsymbol{B}^{-1\top} + \boldsymbol{B}^{-1}\boldsymbol{V}\boldsymbol{V}^\top\boldsymbol{B}^{-1\top}\right) \xrightarrow{\mathrm{P}} 1.$$

We then apply continuous mapping theorem to obtain

$$\left\|\left(\boldsymbol{B}^{-1}\boldsymbol{V}+\boldsymbol{I}_{pd+K+1}\right)^{-1}\right\|_2=\left\{1+\lambda_{\min}\left(\boldsymbol{B}^{-1}\boldsymbol{V}+\boldsymbol{V}^\top\boldsymbol{B}^{-1\top}+\boldsymbol{B}^{-1}\boldsymbol{V}\boldsymbol{V}^\top\boldsymbol{B}^{-1\top}\right)\right\}^{-\frac{1}{2}}\xrightarrow{\text{P}}1.$$

Finally, we can combine all of the above results and obtain

$$\begin{aligned}
&\left\|\boldsymbol{\Omega}^{-1/2}(\boldsymbol{\Xi}^{(\mathcal{S})})\boldsymbol{\Xi}^{(\mathcal{S})}\boldsymbol{B}^{-1}\boldsymbol{V}(\boldsymbol{B}^{-1}\boldsymbol{V}+\boldsymbol{I}_{pd+K+1})^{-1}\boldsymbol{B}^{-1}\tilde{\boldsymbol{Z}}^{-1}\boldsymbol{\psi}(\boldsymbol{\vartheta}^{(0)})\right\|_2\\
&\leq\left\|\boldsymbol{\Omega}^{-1/2}(\boldsymbol{\Xi}^{(\mathcal{S})})\right\|_2\left\|\boldsymbol{\Xi}^{(\mathcal{S})}\boldsymbol{B}^{-1}\boldsymbol{V}\right\|_2\left\|(\boldsymbol{B}^{-1}\boldsymbol{V}+\boldsymbol{I}_{pd+K+1})^{-1}\right\|_2\left\|\boldsymbol{B}^{-1}\right\|_2\left\|\tilde{\boldsymbol{Z}}^{-1}\boldsymbol{\psi}(\boldsymbol{\vartheta}^{(0)})\right\|_2\\
&=O(1)O_{\text{P}}\left(\sqrt{\frac{p}{n}}\right)O_{\text{P}}(1)O(1)O_{\text{P}}(\sqrt{p})\\
&=O_{\text{P}}\left(\frac{p}{\sqrt{n}}\right)=o_{\text{P}}(1),
\end{aligned}$$

as $n\to\infty$ and $p=o(n^{1/2})$. This completes the proof of Lemma S3. □

Proof of Lemma S8. We directly have

$$\sum_{j=1}^q c_j^{1+\tau}\leq\sum_{j=1}^q c_jc_j^\tau\leq\sum_{j=1}^q c_j\left(\sum_{j=1}^q c_j\right)^\tau=\left(\sum_{j=1}^q c_j\right)^{1+\tau},$$

which completes the proof. □

Proof of Lemma S9. For any $s>0$, $\zeta>0$, $\delta>0$ and random variable z, we have the inequality

$$\delta^\zeta E\left(|z|^s I_{\{|z|>\delta\}}\right)\leq E\left(|z|^{s+\zeta}I_{\{|z|>\delta\}}\right)\leq E\,|z|^{s+\zeta},$$

where $I_{\{\cdot\}}$ is the indicator function. Hence, by condition (1) in Lemma S9,

$$E\left(\left|\frac{z_{mj}}{c_{mj}}\right| I_{\left\{\left|\frac{z_{mj}}{c_{mj}}\right|>\delta\right\}}\right) \leq \delta^{-\zeta} E\left|\frac{z_{mj}}{c_{mj}}\right|^{1+\zeta} \leq \delta^{-\zeta} C,$$

which implies $|z_{mj}/c_{mj}|$ is uniformly integrable. This result, together with conditions (2) – (3) of Lemma S9, leads to $\sum_{j=1}^{k_m} z_{mj} \xrightarrow{L^1} 0$ by Theorem 20.11 in Davidson (2021, p.428), which completes the proof by Chebyshev's inequality. □

Proof of Lemma S10. Let $\boldsymbol{H} = (h_{ij})_{a\times b} \in \mathbb{R}^{a\times b}$ and submatrix $\boldsymbol{H}' \in \mathbb{R}^{a'\times b'}$ that consists of the $\mathcal{V} = \{v_1, \cdots, v_{a'}\}$-th rows and $\mathcal{T} = \{\tau_1, \cdots, \tau_{b'}\}$-th columns of $\boldsymbol{H}$, where $a' \leq a$, $b' \leq b$, $\mathcal{V} \subseteq \{1, \cdots, a\}$ and $\mathcal{T} \subseteq \{1, \cdots, b\}$. For any $\boldsymbol{z}' = (z'_1, \cdots, z'_{b'})^\top \in \mathbb{R}^{b'}$ and $\boldsymbol{z}' \neq \boldsymbol{0}_{b'}$, we can construct another non-zero vector $\boldsymbol{z} = (z_1, \cdots, z_b)^\top \in \mathbb{R}^b$ where $z_{\tau_k} = z'_k$ for $k = 1, \cdots, b'$ and $z_j = 0$ for $j \notin \mathcal{T}$. By constructing $\boldsymbol{z}$ in such a way, we have $\|\boldsymbol{z}\|_2 = \|\boldsymbol{z}'\|_2$. Then, we obtain

$$\begin{aligned}\|\boldsymbol{H}'\boldsymbol{z}'\|_2^2 = \sum_{i\in\mathcal{V}}\left(\sum_{j\in\mathcal{T}} h_{ij}z_j\right)^2 &\leq \sum_{i\in\mathcal{V}}\left(\sum_{j\in\mathcal{T}} h_{ij}z_j\right)^2 + \sum_{i\notin\mathcal{V}}\left(\sum_{j\in\mathcal{T}} h_{ij}z_j\right)^2 \\ &= \sum_{i=1}^{a}\left(\sum_{j\in\mathcal{T}} h_{ij}z_j\right)^2 = \sum_{i=1}^{a}\left(\sum_{j=1}^{b} h_{ij}z_j\right)^2 = \|\boldsymbol{H}\boldsymbol{z}\|_2^2,\end{aligned}$$

which implies $\|\boldsymbol{H}'\boldsymbol{z}'\|_2 \leq \|\boldsymbol{H}\boldsymbol{z}\|_2$. This, together with $\|\boldsymbol{z}\|_2 = \|\boldsymbol{z}'\|_2$, implies

$$\frac{\|\boldsymbol{H}'\boldsymbol{z}'\|_2}{\|\boldsymbol{z}'\|_2} \leq \frac{\|\boldsymbol{H}\boldsymbol{z}\|_2}{\|\boldsymbol{z}\|_2} \leq \sup\left\{\frac{\|\boldsymbol{H}\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2} : \boldsymbol{x} \in \mathbb{R}^b \text{ and } \boldsymbol{x} \neq \boldsymbol{0}_b\right\} = \|\boldsymbol{H}\|_2\,,$$

for any $\boldsymbol{z}' \in \mathbb{R}^{b'}$ and $\boldsymbol{z}' \neq \boldsymbol{0}_{b'}$. This further leads to

$$\|\boldsymbol{H}'\|_2 = \sup\left\{\frac{\|\boldsymbol{H}'\boldsymbol{x}'\|_2}{\|\boldsymbol{x}'\|_2} : \boldsymbol{x}' \in \mathbb{R}^{b'} \text{ and } \boldsymbol{x}' \neq \boldsymbol{0}_{b'}\right\} \leq \|\boldsymbol{H}\|_2\,,$$

which completes the proof. □

S6 Proofs of Proposition 1, Theorems 1–2 and Corollary 1

S6.1 Proof of Proposition 1

Suppose we have $\boldsymbol{\alpha} = (\alpha_0, \cdots, \alpha_K)^\top \in \mathscr{A}^+$, this implies $\boldsymbol{\Sigma}(\boldsymbol{\alpha}) = \alpha_0\boldsymbol{I}_p + \sum_{k=1}^K \alpha_k\boldsymbol{W}_k$ is a positive definite matrix and hence is a covariance matrix of some random vector $\boldsymbol{z} \in \mathbb{R}^p$. Let $\boldsymbol{\rho} = (\alpha_1/\alpha_0, \cdots, \alpha_K/\alpha_0)$. Then, we have

$$\boldsymbol{R}(\boldsymbol{\rho}) = \boldsymbol{I}_p + \sum_{k=1}^K \alpha_k/\alpha_0\boldsymbol{W}_k = (\alpha_0\boldsymbol{I}_p)^{-1/2}\boldsymbol{\Sigma}(\boldsymbol{\alpha})(\alpha_0\boldsymbol{I}_p)^{-1/2}. \tag{S6.29}$$

It is worth noting that every element of $\boldsymbol{z}$ has the same variance of α_0 because the diagonals of $\boldsymbol{W}_k$ are all zeros for $k = 1, \cdots, K$ as defined above equation (2.1). This implies $\boldsymbol{R}(\boldsymbol{\rho})$ is a correlation matrix of $\boldsymbol{z}$ since it is obtained via standardizing the covariance matrix as shown in equation (S6.29). Consequently, $\boldsymbol{\rho} = (\alpha_1/\alpha_0, \cdots, \alpha_K/\alpha_0) \in \mathscr{P}^+$, which completes the proof.

S6.2 Proof of Theorem 1

By recalling $\hat{\boldsymbol{\vartheta}} = (\hat{\boldsymbol{\beta}}^\top, \hat{\boldsymbol{\alpha}}^\top)^\top$ is the Z-estimator (van der Vaart, 1998, p. 41) that solves the joint estimating equation $\boldsymbol{\psi}(\boldsymbol{\vartheta}) = \mathbf{0}_{pd+K+1}$ where $\boldsymbol{\psi}(\boldsymbol{\vartheta})$ is defined in Section 2.1, we prove (i) $\hat{\boldsymbol{\vartheta}}$ is consistent; (ii) $\hat{\boldsymbol{\vartheta}}$ is asymptotically normal. Since (i) can be obtained by Theorem 5.9 of van der Vaart (1998, p. 46) given all the conditions in this theorem, we only provide the proof of (ii).

By using Taylor series expansion at $\boldsymbol{\vartheta}^{(0)}$, we have $0 = \boldsymbol{\psi}(\hat{\boldsymbol{\vartheta}}) = \boldsymbol{\psi}(\boldsymbol{\vartheta}^{(0)}) + \{\partial\boldsymbol{\psi}(\boldsymbol{\vartheta}^{(0)})/\partial\boldsymbol{\vartheta}^\top\}(\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}^{(0)}) + \bar{\boldsymbol{r}}$ where

$$\bar{\boldsymbol{r}} = \frac{1}{2}\left\{\boldsymbol{I}_{pd+K+1} \otimes (\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}^{(0)})^\top\right\} \frac{\partial}{\partial\boldsymbol{\vartheta}^\top}\text{vec}\left\{\frac{\partial\boldsymbol{\psi}(\bar{\boldsymbol{\vartheta}})}{\partial\boldsymbol{\vartheta}^\top}\right\}(\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}^{(0)}),$$

and $\bar{\boldsymbol{\vartheta}}$ lies between $\hat{\boldsymbol{\vartheta}}$ and $\boldsymbol{\vartheta}^{(0)}$. This implies

$$\boldsymbol{\Omega}^{-1/2}(\boldsymbol{\Xi}^{(\mathcal{S})})\boldsymbol{Z}\boldsymbol{\Xi}^{(\mathcal{S})}(\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}^{(0)}) = \boldsymbol{u}_1 + \boldsymbol{u}_2 + \boldsymbol{u}_3,$$

where $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ are defined in Lemmas S2 and S3, respectively, and

$$\boldsymbol{u}_3 = -\boldsymbol{\Omega}^{-1/2}(\boldsymbol{\Xi}^{(\mathcal{S})})\boldsymbol{Z}\boldsymbol{\Xi}^{(\mathcal{S})}\left\{\frac{\partial\boldsymbol{\psi}(\boldsymbol{\vartheta}^{(0)})}{\partial\boldsymbol{\vartheta}^\top}\right\}^{-1}\bar{\boldsymbol{r}}.$$

Employing similar techniques to those used in the proofs of Lemmas S2 – S3, we obtain $\boldsymbol{u}_3 = o_{\mathrm{P}}(1)$. This, in conjunction with Lemmas S2 – S3 and Slutsky's theorem, implies

$$\boldsymbol{\Omega}^{-1/2}(\Xi^{(\mathcal{S})})\boldsymbol{Z}\Xi^{(\mathcal{S})}(\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}^{(0)}) = \boldsymbol{u}_1 + o_{\mathrm{P}}(1) \xrightarrow{d} N(\mathbf{0}_{qd+K+1}, \boldsymbol{I}_{qd+K+1}),$$

as $n \to \infty$ and $p = o(n^{1/2})$, which completes the proof.

S6.3 Proof of Theorem 2

By recalling $\hat{\boldsymbol{\theta}} = \boldsymbol{f}(\hat{\boldsymbol{\vartheta}})$ and using similar ideas from the proof of the multivariate delta method in Theorem 3.1 of van der Vaart (1998, p. 26), we can obtain the result for Theorem 2. However, we could not apply Theorem 3.1 of van der Vaart (1998, p. 26) directly as $p \to \infty$ because $\hat{\boldsymbol{\theta}}$ has a diverging dimension in this case. To this end, we combine the techniques from the multivariate delta method and Theorem 1 to show the result.

Based on the proof of Theorem 1, we obtain the asymptotic expansion

$$\begin{aligned}\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}^{(0)} = &-\tilde{\boldsymbol{Z}}^{-1}\boldsymbol{B}^{-1}\tilde{\boldsymbol{Z}}^{-1}\boldsymbol{\psi}(\boldsymbol{\vartheta}^{(0)}) + \tilde{\boldsymbol{Z}}^{-1}\boldsymbol{B}^{-1}\boldsymbol{V}(\boldsymbol{B}^{-1}\boldsymbol{V} + \boldsymbol{I}_{pd+K+1})^{-1}\boldsymbol{B}^{-1}\tilde{\boldsymbol{Z}}^{-1}\boldsymbol{\psi}(\boldsymbol{\vartheta}^{(0)}) \\ &- \left(\frac{\partial \boldsymbol{\psi}(\boldsymbol{\vartheta}^{(0)})}{\partial \boldsymbol{\vartheta}^{\top}}\right)^{-1}\bar{\boldsymbol{r}}. \end{aligned} \tag{S6.30}$$

On the other hand, by Taylor series expansion, we obtain

$$\boldsymbol{f}(\hat{\boldsymbol{\vartheta}}) - \boldsymbol{f}(\boldsymbol{\vartheta}^{(0)}) = \frac{\partial \boldsymbol{f}(\boldsymbol{\vartheta}^{(0)})}{\partial \boldsymbol{\vartheta}^\top}(\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}^{(0)}) + \bar{\boldsymbol{r}}_f, \tag{S6.31}$$

where

$$\bar{\boldsymbol{r}}_f = \frac{1}{2}\left\{\boldsymbol{I}_{pd+K} \otimes (\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}^{(0)})^\top\right\} \frac{\partial}{\partial \boldsymbol{\vartheta}^\top} \text{vec}\left\{\frac{\partial \boldsymbol{f}(\bar{\boldsymbol{\vartheta}}_f)}{\partial \boldsymbol{\vartheta}^\top}\right\}(\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}^{(0)}),$$

and $\bar{\boldsymbol{\vartheta}}_f$ lies between $\hat{\boldsymbol{\vartheta}}$ and $\boldsymbol{\vartheta}^{(0)}$. Note that the Jacobian matrix of the vector-valued function $\boldsymbol{f}(\boldsymbol{\vartheta})$ is given as

$$\frac{\partial \boldsymbol{f}(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}^\top} = \begin{pmatrix} \boldsymbol{I}_{pd} & \mathbf{0}_{pd\times(K+1)} \\ \mathbf{0}_{K\times pd} & \frac{\partial \boldsymbol{\rho}}{\partial \boldsymbol{\alpha}^\top} \end{pmatrix},$$

where $\boldsymbol{\rho} = (\rho_1, \cdots, \rho_K)^\top = (\alpha_1/\alpha_0, \cdots, \alpha_K/\alpha_0)^\top$ and

$$\frac{\partial \boldsymbol{\rho}}{\partial \boldsymbol{\alpha}^\top} = \begin{pmatrix} -\frac{\alpha_1}{\alpha_0^2} & \frac{1}{\alpha_0} & 0 & \cdots & 0 \\ -\frac{\alpha_2}{\alpha_0^2} & 0 & \frac{1}{\alpha_0} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\frac{\alpha_K}{\alpha_0^2} & 0 & 0 & \cdots & \frac{1}{\alpha_0} \end{pmatrix}.$$

By substituting (S6.30) into (S6.31), we obtain

$$\bar{\boldsymbol{\Omega}}^{-1/2}(\bar{\boldsymbol{\Xi}}^{(\mathcal{S})})\bar{\boldsymbol{Z}}\bar{\boldsymbol{\Xi}}^{(\mathcal{S})}\{\boldsymbol{f}(\hat{\boldsymbol{\vartheta}}) - \boldsymbol{f}(\boldsymbol{\vartheta}^{(0)})\} = \bar{\boldsymbol{u}}_1 + \bar{\boldsymbol{u}}_2 + \bar{\boldsymbol{u}}_3 + \bar{\boldsymbol{u}}_4,$$

where

$$\bar{\boldsymbol{u}}_1 = -\bar{\boldsymbol{\Omega}}^{-1/2}(\bar{\Xi}^{(\mathcal{S})})\bar{\Xi}^{(\mathcal{S})}\frac{\partial \boldsymbol{f}(\boldsymbol{\vartheta}^{(0)})}{\partial \boldsymbol{\vartheta}^\top}\boldsymbol{B}^{-1}\tilde{\boldsymbol{Z}}^{-1}\boldsymbol{\psi}(\boldsymbol{\vartheta}^{(0)}),$$

$$\bar{\boldsymbol{u}}_2 = \bar{\boldsymbol{\Omega}}^{-1/2}(\bar{\Xi}^{(\mathcal{S})})\bar{\Xi}^{(\mathcal{S})}\frac{\partial \boldsymbol{f}(\boldsymbol{\vartheta}^{(0)})}{\partial \boldsymbol{\vartheta}^\top}\boldsymbol{B}^{-1}\boldsymbol{V}(\boldsymbol{B}^{-1}\boldsymbol{V} + \boldsymbol{I}_{pd+K+1})^{-1}\boldsymbol{B}^{-1}\tilde{\boldsymbol{Z}}^{-1}\boldsymbol{\psi}(\boldsymbol{\vartheta}^{(0)}),$$

$$\bar{\boldsymbol{u}}_3 = -\bar{\boldsymbol{\Omega}}^{-1/2}(\bar{\Xi}^{(\mathcal{S})})\bar{\boldsymbol{Z}}\bar{\Xi}^{(\mathcal{S})}\frac{\partial \boldsymbol{f}(\boldsymbol{\vartheta}^{(0)})}{\partial \boldsymbol{\vartheta}^\top}\left\{\frac{\partial \boldsymbol{\psi}(\boldsymbol{\vartheta}^{(0)})}{\partial \boldsymbol{\vartheta}^\top}\right\}^{-1}\bar{\boldsymbol{r}}, \text{ and}$$

$$\bar{\boldsymbol{u}}_4 = \bar{\boldsymbol{\Omega}}^{-1/2}(\bar{\Xi}^{(\mathcal{S})})\bar{\boldsymbol{Z}}\bar{\Xi}^{(\mathcal{S})}\bar{\boldsymbol{r}}_f.$$

Under Conditions 1 – 5, we can employ similar techniques to those used in the proofs of Lemmas S2 – S3 to obtain $\bar{\boldsymbol{u}}_1 \xrightarrow{d} N(\boldsymbol{0}_{qd+K}, \boldsymbol{I}_{qd+K})$, $\bar{\boldsymbol{u}}_2 = o_{\mathrm{P}}(1)$, $\bar{\boldsymbol{u}}_3 = o_{\mathrm{P}}(1)$ and $\bar{\boldsymbol{u}}_4 = o_{\mathrm{P}}(1)$ as $n \to \infty$ and $p = o(n^{1/2})$. These, in conjunction with Slutsky's theorem, imply

$$\bar{\boldsymbol{\Omega}}^{-1/2}(\bar{\Xi}^{(\mathcal{S})})\bar{\boldsymbol{Z}}\bar{\Xi}^{(\mathcal{S})}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^{(0)}) \xrightarrow{d} N(\boldsymbol{0}_{qd+K}, \boldsymbol{I}_{qd+K}),$$

since $\hat{\boldsymbol{\theta}} = \boldsymbol{f}(\hat{\boldsymbol{\vartheta}})$ and $\boldsymbol{\theta}^{(0)} = \boldsymbol{f}(\boldsymbol{\vartheta}^{(0)})$. This completes the proof of Theorem 2.

S6.4 Proof of Corollary 1

By Condition 2 and inequality 4.67(e) in Seber (2008), we obtain $\sup_{p\geq 1}\|\boldsymbol{W}_k\|_2 \leq C_W$ for $k = 0, \cdots, K$. This, together with Proposition 1 of Zou et al. (2017), implies the required parameter space $\mathscr{A}^+$ for $\boldsymbol{\alpha}$ is an open set. By recalling the true value $\boldsymbol{\alpha}^{(0)}$ has to satisfy $\boldsymbol{\Sigma}(\boldsymbol{\alpha}^{(0)})$ being positive definite, we have that

$\boldsymbol{\alpha}^{(0)}$ is an interior point of $\mathscr{A}^+$ and hence there exists an open neighborhood $\mathscr{A}^{(0)} \subset \mathscr{A}^+$ around $\boldsymbol{\alpha}^{(0)}$. Furthermore, the conditions in this corollary can imply the estimator $\hat{\boldsymbol{\alpha}}$ is $\sqrt{np}$-consistent based on Theorem 1. Additionally, the event $\{\|\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^{(0)}\| \leq M_\tau/(np)^{1/2}\}$ implies the event $\{\hat{\boldsymbol{\alpha}} \in \mathscr{A}^{(0)} \subset \mathscr{A}^+\}$ as np is sufficiently large and $M_\tau > 0$ is some finite positive constant. Based on the above, we have that for any $\tau > 0$, there exists a finite $M_\tau > 0$ such that

$$
\begin{aligned}
&1 - \tau \leq \mathrm{P}(\left\|\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^{(0)}\right\| \leq M_\tau/\sqrt{np}) \leq \mathrm{P}(\hat{\boldsymbol{\alpha}} \in \mathscr{A}^{(0)} \subset \mathscr{A}^+) \leq \mathrm{P}(\hat{\boldsymbol{\alpha}} \in \mathscr{A}^+) \leq 1 \\
&\implies -\tau \leq \mathrm{P}(\hat{\boldsymbol{\alpha}} \in \mathscr{A}^+) - 1 \leq 0 \leq \tau \\
&\implies |\mathrm{P}(\hat{\boldsymbol{\alpha}} \in \mathscr{A}^+) - 1| \leq \tau,
\end{aligned}
$$

as np is sufficiently large. This leads to $\mathrm{P}(\hat{\boldsymbol{\alpha}} \in \mathscr{A}^+) \to 1$ as $n \to \infty$ and $p = o(n^{1/2})$.

S7 Supplementary Details of Simulation Study

S7.1 Simulation Settings of Section 4

This section provide details of the simulation settings used to obtain the numerical results in Section 4 of the main text. Firstly, we used the covariate vectors $\boldsymbol{x}_i \in \mathbb{R}^4$ from the standardized environmental covariates in the ground beetle dataset of Section 5 for $i = 1, \cdots, n$. The similarity matrices $\boldsymbol{W}_k$ for

$k = 1, \cdots, 5$ were also obtained based on the standardized species trait vectors $\boldsymbol{z}_j \in \mathbb{R}^5$ for $j = 1, \cdots, p$ from the same dataset. For settings where n or p were greater than the number of available sites (i.e., 87 sites) or species (i.e., 38 species) in the application in Section 5, we generated additional covariate vectors or trait vectors whose elements were sampled randomly from the corresponding standardized environmental covariates or species traits, respectively.

The true mean regression coefficients $\beta_{jl}^{(0)}$ and correlation regression parameters $\rho_k^{(0)}$ were generated independently from the Gaussian distribution $N(0, 0.5)$ with mean zero and standard deviation 0.5 and the uniform distribution $U(-0.05, 0.05)$, respectively, where $\beta_{jl}^{(0)}$ and $\rho_k^{(0)}$ are the corresponding elements of $\boldsymbol{\beta}^{(0)}$ and $\boldsymbol{\rho}^{(0)}$, respectively, for $j = 1, \cdots, p$, $l = 1, \cdots, 4$, and $k = 1, \cdots, 5$; for the negative binomial responses, we set $\boldsymbol{\rho}^{(0)} = \mathbf{0}_5$ due to the lack of a publicly available method to simulate correlated multivariate negative binomial responses with $\boldsymbol{\rho}^{(0)} \neq \mathbf{0}_5$. For all four response types, we set the dispersion parameters $\phi_j = 1$ for $j = 1, \cdots, p$, and assumed these dispersion parameters were unknown and needed to be estimated as part of model fitting. Additionally, we used the link and variance functions inherited from the response distribution as discussed below equation (2.1) e.g., the log link $g(\mu) = \log(\mu)$ and variance function $h(\mu; \phi) = \mu + \phi\mu^2$ were used for negative binomial count responses. We then simulated multivariate response vectors $\boldsymbol{Y}_i$ for $i = 1, \cdots, n$ consisting of elements $\{Y_{ij} : j = 1, \cdots, p\}$ that had the same marginal distribution (based on one of the four distributions considered in Section 4), where the means and variances were given by $g^{-1}(\boldsymbol{x}_i^\top \boldsymbol{\beta}_j^{(0)})$ and $h\{g^{-1}(\boldsymbol{x}_i^\top \boldsymbol{\beta}_j^{(0)}); \phi_j\}$, respectively, with $\boldsymbol{\beta}_j^{(0)} = (\beta_{j1}^{(0)}, \cdots, \beta_{j4}^{(0)})^\top$, and the between-response correlation matrix of $\boldsymbol{Y}_i$ was given by the correlation regression model in equation (2.2). Specifically, we generated multivariate negative binomial responses using the `R` function `rnbinom`, while the `R` package `MASS` (Venables and Ripley, 2002) was used to generate multivariate Gaussian responses. Multivariate Bernoulli and Poisson responses in the simulation study presented in Section 4 were generated using the `R` package `PoisBinOrd` (Gao et al., 2019).

S7.2 Simulation Results for Negative Binomial and Gaussian Responses, 95% Confidence Intervals, and Algorithm Runtime

This section presents additional simulation results for the negative binomial and Gaussian responses, together with the empirical coverage probability of 95% confidence intervals for the mean regression coefficients and correlation regression parameters constructed individually as discussed in Section S1. In general, the results for the negative binomial and Gaussian responses from Figure S1 and Table S1 are similar to those of the Bernoulli and Poisson responses in Section 4 e.g., the averaged MSE and the coverage for the 95% confidence regions of the mean regression coefficients only improved when n increased while those of the correlation regression parameters improved when n and/or p increased. For the negative binomial responses, there was undercoverage for the confidence regions of the mean regression coefficients especially when n is small.

Figure S2 presents the empirical coverage probability of 95% confidence intervals for each of the pd mean regression coefficients $\{\beta_{jl}^{(0)} : j = 1, \cdots, p, l = 1, \cdots, d\}$ summarized using boxplots, given the number of coefficients is large and depends on the number of responses p in the simulation settings. Overall, the coverage of the confidence intervals for the mean regression coefficients tended to the nominal level of 95% as n increased, while increasing p had negligible effect on the coverage. Similar to Table S1, there was undercoverage for the case of negative binomial responses particularly when n is small. Tables S2 – S3 show the coverage of 95% confidence intervals for all the correlation regression parameters $\{\rho_k^{(0)} : k = 1, \cdots, 5\}$ were close to the nominal level under various settings of responses distribution, n and p.

Finally, Table S4 shows that Algorithm 1 is computationally very scalable: the average runtime was no more than three seconds under various settings of the simulation study, using an AMD Ryzen 7 CPU @ 3.60 GHz machine.

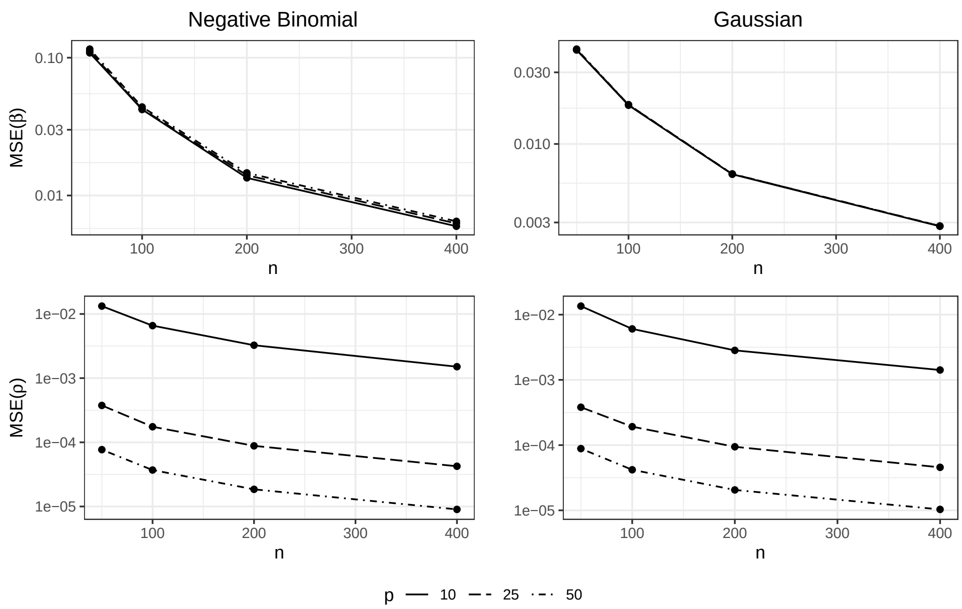

Figure S1: Averaged MSE of the mean regression coefficient estimators (top row) and correlation regression parameter estimators (bottom row) for negative binomial (left column) and Gaussian responses (right column).

Table S1: Empirical coverage probability for 95% confidence regions of $\boldsymbol{\beta}_{\mathcal{S}}^{(0)}$ and $\boldsymbol{\rho}^{(0)}$ for negative binomial and Gaussian responses, where $\mathcal{S} = \{1, \cdots, 5\}$.

		Negative Binomial			Gaussian		
		$p = 10$	$p = 25$	$p = 50$	$p = 10$	$p = 25$	$p = 50$
$\boldsymbol{\beta}_{\mathcal{S}}^{(0)}$	$n = 50$	0.765	0.765	0.757	0.882	0.893	0.892
	$n = 100$	0.884	0.886	0.886	0.924	0.922	0.922
	$n = 200$	0.923	0.923	0.923	0.941	0.942	0.943
	$n = 400$	0.934	0.936	0.937	0.934	0.937	0.936
$\boldsymbol{\rho}^{(0)}$	$n = 50$	0.893	0.886	0.896	0.852	0.888	0.882
	$n = 100$	0.926	0.924	0.921	0.907	0.918	0.922
	$n = 200$	0.928	0.934	0.926	0.929	0.932	0.932
	$n = 400$	0.937	0.948	0.927	0.943	0.933	0.931

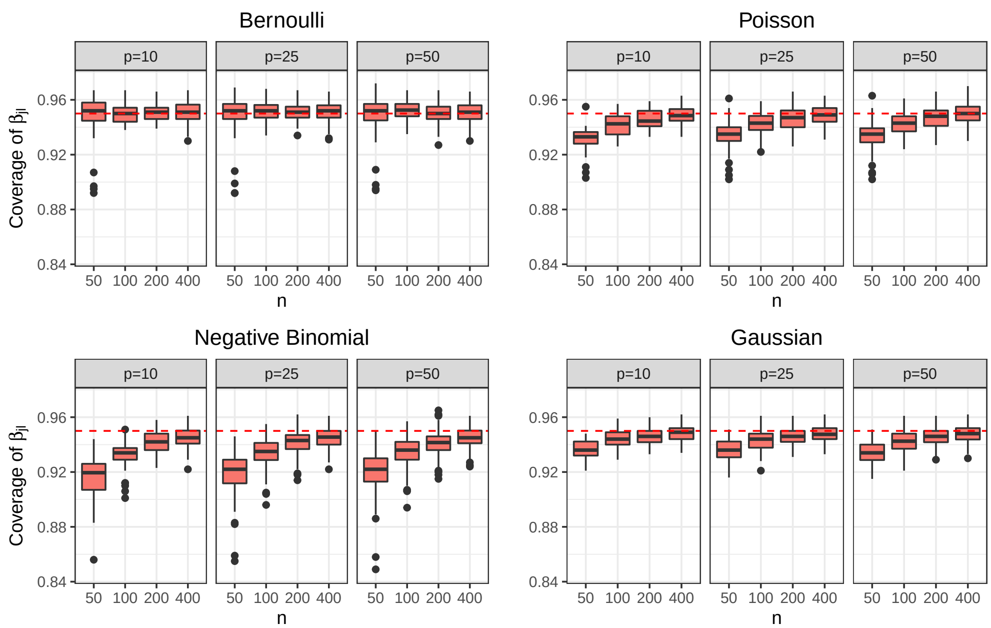

Figure S2: Boxplots summarizing the empirical coverage probability of 95% confidence intervals for each of the mean regression coefficients, where the dashed line represents the nominal level of 95%.

S7.3 Comparison to Other Methods

In this section, we performed a simulation study to compare the performance of the proposed joint estimator (denoted here as CorrReg estimator) to two alternative existing estimators. The first of these is a GEE assuming an independence working correlation matrix (GEE-Ind), which is equivalent to solving a variation of the estimating equation ψ_{β} in equation (2.4) where $\tilde{\boldsymbol{\Sigma}}(\boldsymbol{\alpha})$ is replaced by $\boldsymbol{I}_{np}$. This is implemented by fitting either `glm()` (for Bernoulli, Poisson and Gaussian responses) or `glm.nb()` from the `R` package `MASS` (for negative binomial responses), where $(Y_{1j}, \cdots, Y_{nj})^{\top}$ is the response vector and

Table S2: Empirical coverage probability of 95% confidence intervals for each of the correlation regression parameters with Bernoulli and Poisson responses.

		Bernoulli			Poisson		
		$p = 10$	$p = 25$	$p = 50$	$p = 10$	$p = 25$	$p = 50$
$\rho_1^{(0)}$	$n = 50$	0.951	0.950	0.945	0.934	0.943	0.924
	$n = 100$	0.940	0.945	0.941	0.946	0.940	0.934
	$n = 200$	0.952	0.948	0.947	0.948	0.945	0.944
	$n = 400$	0.950	0.962	0.950	0.951	0.936	0.947
$\rho_2^{(0)}$	$n = 50$	0.921	0.927	0.921	0.940	0.927	0.921
	$n = 100$	0.942	0.933	0.933	0.945	0.929	0.944
	$n = 200$	0.944	0.944	0.942	0.943	0.941	0.941
	$n = 400$	0.956	0.943	0.936	0.944	0.949	0.950
$\rho_3^{(0)}$	$n = 50$	0.941	0.936	0.931	0.933	0.932	0.940
	$n = 100$	0.944	0.942	0.947	0.945	0.937	0.931
	$n = 200$	0.953	0.950	0.938	0.944	0.940	0.937
	$n = 400$	0.951	0.948	0.938	0.949	0.949	0.951
$\rho_4^{(0)}$	$n = 50$	0.933	0.933	0.934	0.932	0.922	0.938
	$n = 100$	0.939	0.937	0.934	0.935	0.933	0.946
	$n = 200$	0.938	0.943	0.945	0.940	0.952	0.955
	$n = 400$	0.945	0.953	0.941	0.953	0.950	0.947
$\rho_5^{(0)}$	$n = 50$	0.933	0.928	0.928	0.917	0.923	0.928
	$n = 100$	0.938	0.934	0.946	0.943	0.937	0.957
	$n = 200$	0.952	0.928	0.944	0.942	0.948	0.950
	$n = 400$	0.960	0.946	0.953	0.939	0.945	0.944

$(\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n)^\top$ is the model matrix to estimate $\boldsymbol{\beta}_j$ separately, for $j = 1, \cdots, p$. The second alternative estimator we consider is a GEE assuming an unstructured working correlation matrix (GEE-Unstruc), which is equivalent to solving a variation of the estimating equation $\boldsymbol{\psi}_{\boldsymbol{\beta}}$ in equation (2.4) where $\tilde{\boldsymbol{\Sigma}}(\boldsymbol{\alpha})$ is replaced by $\boldsymbol{I}_n \otimes \boldsymbol{\Sigma}_{Unstruc}$ and $\boldsymbol{\Sigma}_{Unstruc}$ is modeled as a $p \times p$ unstructured correlation matrix with $p(p-1)/2$ parameters. The GEE-Unstruc estimator is obtained using a slight modification of Algorithm 1, where we iterate between

Table S3: Empirical coverage probability of 95% confidence intervals for each of the correlation regression parameters with negative binomial and Gaussian responses.

		Negative Binomial			Gaussian		
		$p = 10$	$p = 25$	$p = 50$	$p = 10$	$p = 25$	$p = 50$
$\rho_1^{(0)}$	$n = 50$	0.937	0.944	0.929	0.942	0.935	0.928
	$n = 100$	0.938	0.953	0.927	0.951	0.928	0.959
	$n = 200$	0.939	0.950	0.928	0.963	0.942	0.946
	$n = 400$	0.955	0.949	0.943	0.957	0.951	0.957
$\rho_2^{(0)}$	$n = 50$	0.942	0.933	0.937	0.933	0.938	0.923
	$n = 100$	0.951	0.942	0.947	0.936	0.951	0.934
	$n = 200$	0.948	0.944	0.948	0.942	0.947	0.944
	$n = 400$	0.954	0.944	0.938	0.954	0.943	0.952
$\rho_3^{(0)}$	$n = 50$	0.943	0.918	0.932	0.933	0.943	0.938
	$n = 100$	0.932	0.951	0.946	0.951	0.935	0.953
	$n = 200$	0.939	0.942	0.939	0.956	0.935	0.957
	$n = 400$	0.960	0.952	0.941	0.959	0.952	0.947
$\rho_4^{(0)}$	$n = 50$	0.917	0.938	0.948	0.940	0.928	0.947
	$n = 100$	0.942	0.944	0.943	0.929	0.943	0.956
	$n = 200$	0.943	0.942	0.930	0.956	0.955	0.953
	$n = 400$	0.940	0.954	0.941	0.956	0.948	0.950
$\rho_5^{(0)}$	$n = 50$	0.929	0.937	0.946	0.936	0.935	0.911
	$n = 100$	0.943	0.948	0.958	0.929	0.941	0.938
	$n = 200$	0.948	0.942	0.949	0.938	0.945	0.931
	$n = 400$	0.951	0.949	0.948	0.943	0.935	0.950

the two steps of estimating $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}_{Unstruc}$ rather than the correlation regression parameters $\boldsymbol{\rho}$. Specifically, given $\boldsymbol{\Sigma}_{Unstruc}$, we employ a similar Fisher scoring method to solve the modified estimating equation $\boldsymbol{\psi}_{\boldsymbol{\beta}}$ with $\tilde{\boldsymbol{\Sigma}}(\boldsymbol{\alpha})$ replaced by $\boldsymbol{I}_n \otimes \boldsymbol{\Sigma}_{Unstruc}$. Then, given $\boldsymbol{\beta}$, we estimate the unstructured correlation matrix $\boldsymbol{\Sigma}_{Unstruc}$ via the sample correlation matrix of the n vectors $\boldsymbol{\epsilon}_1(\boldsymbol{\beta}), \cdots, \boldsymbol{\epsilon}_n(\boldsymbol{\beta})$. That is, $\hat{\boldsymbol{\Sigma}}_{Unstruc} = \{\mathrm{diag}(\hat{\boldsymbol{\Sigma}}_{sample})\}^{-1/2}\hat{\boldsymbol{\Sigma}}_{sample}\{\mathrm{diag}(\hat{\boldsymbol{\Sigma}}_{sample})\}^{-1/2}$, where $\hat{\boldsymbol{\Sigma}}_{sample} = \sum_{i=1}^{n}\{\boldsymbol{\epsilon}_i(\boldsymbol{\beta}) - \bar{\boldsymbol{\epsilon}}(\boldsymbol{\beta})\}\{\boldsymbol{\epsilon}_i(\boldsymbol{\beta}) - \bar{\boldsymbol{\epsilon}}(\boldsymbol{\beta})\}^\top/(n-1)$ is the sample covari-

Table S4: Average runtime in seconds (over 1000 simulated datasets) of Algorithm 1.

	Bernoulli			Poisson		
	$p = 10$	$p = 25$	$p = 50$	$p = 10$	$p = 25$	$p = 50$
$n = 50$	0.050	0.109	0.341	0.044	0.097	0.306
$n = 100$	0.050	0.131	0.491	0.047	0.132	0.455
$n = 200$	0.061	0.205	0.795	0.062	0.199	0.772
$n = 400$	0.100	0.409	1.432	0.092	0.379	1.429
	Negative Binomial			Gaussian		
	$p = 10$	$p = 25$	$p = 50$	$p = 10$	$p = 25$	$p = 50$
$n = 50$	0.181	0.624	2.205	0.008	0.022	0.071
$n = 100$	0.166	0.636	2.195	0.013	0.036	0.124
$n = 200$	0.176	0.705	2.430	0.019	0.063	0.241
$n = 400$	0.195	0.839	3.028	0.034	0.131	0.513

ance matrix, $\bar{\boldsymbol{\epsilon}}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \boldsymbol{\epsilon}_i(\boldsymbol{\beta})/n$, $\mathrm{diag}(\hat{\boldsymbol{\Sigma}}_{sample})$ is a $p \times p$ diagonal matrix consisting of the diagonal elements of $\hat{\boldsymbol{\Sigma}}_{sample}$, and the definition of $\boldsymbol{\epsilon}_i(\boldsymbol{\beta})$ is given below equation (2.4).

Both GEE-Ind and GEE-Unstruc estimators were applied to the same 1000 datasets simulated under different combinations of n, p and response types in Sections 4 and S7.2, so as to facilitate comparison to our proposed joint estimator. However, `glm.nb()` tended to have non-convergence issues especially when n is small, so the results for negative binomial responses using GEE-Ind estimator were based on only a subset of the 1000 simulated datasets under which `glm.nb()` converged; see Table S5 for the number of datasets with converged `glm.nb()` results under different settings of n and p. The aforementioned modified algorithm for obtaining GEE-Unstruc estimator also occasionally did not converge when n is small, and so its results were based on subsets of 1000 simulated datasets under which the modified algorithm converged; see Table S6 for more details.

Figures S3 and S4 show that both GEE-Ind and GEE-Unstruc estimators offered comparable performance to our proposed CorrReg estimator in estimating the mean regression coefficients under different combinations of n, p and response types. This provides empirical support to the robustness of the mean regression coefficient estimator to misspecification of the working correlation structure, as is well known under the GEE framework. Next, we compared the performance of our proposed joint estimator and GEE-Unstruc estimator in recovering the true correlation matrix $\boldsymbol{R}^{(0)} = \boldsymbol{I}_p + \sum_{k=1}^{5} \rho_k^{(0)} \boldsymbol{W}_k$ by computing two types of average estimation errors, spectral-error and Frobenius-error, of correlation matrices measured under the spectral norm and the Frobenius norm, i.e., $\|\hat{\boldsymbol{R}} - \boldsymbol{R}^{(0)}\|_2$ and $p^{-1/2}\|\hat{\boldsymbol{R}} - \boldsymbol{R}^{(0)}\|_F$, where $\hat{\boldsymbol{R}} = \boldsymbol{I}_p + \sum_{k=1}^{5} \hat{\rho}_k \boldsymbol{W}_k$ for the proposed CorrReg estimator and $\hat{\boldsymbol{R}} = \hat{\Sigma}_{Unstruc}$ for GEE-Unstruc estimator. Figures S5 – S6 demonstrate the superior performance of the CorrReg estimator over the GEE-Unstruc estimator in recovering the true correlation matrix: this is not overly surprising since the latter does not directly utilize the extra information from the similarity matrices $\boldsymbol{W}_k$.

Table S5: Number of simulated datasets (out of 1000) with converged results for GEE-Ind estimator using `glm.nb()` for negative binomial responses.

	$p = 10$	$p = 25$	$p = 50$
$n = 50$	894	688	459
$n = 100$	982	933	865
$n = 200$	1000	999	998
$n = 400$	1000	1000	1000

Table S6: Number of simulated datasets (out of 1000) with converged results for GEE-Unstruc estimator using the modified iterative algorithm. Dashes indicate procedures that are not executed due to the sample correlation matrix being singular since $n = p$.

	Bernoulli			Poisson		
	$p = 10$	$p = 25$	$p = 50$	$p = 10$	$p = 25$	$p = 50$
$n = 50$	953	501	—	991	789	—
$n = 100$	1000	991	879	1000	1000	945
$n = 200$	1000	1000	1000	1000	1000	999
$n = 400$	1000	1000	1000	1000	1000	1000
	Negative Binomial			Gaussian		
	$p = 10$	$p = 25$	$p = 50$	$p = 10$	$p = 25$	$p = 50$
$n = 50$	985	880	—	1000	1000	—
$n = 100$	997	995	970	1000	1000	1000
$n = 200$	1000	998	999	1000	1000	1000
$n = 400$	1000	1000	998	1000	1000	1000

S7.4 Impact of Misspecified Similarity Measures

In this section, we conducted additional numerical studies to investigate the effect of misspecified similarity measures on the empirical performance of the proposed joint estimator. Recalling the off-diagonals of $\boldsymbol{W}_1$ used for simulating the datasets are given as $w^{(1)}_{j_1 j_2} = \exp(-|z_{j_1 1} - z_{j_2 1}|^2)$, we considered fitting the joint mean and correlation regression model to the same 1000 datasets under all combinations of n, p and response types but using a misspecified $\boldsymbol{W}_1$ matrix

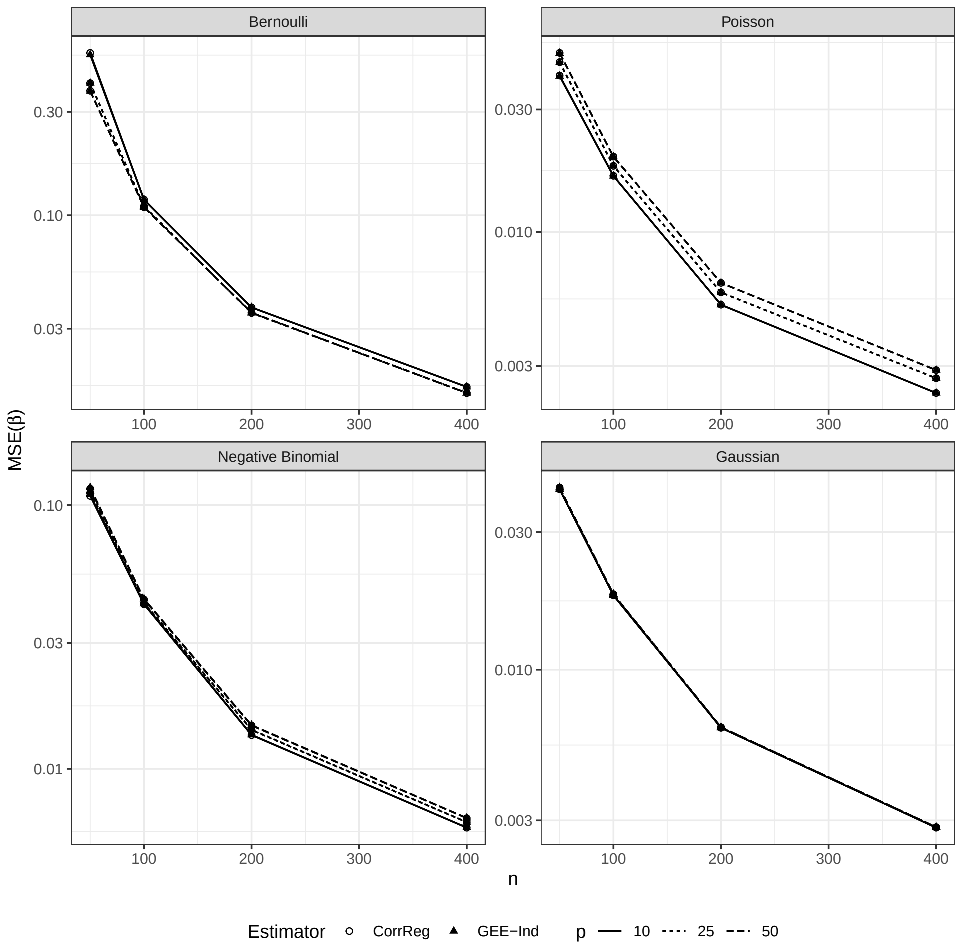

Figure S3: Averaged MSE of the mean regression coefficients for the proposed CorrReg estimator (circle) and GEE-Ind estimator (triangle).

with off-diagonals defined as $w^{(1)}_{j_1 j_2} = \exp(-|z_{j_1 1} - z_{j_2 1}|)$ while the other similarity matrices are correctly specified. This estimator is denoted as the CorrReg-Misspec estimator.

It can be seen from Figure S7 that the estimation performance of the CorrReg-Misspec estimator for the mean regression coefficients was largely unaffected by

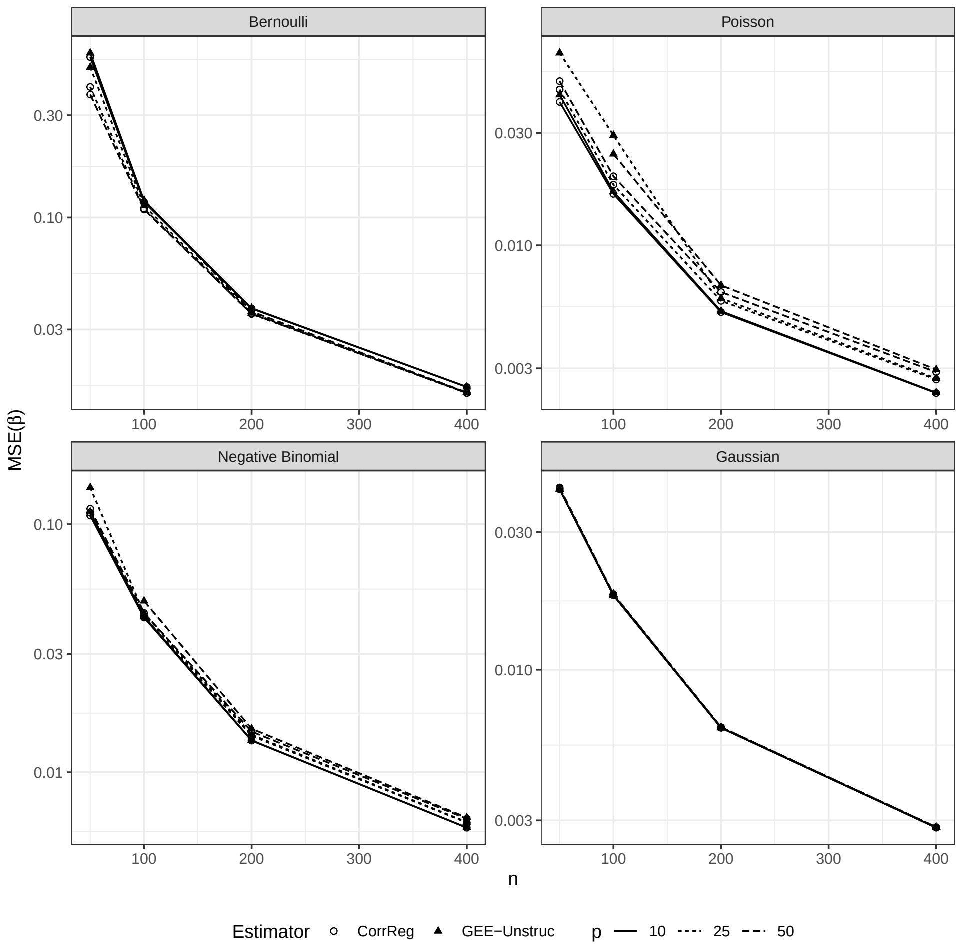

Figure S4: Averaged MSE of the mean regression coefficients for the proposed CorrReg estimator (circle) and GEE-Unstruc estimator (triangle).

the misspecified similarity measure; this again supported the robustness property of the mean regression coefficient estimator to misspecification of correlation component of the joint model. On the other hand, and as expected, Figure S8 shows that the CorrReg-Misspec estimator had worse estimation performance for the correlation regression parameters when p was large. Further investiga-

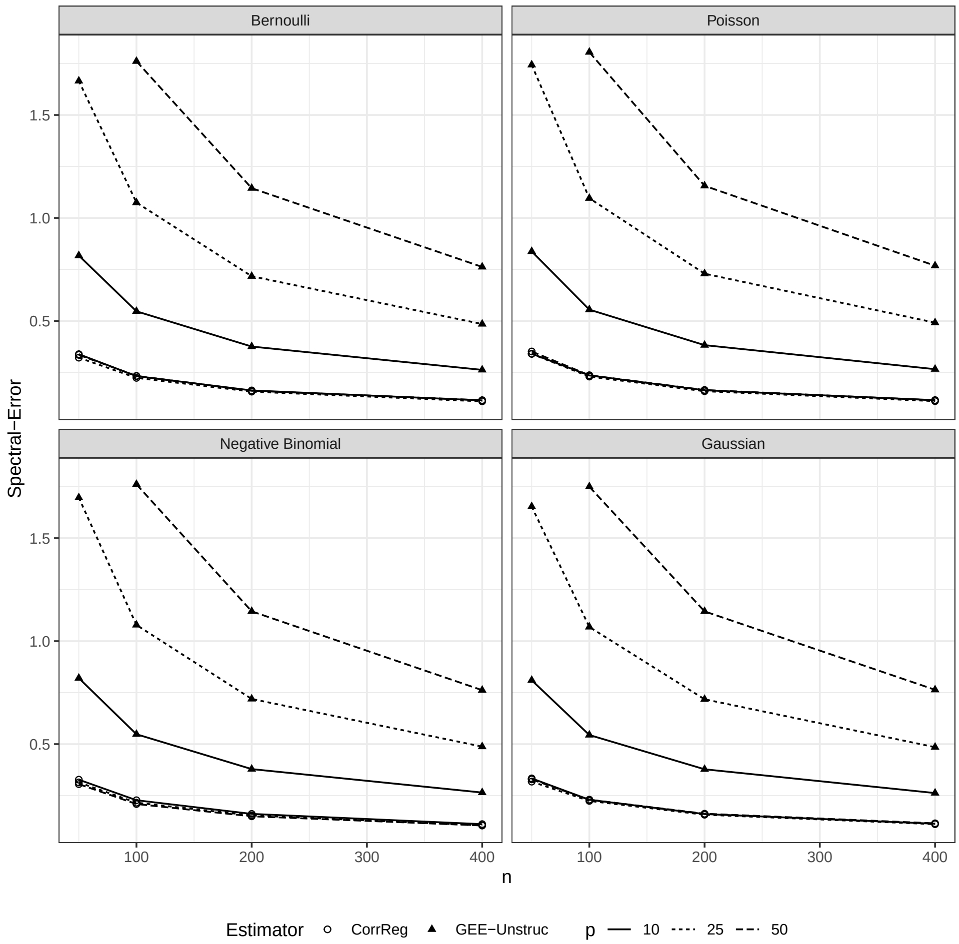

Figure S5: Averaged spectral-error for the proposed CorrReg estimator (circle) and GEE-Unstruc estimator (triangle).

tion revealed that these differences between the performances of CorrReg estimator and CorrReg-Misspec estimator of the correlation regression parameters were primarily driven by the estimation of ρ_1, i.e., the parameter associated with $\boldsymbol{W}_1$ that was misspecified for CorrReg-Misspec estimator. This also led to the poorer estimation of the correlation matrix using the CorrReg-Misspec estimator

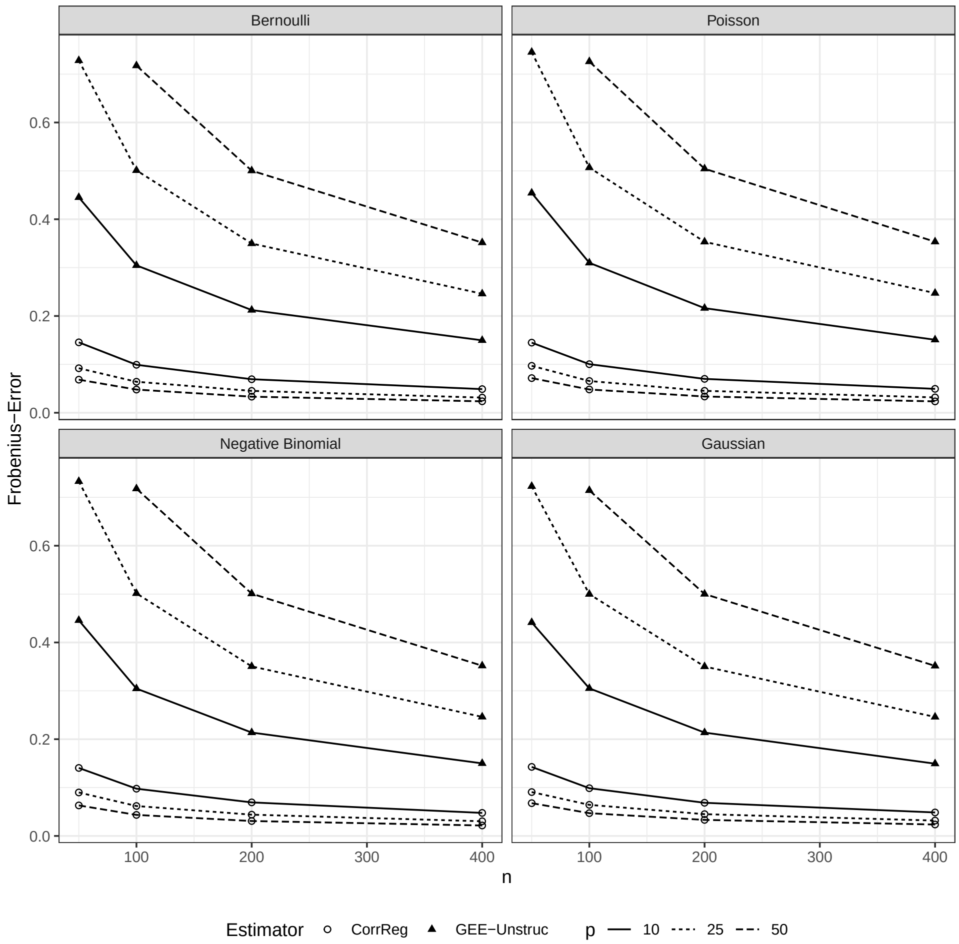

Figure S6: Averaged Frobenius-error for the proposed CorrReg estimator (circle) and GEE-Unstruc estimator (triangle).

as demonstrated in Figures S9 – S10.

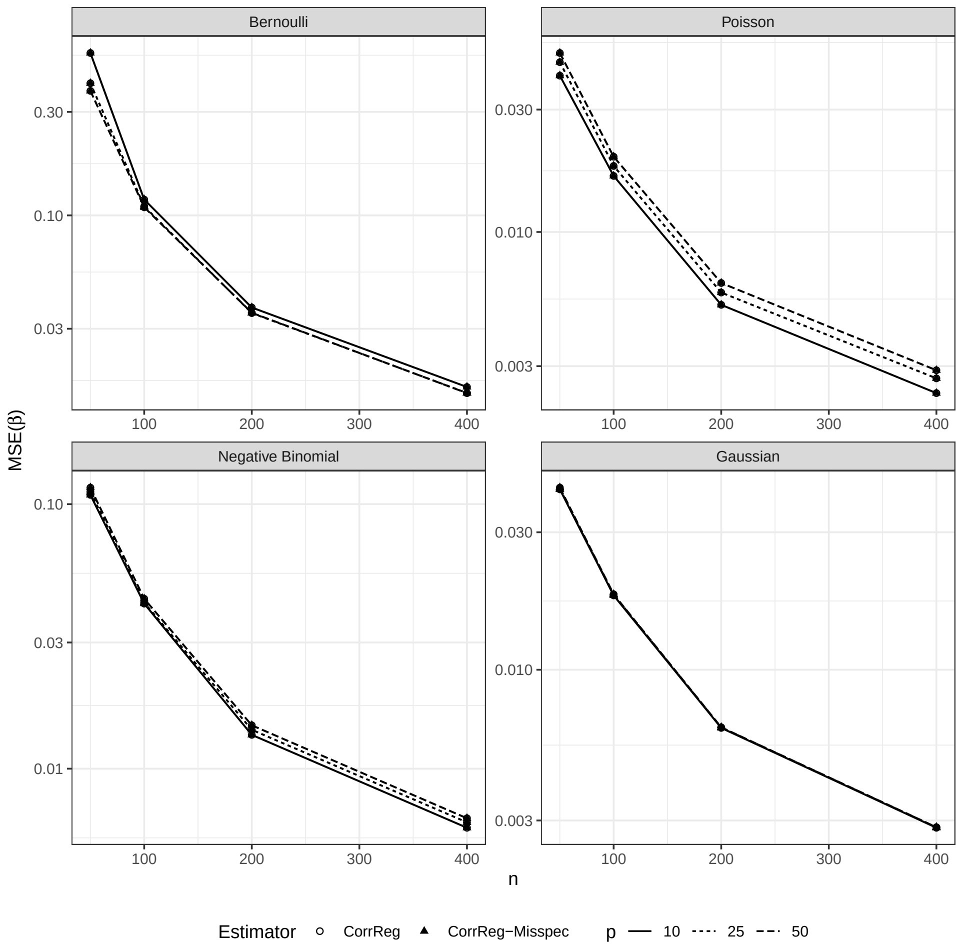

Figure S7: Averaged MSE of the mean regression coefficients for the proposed CorrReg estimator (circle) and CorrReg-Misspec estimator (triangle).

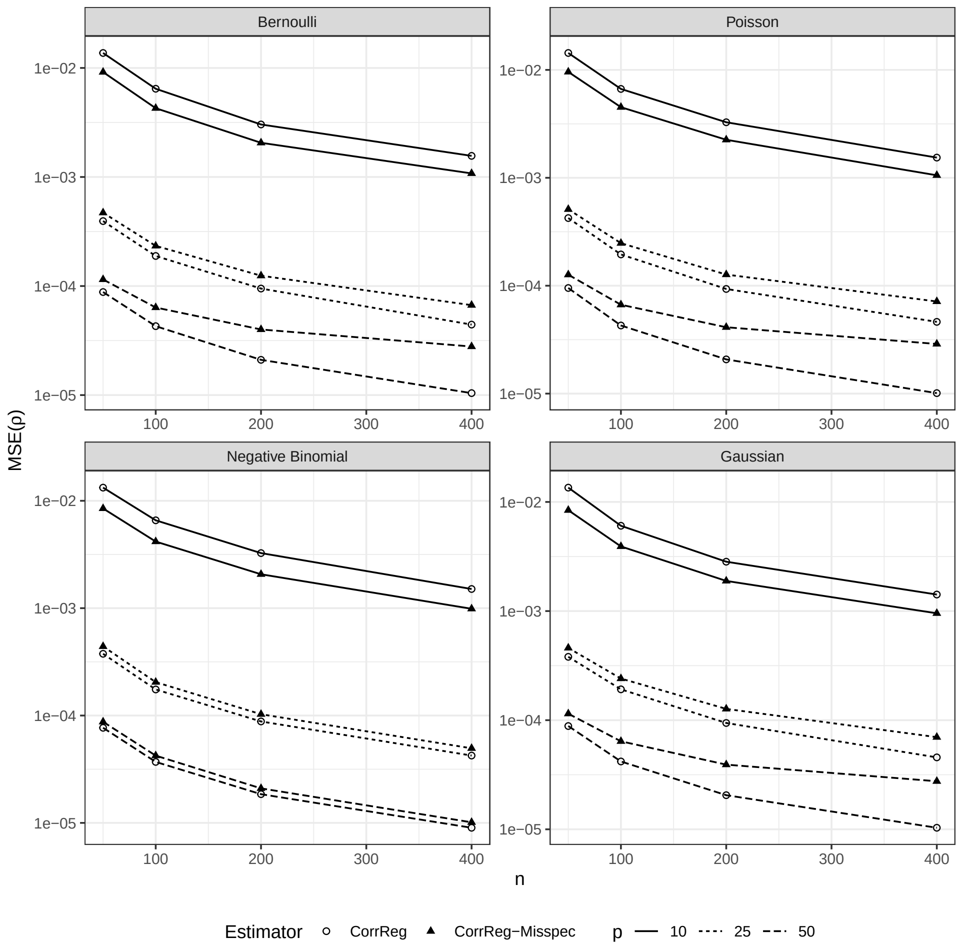

Figure S8: Averaged MSE of the correlation regression parameters for the proposed CorrReg estimator (circle) and CorrReg-Misspec estimator (triangle).

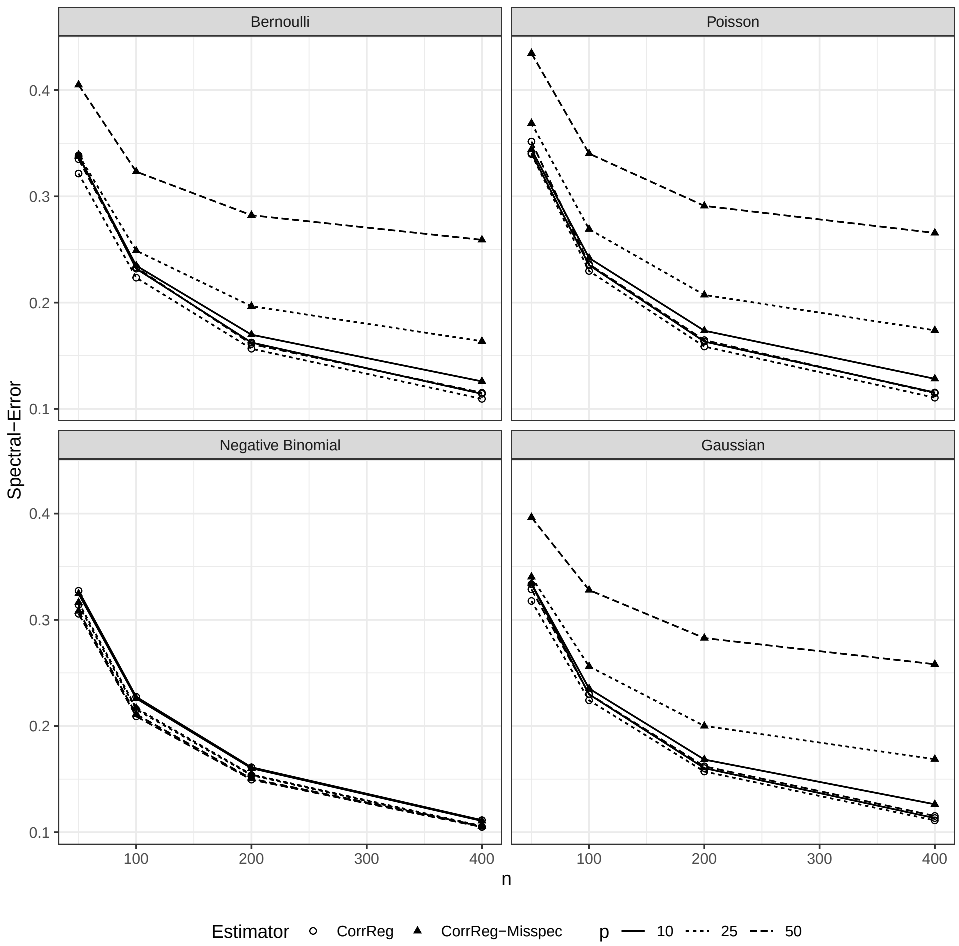

Figure S9: Averaged spectral-error for the proposed CorrReg estimator (circle) and CorrReg-Misspec estimator (triangle).

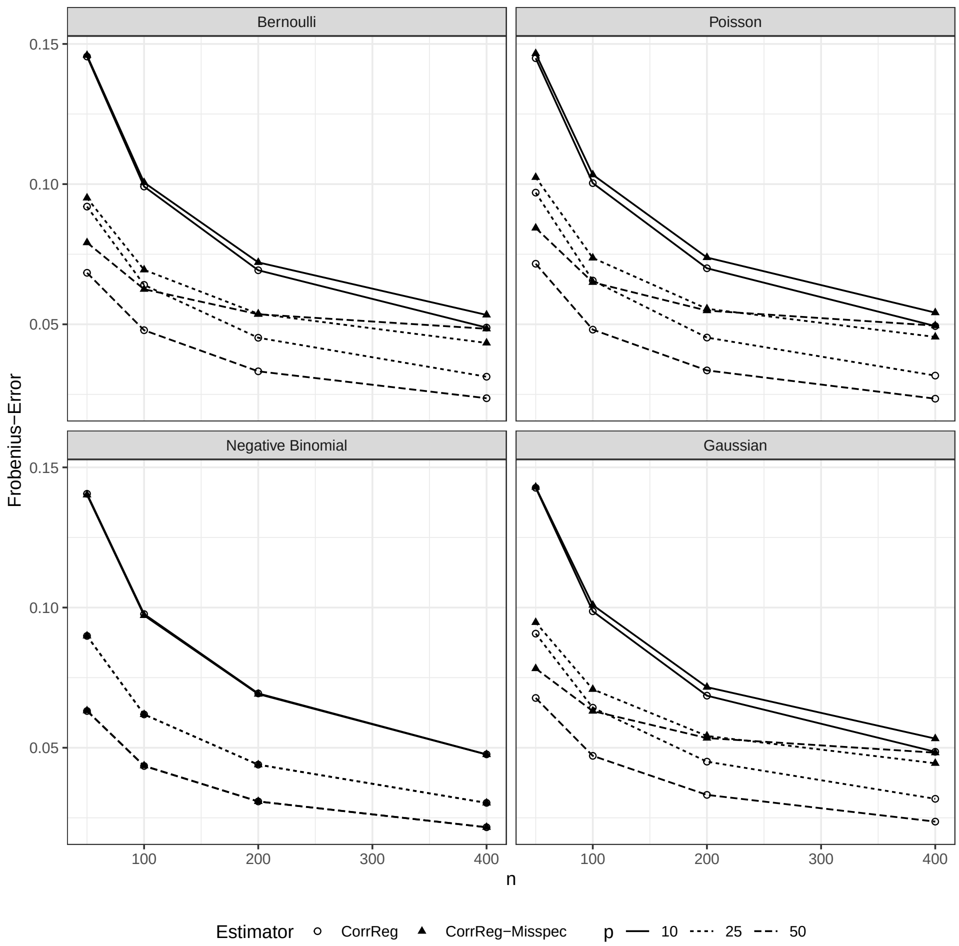

Figure S10: Averaged Frobenius-error for the proposed CorrReg estimator (circle) and CorrReg-Misspec estimator (triangle).

S8 Supplementary Details of Application to Scotland Carabidae Ground Beetle Dataset

We present additional details for the analysis of the ground beetle dataset in Section 5. Figure S11 shows most of the sample variances of species' counts were much larger than their corresponding sample means, implying the carabid beetle species counts are overdispersed. Additionally, the straight line pattern observed in Figure S11 indicates a quadratic mean-variance relationship is reasonable (Warton et al., 2012). Table S7 presents the estimated mean regression coefficients for all 38 carabid ground beetle species and the estimated correlation regression parameters for all five trait variables, which demonstrates different species had distinct preferences for the environmental conditions.

We also investigated the sensitivity of the proposed model to three alternative ways of specifying the similarity measures/correlation component of the joint model, namely, (i) by defining $w^{(1)}_{j_1 j_2} = \exp(-|z_{j_1 1} - z_{j_2 1}|)$ rather than $w^{(1)}_{j_1 j_2} = \exp(-|z_{j_1 1} - z_{j_2 1}|^2)$ for the similarity matrix $\boldsymbol{W}_1$ associated with beetle total length, (ii) only considering two trait similarity matrices with significant correlation regression parameters in Table S7, i.e., beetle total length and breeding season, rather than all five trait similarity matrices as in the main text, or (iii) dropping all five trait similarity matrices that is equivalent to fitting GEE with an independence working correlation matrix as in Section S7.3. Figure S12 indicates that the estimated values of the mean regression coefficients remained mostly unchanged, while the inference results for the mean regression coefficients were also found to be largely similar to those in Table S7, again supporting the robustness of our proposed mean regression coefficient estimators to alternative specifications of the correlation component of the joint model. Interestingly, Table S8 and Figure S13 also suggest that estimation results for the correlation regression parameters and between-species correlation matrix, as well as the inference results for the correlation regression parameters, were largely unaffected under the first two alternative choices of similarity measures in this application. This lends further support to the original results in Table S7. Finally, we note that GEE with an independence working correlation matrix could not be used to study the relationship between trait similarity matrices and the between-species correlation matrix due to its implicit independence assumption.

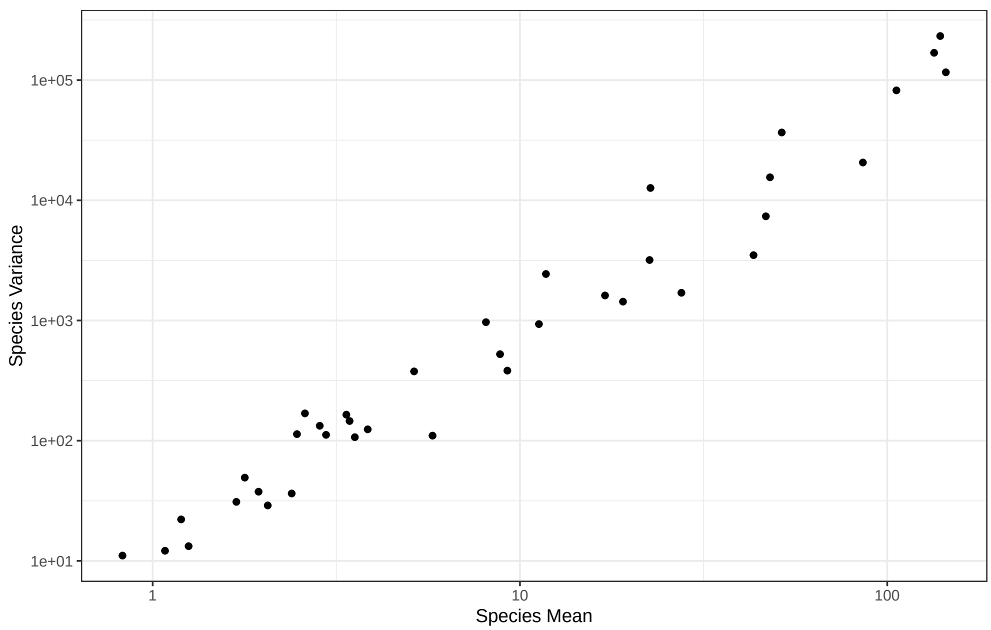

Figure S11: Log-log plot of sample variances for each of the species' counts against their corresponding sample means for the Carabidae ground beetle dataset.

Table S7: Point estimates and 95% confidence intervals (in parentheses) for mean regression coefficients of all 38 carabid beetle species, and correlation regression parameters for all five trait variables, based on analysis of the Carabidae ground beetle dataset. Estimates whose confidence interval excludes zero are bolded.

Estimation of $\boldsymbol{\beta}_j$				
Species	Intercept	Soil pH	Elevation	Land Management
A.Muelleri	**2.444** **(2.015, 2.872)**	**1.382** **(0.631, 2.132)**	**-0.730** **(-1.396, -0.064)**	-0.022 (-0.725, 0.682)
A.Apricaria	-0.697 (-1.528, 0.135)	-0.151 (-1.186, 0.883)	**-1.649** **(-2.904, -0.393)**	**1.615** **(0.692, 2.538)**
A.Bifrons	**-1.349** **(-2.462, -0.236)**	**1.415** **(0.134, 2.696)**	**-2.145** **(-3.720, -0.570)**	0.314 (-0.775, 1.404)
A.Communis	**0.604** **(0.145, 1.062)**	-0.080 (-0.876, 0.715)	-0.439 (-1.058, 0.181)	**-0.970** **(-1.741, -0.198)**
A.Familiaris	0.221 (-0.382, 0.823)	0.454 (-0.593, 1.500)	-0.371 (-1.287, 0.544)	0.474 (-0.509, 1.458)
A.Lunicollis	-0.016 (-0.533, 0.502)	-0.714 (-1.644, 0.217)	**-0.841** **(-1.600, -0.082)**	-0.542 (-1.401, 0.317)
A.Plebeja	**3.074** **(2.667, 3.480)**	**1.576** **(0.860, 2.291)**	**-0.933** **(-1.567, -0.298)**	-0.390 (-1.063, 0.283)
A.Dorsalis	0.106 (-0.513, 0.725)	0.305 (-0.544, 1.153)	-0.144 (-1.064, 0.776)	**2.557** **(1.720, 3.394)**
B.Aeneum	**2.010** **(1.265, 2.756)**	**1.467** **(0.262, 2.673)**	**-2.993** **(-4.314, -1.672)**	-0.236 (-1.345, 0.873)
B.Guttula	**1.948** **(1.560, 2.336)**	**0.921** **(0.264, 1.577)**	**-1.024** **(-1.672, -0.376)**	0.448 (-0.159, 1.055)
B.Lampros	**2.160** **(1.650, 2.670)**	0.877 (-0.011, 1.764)	**0.721** **(0.023, 1.419)**	**0.985** **(0.132, 1.837)**
B.Tetracolum	**-4.375** **(-6.499, -2.251)**	0.715 (-0.289, 1.719)	**-4.042** **(-6.283, -1.800)**	**3.321** **(2.182, 4.459)**
C.Fuscipes	**4.368** **(3.830, 4.906)**	0.513 (-0.429, 1.455)	**-0.787** **(-1.534, -0.040)**	-0.005 (-0.907, 0.896)
C.Melanocephalus	**3.655** **(3.296, 4.014)**	0.130 (-0.497, 0.757)	-0.043 (-0.535, 0.449)	0.465 (-0.137, 1.066)
C.Problematicus	0.298 (-0.183, 0.779)	**-0.679** **(-1.323, -0.035)**	0.148 (-0.294, 0.590)	**-1.886** **(-2.622, -1.150)**
C.Violaceus	**-1.112** **(-1.875, -0.349)**	-0.180 (-1.190, 0.830)	**1.556** **(0.835, 2.276)**	-0.796 (-1.944, 0.352)
C.Fossor	**2.940** **(2.614, 3.267)**	**0.649** **(0.079, 1.220)**	-0.223 (-0.681, 0.235)	0.515 (-0.029, 1.059)
C.Caraboides	**-1.856** **(-3.075, -0.636)**	-0.534 (-2.222, 1.154)	-0.590 (-1.770, 0.591)	**-2.085** **(-3.980, -0.191)**
H.Latus	**-1.436** **(-2.303, -0.569)**	-0.386 (-1.394, 0.623)	0.552 (-0.120, 1.224)	**-1.498** **(-2.737, -0.258)**

Continued on next page

Table S7 – *Continued from previous page*

Species	Intercept	Soil pH	Elevation	Land Management
H.Rufipes	-0.648 (-1.402, 0.106)	-0.260 (-1.292, 0.772)	**-1.624** **(-2.822, -0.425)**	**1.272** **(0.357, 2.187)**
L.Terminatus	**-1.388** **(-2.393, -0.384)**	**-2.070** **(-3.539, -0.600)**	**-1.204** **(-2.240, -0.168)**	**-1.593** **(-3.019, -0.168)**
L.Pilicornis	**3.615** **(3.317, 3.913)**	**0.865** **(0.344, 1.386)**	0.054 (-0.356, 0.463)	-0.048 (-0.547, 0.452)
N.Brevicollis	**4.780** **(4.120, 5.440)**	0.584 (-0.570, 1.738)	**-1.237** **(-2.162, -0.313)**	-0.433 (-1.538, 0.672)
N.Salina	**1.855** **(1.119, 2.591)**	-1.095 (-2.386, 0.196)	0.312 (-0.683, 1.306)	-0.572 (-1.803, 0.659)
N.Aquaticus	-0.423 (-1.172, 0.327)	-0.953 (-2.119, 0.213)	0.286 (-0.534, 1.106)	**-1.233** **(-2.423, -0.044)**
N.Biguttatus	**1.668** **(1.328, 2.008)**	-0.098 (-0.692, 0.496)	0.076 (-0.392, 0.543)	0.564 (-0.006, 1.134)
N.Substriatus	0.599 (-0.044, 1.243)	**1.295** **(0.184, 2.407)**	0.096 (-0.816, 1.007)	0.673 (-0.366, 1.712)
P.Assimilis	**-1.788** **(-3.086, -0.489)**	**-1.392** **(-2.519, -0.266)**	0.263 (-0.451, 0.977)	**-2.696** **(-4.333, -1.059)**
P.Diligens	-0.618 (-1.439, 0.204)	-0.789 (-1.883, 0.304)	0.173 (-0.578, 0.925)	**-1.975** **(-3.228, -0.722)**
P.Madidus	**4.507** **(3.859, 5.156)**	-0.183 (-1.317, 0.951)	**-1.314** **(-2.218, -0.411)**	-0.505 (-1.591, 0.581)
P.Melanarius	**3.615** **(3.183, 4.046)**	0.003 (-0.756, 0.762)	**-1.083** **(-1.759, -0.408)**	**1.237** **(0.522, 1.953)**
P.Nigrita	**1.079** **(0.480, 1.678)**	0.055 (-0.989, 1.098)	-0.317 (-1.130, 0.495)	**-1.100** **(-2.108, -0.093)**
P.Strenuus	**1.938** **(1.474, 2.402)**	**1.028** **(0.211, 1.845)**	**-0.780** **(-1.466, -0.094)**	**-0.880** **(-1.656, -0.104)**
P.Vernalis	-0.051 (-0.668, 0.566)	0.531 (-0.551, 1.614)	-0.585 (-1.505, 0.334)	-0.645 (-1.674, 0.384)
S.Vivalis	**0.881** **(0.274, 1.488)**	0.579 (-0.415, 1.572)	**-2.064** **(-3.129, -0.999)**	-0.083 (-0.995, 0.830)
T.Micros	0.049 (-0.858, 0.956)	-0.095 (-1.671, 1.480)	-1.024 (-2.542, 0.494)	1.048 (-0.409, 2.504)
T.Obtusus	**1.565** **(0.940, 2.191)**	**-2.483** **(-3.664, -1.301)**	**-1.381** **(-2.334, -0.427)**	0.850 (-0.202, 1.901)
T.Quadristriatus	**2.468** **(2.064, 2.873)**	-0.141 (-0.826, 0.544)	**-1.298** **(-1.988, -0.608)**	**1.911** **(1.280, 2.542)**
		Estimation of ρ_k		
Total Length	Leg Color	Wing Development	Overwintering	Breeding Season
0.061 **(0.023, 0.098)**	-0.002 (-0.022, 0.018)	0.021 (-0.001, 0.044)	0.004 (-0.018, 0.026)	**0.036** **(0.008, 0.064)**

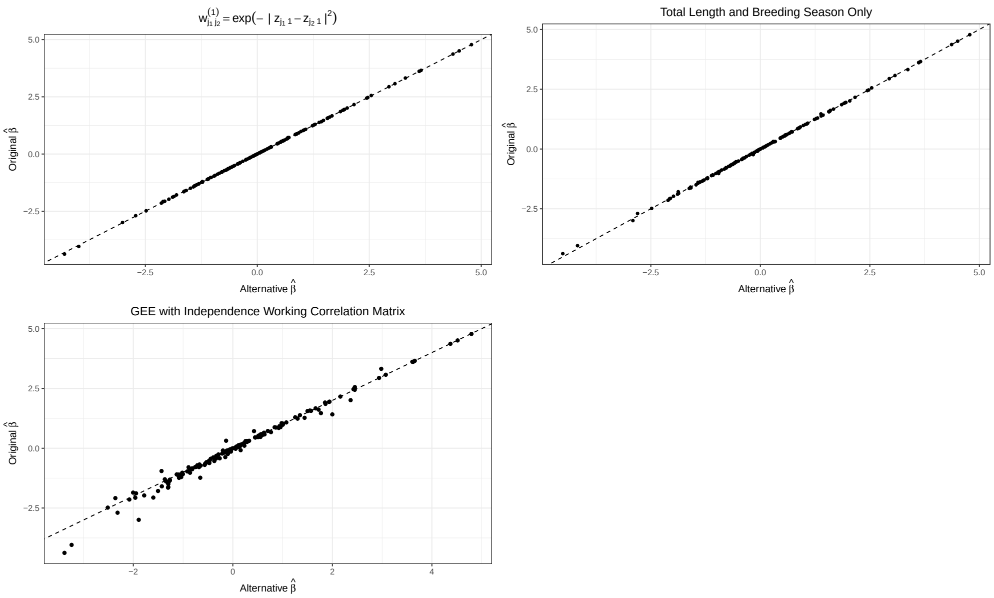

Figure S12: Scatterplots of original estimated mean regression coefficients from Table S7 against those estimated using alternative specifications of similarity measures/correlation component of the joint model and dashed lines represent the lines for $y = x$. Top left: Off-diagonals of $\boldsymbol{W}_1$ defined as $w^{(1)}_{j_1 j_2} = \exp(-|z_{j_1 1} - z_{j_2 1}|)$; Top right: Only including similarity matrices associated with beetle total length and breeding season; Bottom left: GEE with an independence working correlation matrix.

Table S8: Estimated correlation regression parameters and 95% confidence intervals (in parentheses) using different specifications of similarity measures. Estimates whose confidence interval excludes zero are bolded. First row: Original specification; Second row: Off-diagonals of $\boldsymbol{W}_1$ defined as $w^{(1)}_{j_1 j_2} = \exp(-|z_{j_1 1} - z_{j_2 1}|)$; Third row: Only including similarity matrices associated with beetle total length and breeding season.

Specification	Estimation of ρ_k				
	Total Length	Leg Color	Wing Development	Overwintering	Breeding Season
Original	**0.061**	-0.002	0.021	0.004	**0.036**
	(0.023, 0.098)	(-0.022, 0.018)	(-0.001, 0.044)	(-0.018, 0.026)	**(0.008, 0.064)**
$w^{(1)}_{j_1 j_2} = \exp(-\lvert z_{j_1 1} - z_{j_2 1}\rvert)$	**0.074**	-0.005	0.020	0.001	**0.036**
	(0.028, 0.120)	(-0.025, 0.015)	(-0.002, 0.042)	(-0.020, 0.023)	**(0.008, 0.063)**
Total length and breeding season only	**0.071**	—	—	—	**0.041**
	(0.034, 0.108)	—	—	—	**(0.014, 0.067)**

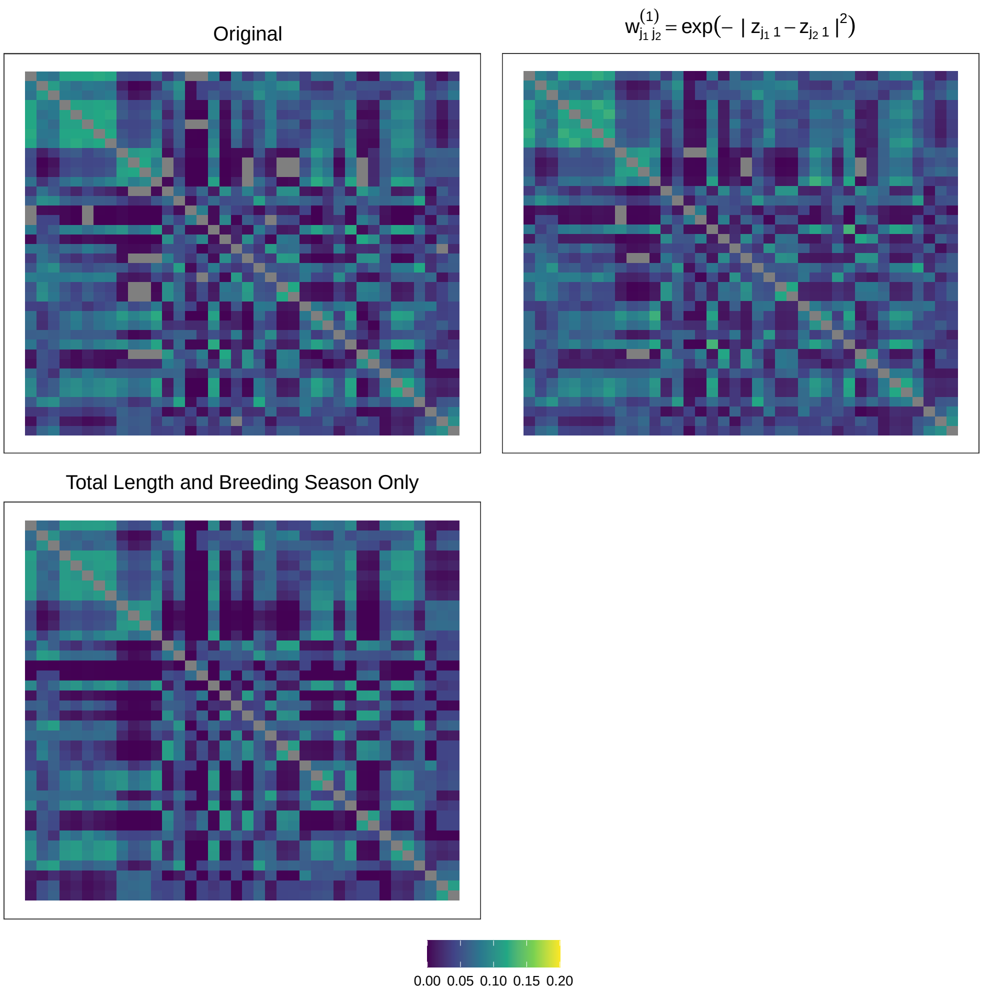

Figure S13: Heatmap of the estimated between-species correlation matrices under different specifications of similarity measures. Diagonal elements of ones and off-diagonal elements that are negative are greyed out.

References

Davidson, J. (2021). *Stochastic Limit Theory: An Introduction for Econometricians*. Oxford, U.K.: Oxford University Press.

Gao, R., G. Inan, and H. Demirtas (2019). *PoisBinOrd: Data Generation with Poisson, Binary and Ordinal Components*. R package version 1.4.1, https://CRAN.R-project.org/package=PoisBinOrd.

Hall, P. and C. Heyde (2014). *Martingale Limit Theory and Its Application*. Cambridge, Massachusetts: Academic Press.

McGuire, T. W., J. U. Farley, R. E. Lucas, and L. W. Ring (1968). Estimation and inference for linear models in which subsets of the dependent variable are constrained. *J. Am. Statist. Assoc. 63*, 1201–1213.

Seber, G. A. (2008). *A Matrix Handbook for Statisticians*. Hoboken, New Jersey: Wiley.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge, U.K.: Cambridge University Press.

Venables, W. N. and B. D. Ripley (2002). *Modern Applied Statistics with S*. New York: Springer.

Warton, D. I., S. T. Wright, and Y. Wang (2012). Distance-based multivariate analyses confound location and dispersion effects. *Methods Ecol. Evol. 3*, 89–101.

Xue, L., S. Ma, and H. Zou (2012). Positive-definite ℓ_1-penalized estimation of large covariance matrices. *J. Am. Statist. Assoc. 107*, 1480–1491.

Zou, T., W. Lan, R. Li, and C.-L. Tsai (2022). Inference on covariance-mean regression. *J. Econometrics 230*, 318–338.

Zou, T., W. Lan, H. Wang, and C.-L. Tsai (2017). Covariance regression analysis. *J. Am. Statist. Assoc. 112*, 266–281.

Zou, T., R. Luo, W. Lan, and C.-L. Tsai (2021). Network influence analysis. *Stat. Sin. 31*, 1727–1748.