# Inference on Large-scale Partially Functional Linear Model with Heterogeneous Errors

*Kaijie Xue and Fang Yao*

*Shanghai University of International Business and Economics, Peking University*

**Supplementary Material**

## S1 Notations

To be concise, we first set up basic notations used throughout the supplement as well as the main paper as follows. For any two vectors $a, b \in \mathbb{R}^p$, we write $\langle a, b\rangle = a'b$. For a vector $v = (v_1, \ldots, v_p)' \in \mathbb{R}^p$, we write its $\ell_q$-norms as $\|v\|_q = (\sum_{l=1}^p |v_l|^q)^{1/q}$ for $q \geq 1$, $\|v\|_\infty = \max_{l\leq p} |v_l|$, and $\|v\|_0$ stands for the cardinality of supp$(v)$, with supp$(v) = \{l : v_l \neq 0\}$. For a vector $v = (v_1, \ldots, v_p)' \in \mathbb{R}^p$ and a subset $S \subseteq \{1, \ldots, p\}$, we use $v_S \in \mathbb{R}^{\text{card}(S)}$ to represent the vector $v$ restricted to $S$. For a matrix $A = [a_{ij}]_{p\times q}$, we denote the elementwise $\ell_\infty$-norm as $\|A\|_\infty = \max_{i,j} |a_{ij}|$. For a symmetric matrix $A = [a_{ij}]_{p\times p}$, we write $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ to denote the minimal and maximal eigenvalues. For any two real sequences $a_n$ and $b_n$, we write $a_n \lesssim b_n$ to mean

that $a_n \leq cb_n$ for some universal constant $c > 0$, and analogously, $a_n \gtrsim b_n$ when $a_n \geq c_1 b_n$ for some universal constant $c_1 > 0$. We write $a_n \asymp b_n$ provided that $|a_n| \lesssim |b_n|$ and $|a_n| \gtrsim |b_n|$.

Note that we write $\mathcal{P}_n = \{1, \ldots, p_n\}$ to denote the index set of all functional predictors, and similarly, $\mathcal{D}_n = \{1, \ldots, d_n\}$ to represent all scalar covariates. For any nonempty subset $\mathcal{H}_n \subseteq \mathcal{P}_n$ with cardinality $|\mathcal{H}_n| = h_n \leq p_n$, we write its complement as $\mathcal{H}_n^c = \mathcal{P}_n \setminus \mathcal{H}_n$. For any nonempty subset $\mathcal{K}_n \subseteq \mathcal{D}_n$ with cardinality $|\mathcal{K}_n| = k_n \leq d_n$, we write its complement as $\mathcal{K}_n^c = \mathcal{D}_n \setminus \mathcal{K}_n$. We denote the response vector as $Y = (Y_1, \ldots, Y_n)'$. For functional predictors, we write the vector $\eta_{\mathcal{H}_n}$ as stacking $\{\eta_j : j \in \mathcal{H}_n\}$ in a column, and similarly for $\hat{\eta}_{\mathcal{H}_n}$. In addition, we abbreviate $\beta_{\mathcal{H}_n} = \{\beta_j : j \in \mathcal{H}_n\}$ as the sequence of regression functions. Given the fixed basis $\{b_k : k \geq 1\}$ and the truncation size $s_n$, we define the function $F_{\{b_k : k \leq s_n\}}(\beta_{\mathcal{H}_n}) = \eta_{\mathcal{H}_n}$ as mapping the regression curves $\beta_{\mathcal{H}_n}$ to the projection vector $\eta_{\mathcal{H}_n}$. For scalar covariates, we denote $\gamma_{\mathcal{K}_n}$ as restricting the vector $\gamma$ to $\mathcal{K}_n$, and similarly for $\hat{\gamma}_{\mathcal{K}_n}$. For functional predictors, we denote its design matrix $\Theta$ as stacking $\{\Theta_j : j \leq p_n\}$ in a row, and similarly, the matrix $\Theta_{\mathcal{H}_n}$ as stacking $\{\Theta_j : j \in \mathcal{H}_n\}$ in a row. For scalar covariates, we denote its design matrix as $Z = (Z_1, \ldots, Z_n)'$, and similarly, the matrix $Z_{\mathcal{K}_n}$ as restricting the columns of $Z$ to $\mathcal{K}_n$. Estimating the eigenvalues $\omega_{jk}$ by $\hat{\omega}_{jk} = n^{-1} \sum_{i=1}^n \theta_{ijk}^2$, we denote several block diagonal matrices as $\Lambda = diag\{\Lambda_j :$

$j \in \mathcal{P}_n\}$, $\Lambda_{\mathcal{H}_n} = diag\{\Lambda_j : j \in \mathcal{H}_n\}$, $\hat{\Lambda} = diag\{\hat{\Lambda}_j : j \in \mathcal{P}_n\}$, $\hat{\Lambda}_{\mathcal{H}_n} = diag\{\hat{\Lambda}_j : j \in \mathcal{H}_n\}$ with submatrices $\Lambda_j = diag\{\omega_{j1}, \ldots, \omega_{js_n}\}$ and $\hat{\Lambda}_j = diag\{\hat{\omega}_{j1}, \ldots, \hat{\omega}_{js_n}\}$. We express several matrices as

$$(\Theta, Z) = (G_1, \ldots, G_n)', \qquad (\tilde{\Theta}, Z) = (\Theta\Lambda^{-1/2}, Z) = (\tilde{G}_1, \ldots, \tilde{G}_n)',$$

$$(\breve{\Theta}, Z) = (\Theta\hat{\Lambda}^{-1/2}, Z) = (\breve{G}_1, \ldots, \breve{G}_n)', \qquad (\Theta_{\mathcal{H}_n}, Z_{\mathcal{K}_n}) = (E_1, \ldots, E_n)',$$

$$(\tilde{\Theta}_{\mathcal{H}_n}, Z_{\mathcal{K}_n}) = (\Theta_{\mathcal{H}_n}\Lambda_{\mathcal{H}_n}^{-1/2}, Z_{\mathcal{K}_n}) = (\tilde{E}_1, \ldots, \tilde{E}_n)',$$

$$(\breve{\Theta}_{\mathcal{H}_n}, Z_{\mathcal{K}_n}) = (\Theta_{\mathcal{H}_n}\hat{\Lambda}_{\mathcal{H}_n}^{-1/2}, Z_{\mathcal{K}_n}) = (\breve{E}_1, \ldots, \breve{E}_n)', \;\; (\Theta_{\mathcal{H}_n^c}, Z_{\mathcal{K}_n^c}) = (F_1, \ldots, F_n)',$$

$$(\tilde{\Theta}_{\mathcal{H}_n^c}, Z_{\mathcal{K}_n^c}) = (\Theta_{\mathcal{H}_n^c}\Lambda_{\mathcal{H}_n^c}^{-1/2}, Z_{\mathcal{K}_n^c}) = (\tilde{F}_1, \ldots, \tilde{F}_n)',$$

$$(\breve{\Theta}_{\mathcal{H}_n^c}, Z_{\mathcal{K}_n^c}) = (\Theta_{\mathcal{H}_n^c}\hat{\Lambda}_{\mathcal{H}_n}^{-1/2}, Z_{\mathcal{K}_n^c}) = (\breve{F}_1, \ldots, \breve{F}_n)'$$

with $G_i, \tilde{G}_i, \breve{G}_i, E_i, \tilde{E}_i, \breve{E}_i, F_i, \tilde{F}_i, \breve{F}_i$ as the transpose of row vectors. Likewise, we express several scaled-vectors as

$$\tilde{\eta} = \Lambda^{1/2}\eta, \quad \breve{\eta} = \hat{\Lambda}^{1/2}\eta, \quad \tilde{\eta}_{\mathcal{H}_n} = \Lambda_{\mathcal{H}_n}^{1/2}\eta_{\mathcal{H}_n}, \quad \breve{\eta}_{\mathcal{H}_n} = \hat{\Lambda}_{\mathcal{H}_n}^{1/2}\eta_{\mathcal{H}_n}.$$

With some abuse of notation, we sometimes denote $(\eta^*, \gamma^*)$ as the true version of $(\eta, \gamma)$, and writes their differences as $\nu = [(\eta - \eta^*)', (\gamma - \gamma^*)']'$ and $\tilde{\nu} = diag\{\Lambda^{1/2}, I_{d_n}\}\nu$. Recall that we denote the unknown matrix $w$ as

$$w = \{E(F_iF_i')\}^{-1}E(F_i\tilde{E}_i') = (w_1, \ldots, w_{h_ns_n+k_n}) \in \mathbb{R}^{(p_n-h_n)s_n+(d_n-k_n)\times(h_ns_n+k_n)},$$

whose sparsity level is measured by the parameter $\rho_n = \sup_{j\le h_ns_n+k_n} \rho_{nj}$, with each $\rho_{nj} = \|w_j\|_0$. Using the regularized estimators $(\hat{\eta}, \hat{\gamma})$ and $\hat{w}$ from (2.3) and

(3.5) of the main paper, we denote several random vectors for $i = 1, \ldots, n$ as

$$\tilde{S}_i = (\tilde{S}_{i1}, \ldots, \tilde{S}_{i,h_n s_n + k_n})' = \big(w' F_i - \tilde{E}_i\big)\big(Y_i - G_i'\big(\tbinom{\eta}{\gamma}\big)\big),$$

$$S_i = (S_{i1}, \ldots, S_{i,h_n s_n + k_n})' = \big(\hat{w}' F_i - \breve{E}_i\big)\big(Y_i - E_i'\big(\tbinom{\eta_{\mathcal{H}_n}}{\gamma_{\mathcal{K}_n}}\big) - F_i'\big(\tbinom{\hat{\eta}_{\mathcal{H}_n^c}}{\hat{\gamma}_{\mathcal{K}_n^c}}\big)\big),$$

$$\hat{S}_i = (\hat{S}_{i1}, \ldots, \hat{S}_{i,h_n s_n + k_n})' = \big(\hat{w}' F_i - \breve{E}_i\big)\big(Y_i - G_i'\big(\tbinom{\hat{\eta}}{\hat{\gamma}}\big)\big),$$

$$S_i^* = (S_{i1}^*, \ldots, S_{i,h_n s_n + k_n}^*)' = \big(w' F_i - \tilde{E}_i\big)\epsilon_i,$$

$$S(\tilde{\eta}_{\mathcal{H}_n}, \gamma_{\mathcal{K}_n}; \eta_{\mathcal{H}_n^c}, \gamma_{\mathcal{K}_n^c}) = n^{-1} \sum_{i=1}^n \tilde{S}_i, \qquad \hat{S}(\hat{\Lambda}_{\mathcal{H}_n}^{1/2} \eta_{\mathcal{H}_n}, \gamma_{\mathcal{K}_n}; \hat{\eta}_{\mathcal{H}_n^c}, \hat{\gamma}_{\mathcal{K}_n^c}) = n^{-1} \sum_{i=1}^n S_i,$$

$$\hat{S}(\hat{\Lambda}_{\mathcal{H}_n}^{1/2} \hat{\eta}_{\mathcal{H}_n}, \hat{\gamma}_{\mathcal{K}_n}; \hat{\eta}_{\mathcal{H}_n^c}, \hat{\gamma}_{\mathcal{K}_n^c}) = n^{-1} \sum_{i=1}^n \hat{S}_i, \quad \hat{T}(\beta_{\mathcal{H}_n}, \gamma_{\mathcal{K}_n}) = n^{-1/2} \sum_{i=1}^n S_i,$$

$$\hat{T}_e = n^{-1/2} \sum_{i=1}^n e_i \hat{S}_i, \quad T^* = n^{-1/2} \sum_{i=1}^n S_i^*, \quad T_e^* = n^{-1/2} \sum_{i=1}^n e_i S_i^*,$$

$$c_B(\alpha) = \inf\{t \in \mathbb{R} : P_e(\|\hat{T}_e\|_\infty \le t) \ge 1 - \alpha\}, \quad \alpha \in (0, 1),$$

where $e = (e_1, \ldots, e_n)'$ denotes a set of i.i.d. standard normals independent of the data, $P_e(\cdot)$ indicates the probability measure with respect to $e$ only, and $c_B(\alpha)$ means the $100(1-\alpha)$th percentile of $\|\hat{T}_e\|_\infty$.

In the next section, we will present a series of auxiliary lemmas used in deriving the main theorems, as well as their proofs.

## S2 Auxiliary Lemmas and Proofs

**Lemma 1.** *1) Under conditions (B1)–(B4), one has*

$$|\rho_\lambda(t_1) - \rho_\lambda(t_2)| \le \lambda L |t_1 - t_2|, \quad \textit{for any } \ t_1, t_2 \in \mathbb{R}.$$

2) *Under conditions (B1)–(B4), one has* $|\rho'_\lambda(t)| \leq \lambda L$*, for any* $t \neq 0$.

3) *Under conditions (B1)–(B5), one has*

$$\lambda L|t| \leq \rho_\lambda(t) + 2^{-1}\mu t^2, \quad \textit{for any } \ t \in \mathbb{R}.$$

4) *Under conditions (B1)–(B5), if* $P_{\lambda_n}(\eta^*, \gamma^*) - P_{\lambda_n}(\eta, \gamma) \geq 0$*, where* $(\eta^*, \gamma^*)$ *stands for the true version of* $(\eta, \gamma)$ *and* $P_{\lambda_n}(\eta, \gamma) = \sum_{j=1}^{p_n} \rho_{\lambda_n}\big(n^{-5/9}\|\Theta_j \eta_j\|_2\big) + \sum_{l=1}^{d_n} \rho_{\lambda_n}\big(n^{-5/9}(\sum_{i=1}^n Z_{il}^2)^{1/2}|\gamma_l|\big)$*, then one has*

$$\begin{aligned}
0 &\leq P_{\lambda_n}(\eta^*, \gamma^*) - P_{\lambda_n}(\eta, \gamma) \\
&\leq \lambda_n L\Big\{ \sum_{j \in \mathcal{A}_n} n^{-5/9}\|\Theta_j(\eta_j - \eta_j^*)\|_2 - \sum_{j \in \mathcal{A}_n^c} n^{-5/9}\|\Theta_j(\eta_j - \eta_j^*)\|_2 \Big\} \\
&+ \lambda_n L\Big\{ \sum_{l \in \mathcal{B}_n} n^{-5/9}(\sum_{i=1}^n Z_{il}^2)^{1/2}|\gamma_l - \gamma_l^*| - \sum_{l \in \mathcal{B}_n^c} n^{-5/9}(\sum_{i=1}^n Z_{il}^2)^{1/2}|\gamma_l - \gamma_l^*| \Big\},
\end{aligned}$$

*where the subset* $\mathcal{A}_n \subseteq \mathcal{P}_n$ *denotes the index set corresponding to the largest* $q_n$ *elements of* $\{n^{-5/9}\|\Theta_j(\eta_j - \eta_j^*)\|_2 : j \leq p_n\}$ *in magnitude, with* $\mathcal{A}_n^c = \mathcal{P}_n \setminus \mathcal{A}_n$*. Likewise, the subset* $\mathcal{B}_n \subseteq \mathcal{D}_n$ *denotes the index set corresponding to the largest* $r_n$ *elements of* $\{n^{-5/9}(\sum_{i=1}^n Z_{il}^2)^{1/2}|\gamma_l - \gamma_l^*| : l \leq d_n\}$ *in magnitude, with* $\mathcal{B}_n^c = \mathcal{D}_n \setminus \mathcal{B}_n$.

*Proof.* First of all, parts 1) to 3) are established via Lemma 4 in Loh and Wainwright (2015). To show part 4), we first define a function $f_n(t)$ as

$$f_n(t) = \begin{cases} t/\rho_{\lambda_n}(t), & \text{for } t > 0 \\ (\lambda_n L)^{-1}, & \text{for } t = 0 \end{cases}$$

which is nondecreasing in $t \in [0, \infty)$ by conditions (B1)–(B4). Thus, we have

$$\sum_{j \in \mathcal{A}_n^c} n^{-5/9} \|\Theta_j(\eta_j - \eta_j^*)\|_2 = \sum_{j \in \mathcal{A}_n^c} \rho_{\lambda_n}(n^{-5/9}\|\Theta_j(\eta_j - \eta_j^*)\|_2) f_n(n^{-5/9}\|\Theta_j(\eta_j - \eta_j^*)\|_2)$$
$$\leq f_n(\max_{j \in \mathcal{A}_n^c} n^{-5/9}\|\Theta_j(\eta_j - \eta_j^*)\|_2) \sum_{j \in \mathcal{A}_n^c} \rho_{\lambda_n}(n^{-5/9}\|\Theta_j(\eta_j - \eta_j^*)\|_2), \tag{S2.1}$$

and

$$\sum_{j \in \mathcal{A}_n} n^{-5/9} \|\Theta_j(\eta_j - \eta_j^*)\|_2 = \sum_{j \in \mathcal{A}_n} \rho_{\lambda_n}(n^{-5/9}\|\Theta_j(\eta_j - \eta_j^*)\|_2) f_n(n^{-5/9}\|\Theta_j(\eta_j - \eta_j^*)\|_2)$$
$$\geq f_n(\max_{j \in \mathcal{A}_n^c} n^{-5/9}\|\Theta_j(\eta_j - \eta_j^*)\|_2) \sum_{j \in \mathcal{A}_n} \rho_{\lambda_n}(n^{-5/9}\|\Theta_j(\eta_j - \eta_j^*)\|_2). \tag{S2.2}$$

Likewise, we have

$$\sum_{l \in \mathcal{B}_n^c} n^{-5/9} (\sum_{i=1}^n Z_{il}^2)^{1/2} |\gamma_l - \gamma_l^*| \leq f_n(\max_{l \in \mathcal{B}_n^c} n^{-5/9} (\sum_{i=1}^n Z_{il}^2)^{1/2} |\gamma_l - \gamma_l^*|)$$
$$\cdot \sum_{l \in \mathcal{B}_n^c} \rho_{\lambda_n}(n^{-5/9} (\sum_{i=1}^n Z_{il}^2)^{1/2} |\gamma_l - \gamma_l^*|), \tag{S2.3}$$

and

$$\sum_{l \in \mathcal{B}_n} n^{-5/9} (\sum_{i=1}^n Z_{il}^2)^{1/2} |\gamma_l - \gamma_l^*| \geq f_n(\max_{l \in \mathcal{B}_n^c} n^{-5/9} (\sum_{i=1}^n Z_{il}^2)^{1/2} |\gamma_l - \gamma_l^*|)$$
$$\cdot \sum_{l \in \mathcal{B}_n} \rho_{\lambda_n}(n^{-5/9} (\sum_{i=1}^n Z_{il}^2)^{1/2} |\gamma_l - \gamma_l^*|). \tag{S2.4}$$

By combining (S2.1), (S2.2), (S2.3), (S2.4) with $P_{\lambda_n}(\eta^*, \gamma^*) - P_{\lambda_n}(\eta, \gamma) \geq 0$,

we have

$$
\begin{aligned}
0 \le P_{\lambda_n}(\eta^*, \gamma^*) - P_{\lambda_n}(\eta, \gamma) &= \Big\{ \sum_{j=1}^{q_n} \rho_{\lambda_n}(n^{-5/9}\|\Theta_j \eta_j^*\|_2) - \sum_{j=1}^{p_n} \rho_{\lambda_n}(n^{-5/9}\|\Theta_j \eta_j\|_2) \Big\} \\
&+ \Big\{ \sum_{l=1}^{r_n} \rho_{\lambda_n}\big(n^{-5/9}(\sum_{i=1}^{n} Z_{il}^2)^{1/2}|\gamma_l^*|\big) - \sum_{l=1}^{d_n} \rho_{\lambda_n}\big(n^{-5/9}(\sum_{i=1}^{n} Z_{il}^2)^{1/2}|\gamma_l|\big) \Big\} \\
&\le \Big\{ \sum_{j=1}^{q_n} \rho_{\lambda_n}(n^{-5/9}\|\Theta_j(\eta_j - \eta_j^*)\|_2) - \sum_{j=q_n+1}^{p_n} \rho_{\lambda_n}(n^{-5/9}\|\Theta_j \eta_j\|_2) \Big\} \\
&+ \Big\{ \sum_{l=1}^{r_n} \rho_{\lambda_n}\big(n^{-5/9}(\sum_{i=1}^{n} Z_{il}^2)^{1/2}|\gamma_l - \gamma_l^*|\big) - \sum_{l=r_n+1}^{d_n} \rho_{\lambda_n}\big(n^{-5/9}(\sum_{i=1}^{n} Z_{il}^2)^{1/2}|\gamma_l|\big) \Big\} \\
&\le \Big\{ \sum_{j\in\mathcal{A}_n} \rho_{\lambda_n}(n^{-5/9}\|\Theta_j(\eta_j - \eta_j^*)\|_2) - \sum_{j\in\mathcal{A}_n^c} \rho_{\lambda_n}(n^{-5/9}\|\Theta_j(\eta_j - \eta_j^*)\|_2) \Big\} \\
&+ \Big\{ \sum_{l\in\mathcal{B}_n} \rho_{\lambda_n}\big(n^{-5/9}(\sum_{i=1}^{n} Z_{il}^2)^{1/2}|\gamma_l - \gamma_l^*|\big) - \sum_{l\in\mathcal{B}_n^c} \rho_{\lambda_n}\big(n^{-5/9}(\sum_{i=1}^{n} Z_{il}^2)^{1/2}|\gamma_l - \gamma_l^*|\big) \Big\} \\
&\le \big\{ f_n(\max_{j\in\mathcal{A}_n^c} n^{-5/9}\|\Theta_j(\eta_j - \eta_j^*)\|_2) \big\}^{-1} \Big\{ \sum_{j\in\mathcal{A}_n} n^{-5/9}\|\Theta_j(\eta_j - \eta_j^*)\|_2 - \sum_{j\in\mathcal{A}_n^c} n^{-5/9}\|\Theta_j(\eta_j - \eta_j^*)\|_2 \Big\} \\
&+ \big\{ f_n(\max_{l\in\mathcal{B}_n^c} n^{-5/9}(\sum_{i=1}^{n} Z_{il}^2)^{1/2}|\gamma_l - \gamma_l^*|) \big\}^{-1} \Big\{ \sum_{l\in\mathcal{B}_n} n^{-5/9}(\sum_{i=1}^{n} Z_{il}^2)^{1/2}|\gamma_l - \gamma_l^*| - \\
&\qquad \sum_{l\in\mathcal{B}_n^c} n^{-5/9}(\sum_{i=1}^{n} Z_{il}^2)^{1/2}|\gamma_l - \gamma_l^*| \Big\} \\
&\le \lambda_n L \Big\{ \sum_{j\in\mathcal{A}_n} n^{-5/9}\|\Theta_j(\eta_j - \eta_j^*)\|_2 - \sum_{j\in\mathcal{A}_n^c} n^{-5/9}\|\Theta_j(\eta_j - \eta_j^*)\|_2 \Big\} \\
&+ \lambda_n L \Big\{ \sum_{l\in\mathcal{B}_n} n^{-5/9}(\sum_{i=1}^{n} Z_{il}^2)^{1/2}|\gamma_l - \gamma_l^*| - \sum_{l\in\mathcal{B}_n^c} n^{-5/9}(\sum_{i=1}^{n} Z_{il}^2)^{1/2}|\gamma_l - \gamma_l^*| \Big\},
\end{aligned}
$$

which completes the proof. □

**Lemma 2.** *Under conditions (A1.1), (A2.1) and (A3.1), denoting $I$ as the identity matrix and $\epsilon = (\epsilon_1, \ldots, \epsilon_n)'$ as the error vector, we have that with probability*

*tending to 1:*

1) $\hat{\Lambda}$ *is positive definite.*

2) $\|\hat{\Lambda}\Lambda^{-1} - I\|_\infty \leq c_1\{\log(np_ns_n)/n\}^{1/2}$, *for some universal constant* $c_1 > 0$.

3) $\|\Lambda\hat{\Lambda}^{-1} - I\|_\infty \leq c_2\{\log(np_ns_n)/n\}^{1/2}$, *for some universal constant* $c_2 > 0$.

4) $\|\hat{\Lambda}^{1/2}\Lambda^{-1/2} - I\|_\infty \leq c_3\{\log(np_ns_n)/n\}^{1/2}$, *for some universal constant* $c_3 > 0$.

5) $\|\Lambda^{1/2}\hat{\Lambda}^{-1/2} - I\|_\infty \leq c_4\{\log(np_ns_n)/n\}^{1/2}$, *for some universal constant* $c_4 > 0$.

6) $\|n^{-1}(\Theta\Lambda^{-1/2}, Z)'(\Theta\Lambda^{-1/2}, Z) - E(\tilde{G}_i\tilde{G}_i')\|_\infty \leq c_5[\log\{n(p_ns_n + d_n)\}/n]^{1/2}$, *for some universal constant* $c_5 > 0$.

7) $\|n^{-1}(\Theta, Z)'(\Theta, Z) - E(G_iG_i')\|_\infty \leq c_6[\log\{n(p_ns_n + d_n)\}/n]^{1/2}$, *for some universal constant* $c_6 > 0$.

8) $\|n^{-1}(\Theta\hat{\Lambda}^{-1/2}, Z)'(\Theta\hat{\Lambda}^{-1/2}, Z) - E(\tilde{G}_i\tilde{G}_i')\|_\infty \leq c_7[\log\{n(p_ns_n + d_n)\}/n]^{1/2}$, *for some universal constant* $c_7 > 0$.

9) $\|n^{-1}(\Theta\Lambda^{-1/2}, Z)'\epsilon\|_\infty \leq c_8[\log\{n(p_ns_n + d_n)\}/n]^{1/2}$, *for some universal constant* $c_8 > 0$.

10) $\|n^{-1}(\Theta, Z)'\epsilon\|_\infty \leq c_9[\log\{n(p_ns_n + d_n)\}/n]^{1/2}$, *for some universal constant* $c_9 > 0$.

*11)* $\|n^{-1}(\Theta\hat{\Lambda}^{-1/2}, Z)'\epsilon\|_\infty \leq c_{10}[\log\{n(p_n s_n + d_n)\}/n]^{1/2}$, *for some universal constant* $c_{10} > 0$.

*12)* $\max_{j\leq h_n s_n+k_n} \|n^{-1}\sum_{i=1}^n \tilde{F}_i(\tilde{E}_{ij}-F_i' w_j)\|_\infty \leq c_{11}[\log\{n(p_n s_n+d_n)\}/n]^{1/2}$, *for some universal constant* $c_{11} > 0$.

*13)* $\max_{j\leq h_n s_n+k_n} \|n^{-1}\sum_{i=1}^n \tilde{F}_i(\breve{E}_{ij}-F_i' w_j)\|_\infty \leq c_{12}[\log\{n(p_n s_n+d_n)\}/n]^{1/2}$, *for some universal constant* $c_{12} > 0$.

*14)* $\max_{j\leq h_n s_n+k_n} \|n^{-1}\sum_{i=1}^n \breve{F}_i(\breve{E}_{ij}-F_i' w_j)\|_\infty \leq c_{13}[\log\{n(p_n s_n+d_n)\}/n]^{1/2}$, *for some universal constant* $c_{13} > 0$.

*15)* $\|n^{-1}\sum_{i=1}^n \tilde{G}_i\|_\infty \leq c_{14}[\log\{n(p_n s_n + d_n)\}/n]^{1/2}$, *for some universal constant* $c_{14} > 0$.

*16)* $\|n^{-1}\sum_{i=1}^n G_i\|_\infty \leq c_{15}[\log\{n(p_n s_n + d_n)\}/n]^{1/2}$, *for some universal constant* $c_{15} > 0$.

*17)* $\|n^{-1}\sum_{i=1}^n \breve{G}_i\|_\infty \leq c_{16}[\log\{n(p_n s_n + d_n)\}/n]^{1/2}$, *for some universal constant* $c_{16} > 0$.

*18)* $\max_{j\leq p_n s_n+d_n} \max_{i\leq n} |\tilde{G}_{ij}| \leq c_{17}[\log\{n(p_n s_n + d_n)\}]^{1/2}$, *for some universal constant* $c_{17} > 0$.

*19)* $\max_{j\leq h_n s_n+k_n} \max_{i\leq n} |w_j' F_i| \leq c_{18}[\log\{n(p_n s_n + d_n)\}]^{1/2}$, *for some universal constant* $c_{18} > 0$.

*20)* $\max_{i\le n}|\epsilon_i| \le c_{19}\{\log(n)\}^{1/2}$*, for some universal constant* $c_{19}>0$.

*Proof.* First of all, note that for any $t>0$,

$$P\big(\|\hat\Lambda\Lambda^{-1}-I\|_\infty \ge t\big) = P\big\{\max_{j\le p_n}\max_{k\le s_n}|n^{-1}\sum_{i=1}^n(\omega_{jk}^{-1}\theta_{ijk}^2-1)|\ge t\big\}$$
$$\le \sum_{j=1}^{p_n}\sum_{k=1}^{s_n} P\big\{|n^{-1}\sum_{i=1}^n(\omega_{jk}^{-1}\theta_{ijk}^2-1)|\ge t\big\}$$
$$\le 2p_ns_n\exp\{-n\min(c_1^{-2}t^2, c_1^{-1}t)\}, \tag{S2.5}$$

for some universal constant $c_1>0$, where the first inequality is by union bound inequality and the second inequality holds from (A1.1) and Bernstein's inequality. Plugging $t=c_1\{\log(np_ns_n)/n\}^{1/2}$ into (S2.5) yields

$$P\big[\|\hat\Lambda\Lambda^{-1}-I\|_\infty \le c_1\{\log(np_ns_n)/n\}^{1/2}\big] \ge 1-2n^{-1}\to 1, \tag{S2.6}$$

which completes the proof of part 2). To show part 1), notice that

$$\lambda_{\min}(\hat\Lambda) = \lambda_{\min}(\Lambda\hat\Lambda\Lambda^{-1}) \ge \lambda_{\min}(\Lambda)\lambda_{\min}(\hat\Lambda\Lambda^{-1}) = \lambda_{\min}(\Lambda)\lambda_{\min}(\hat\Lambda\Lambda^{-1}-I+I)$$
$$\ge (1-\|\hat\Lambda\Lambda^{-1}-I\|_\infty)\lambda_{\min}(\Lambda). \tag{S2.7}$$

By combining (S2.6), (S2.7), (A3.1) with $\lambda_{\min}(\Lambda)>0$, it can be deduced that

$$P\{\lambda_{\min}(\hat\Lambda)>0\} \ge 1-2n^{-1}\to 1, \tag{S2.8}$$

which completes the proof of part 1). To show part 3), note that on the event

$\{\lambda_{\min}(\hat{\Lambda}) > 0\} \cap \{\|\hat{\Lambda}\Lambda^{-1} - I\|_\infty \leq c_1\{\log(np_ns_n)/n\}^{1/2}\}$, we have

$$\begin{aligned}
&\|\Lambda\hat{\Lambda}^{-1} - I\|_\infty = \|(\hat{\Lambda}\Lambda^{-1} - I + I)^{-1}(\hat{\Lambda}\Lambda^{-1} - I)\|_\infty \\
\leq&(1 - \|\hat{\Lambda}\Lambda^{-1} - I\|_\infty)^{-1}\|\hat{\Lambda}\Lambda^{-1} - I\|_\infty \leq 2\|\hat{\Lambda}\Lambda^{-1} - I\|_\infty \\
\leq&2c_1\{\log(np_ns_n)/n\}^{1/2}.
\end{aligned}$$

Together with (S2.6) and (S2.8), it is apparent that

$$P\big[\|\Lambda\hat{\Lambda}^{-1} - I\|_\infty \leq 2c_1\{\log(np_ns_n)/n\}^{1/2}\big] \geq 1 - 4n^{-1} \to 1, \tag{S2.9}$$

which completes the proof of part 3). To show part 4), note that

$$\|\hat{\Lambda}^{1/2}\Lambda^{-1/2} - I\|_\infty \leq \|(\hat{\Lambda}^{1/2}\Lambda^{-1/2} - I)(\hat{\Lambda}^{1/2}\Lambda^{-1/2} + I)\|_\infty = \|\hat{\Lambda}\Lambda^{-1} - I\|_\infty.$$

Together with (S2.6), it is clear that

$$P\big[\|\hat{\Lambda}^{1/2}\Lambda^{-1/2} - I\|_\infty \leq c_1\{\log(np_ns_n)/n\}^{1/2}\big] \geq 1 - 2n^{-1} \to 1, \tag{S2.10}$$

which completes the proof of part 4). To show part 5), note that on the event $\{\lambda_{\min}(\hat{\Lambda}) > 0\} \cap \{\|\Lambda\hat{\Lambda}^{-1} - I\|_\infty \leq 2c_1\{\log(np_ns_n)/n\}^{1/2}\}$, we have

$$\|\Lambda^{1/2}\hat{\Lambda}^{-1/2} - I\|_\infty \leq \|\Lambda\hat{\Lambda}^{-1} - I\|_\infty \leq 2c_1\{\log(np_ns_n)/n\}^{1/2}.$$

Together with (S2.8) and (S2.9), it is obvious that

$$P\big[\|\Lambda^{1/2}\hat{\Lambda}^{-1/2} - I\|_\infty \leq 2c_1\{\log(np_ns_n)/n\}^{1/2}\big] \geq 1 - 6n^{-1} \to 1, \tag{S2.11}$$

which completes the proof of part 5). To show part 6), note that for any $t > 0$,

$$
\begin{aligned}
&P\big(\|n^{-1}(\Theta\Lambda^{-1/2}, Z)'(\Theta\Lambda^{-1/2}, Z) - E(\tilde{G}_i\tilde{G}_i')\|_\infty \geq t\big) \\
=&P\big[\max_{l_1 \leq p_n s_n + d_n} \max_{l_2 \leq p_n s_n + d_n} |n^{-1}\sum_{i=1}^n \{\tilde{G}_{il_1}\tilde{G}_{il_2} - E(\tilde{G}_{il_1}\tilde{G}_{il_2})\}| \geq t\big] \\
\leq& \sum_{l_1=1}^{p_n s_n + d_n} \sum_{l_2=1}^{p_n s_n + d_n} P\big[|n^{-1}\sum_{i=1}^n \{\tilde{G}_{il_1}\tilde{G}_{il_2} - E(\tilde{G}_{il_1}\tilde{G}_{il_2})\}| \geq t\big] \\
\leq& 2(p_n s_n + d_n)^2 \exp\{-n\min(c_2^{-2}t^2, c_2^{-1}t)\},
\end{aligned} \tag{S2.12}
$$

for some universal constant $c_2 > 0$, where the last inequality follows from (A1.1) and Bernstein's inequality. Plugging $t = 2c_2[\log\{n(p_n s_n + d_n)\}/n]^{1/2}$ into (S2.12) yields

$$
\begin{aligned}
&P\Big(\|n^{-1}(\Theta\Lambda^{-1/2}, Z)'(\Theta\Lambda^{-1/2}, Z) - E(\tilde{G}_i\tilde{G}_i')\|_\infty \leq 2c_2[\log\{n(p_n s_n + d_n)\}/n]^{1/2}\Big) \\
\geq& 1 - 2(p_n s_n + d_n)^{-2}n^{-4} \to 1,
\end{aligned} \tag{S2.13}
$$

which completes the proof of part 6). To show part 7), note that

$$
\begin{aligned}
&\|n^{-1}(\Theta, Z)'(\Theta, Z) - E(G_i G_i{}')\|_\infty \\
\leq& \{1 + \lambda_{\max}(\Lambda)\}\|n^{-1}(\Theta\Lambda^{-1/2}, Z)'(\Theta\Lambda^{-1/2}, Z) - E(\tilde{G}_i\tilde{G}_i')\|_\infty \\
\leq& c_3\|n^{-1}(\Theta\Lambda^{-1/2}, Z)'(\Theta\Lambda^{-1/2}, Z) - E(\tilde{G}_i\tilde{G}_i')\|_\infty,
\end{aligned}
$$

for some universal constant $c_3 > 0$, where the last inequality is based on (A2.1).

Together with (S2.13), it is obvious that

$$P\Big(\|n^{-1}(\Theta,Z)'(\Theta,Z)-E(G_iG_i{}')\|_\infty \le 2c_2c_3[\log\{n(p_ns_n+d_n)\}/n]^{1/2}\Big)$$
$$\ge 1-2(p_ns_n+d_n)^{-2}n^{-4}\to 1, \tag{S2.14}$$

which completes the proof of part 7). To show part 8), note that

$$\|n^{-1}(\Theta\hat\Lambda^{-1/2},Z)'(\Theta\hat\Lambda^{-1/2},Z)-E(\tilde G_i\tilde G_i')\|_\infty = \|n^{-1}\sum_{i=1}^n \breve G_i\breve G_i'-E(\tilde G_i\tilde G_i')\|_\infty$$
$$=\|diag\{\hat\Lambda^{-1/2}\Lambda^{1/2},I_{d_n}\}\{n^{-1}\sum_{i=1}^n\tilde G_i\tilde G_i'-E(\tilde G_i\tilde G_i')\}diag\{\hat\Lambda^{-1/2}\Lambda^{1/2},I_{d_n}\}$$
$$+diag\{\hat\Lambda^{-1/2}\Lambda^{1/2},I_{d_n}\}\{E(\tilde G_i\tilde G_i')\}diag\{\hat\Lambda^{-1/2}\Lambda^{1/2}-I,0_{d_n\times d_n}\}$$
$$+diag\{\hat\Lambda^{-1/2}\Lambda^{1/2}-I,0_{d_n\times d_n}\}\{E(\tilde G_i\tilde G_i')\}\|_\infty$$
$$\le\|diag\{\hat\Lambda^{-1/2}\Lambda^{1/2},I_{d_n}\}\{n^{-1}\sum_{i=1}^n\tilde G_i\tilde G_i'-E(\tilde G_i\tilde G_i')\}diag\{\hat\Lambda^{-1/2}\Lambda^{1/2},I_{d_n}\}\|_\infty$$
$$+\|diag\{\hat\Lambda^{-1/2}\Lambda^{1/2},I_{d_n}\}\{E(\tilde G_i\tilde G_i')\}diag\{\hat\Lambda^{-1/2}\Lambda^{1/2}-I,0_{d_n\times d_n}\}\|_\infty$$
$$+\|diag\{\hat\Lambda^{-1/2}\Lambda^{1/2}-I,0_{d_n\times d_n}\}\{E(\tilde G_i\tilde G_i')\}\|_\infty$$
$$\le(2+\|\hat\Lambda^{-1}\Lambda-I\|_\infty)\|n^{-1}\sum_{i=1}^n\tilde G_i\tilde G_i'-E(\tilde G_i\tilde G_i')\|_\infty+\|\hat\Lambda^{-1/2}\Lambda^{1/2}-I\|_\infty\|E(\tilde G_i\tilde G_i')\|_\infty$$
$$+(2+\|\hat\Lambda^{-1/2}\Lambda^{1/2}-I\|_\infty)\|\hat\Lambda^{-1/2}\Lambda^{1/2}-I\|_\infty\|E(\tilde G_i\tilde G_i')\|_\infty. \tag{S2.15}$$

By combining parts 1–6), (A1.1), and (A3.1) with (S2.15), we have that with probability tending to 1:

$$\|n^{-1}(\Theta\hat\Lambda^{-1/2},Z)'(\Theta\hat\Lambda^{-1/2},Z)-E(\tilde G_i\tilde G_i')\|_\infty \le c_4[\log\{n(p_ns_n+d_n)\}/n]^{1/2},$$

for some universal constant $c_4 > 0$, which completes the proof of part 8). In a similar fashion to the proofs of parts 6–8), we can show parts 9–11). To show part 12), note that for any $t > 0$,

$$
\begin{aligned}
&P\big\{\max_{j\le h_n s_n+k_n}\|n^{-1}\sum_{i=1}^{n}\tilde{F}_i(\tilde{E}_{ij}-F_i'w_j)\|_\infty \ge t\big\}\\
\le&\sum_{j=1}^{h_n s_n+k_n}\sum_{l=1}^{(p_n-h_n)s_n+(d_n-k_n)} P\big\{|n^{-1}\sum_{i=1}^{n}\tilde{F}_{il}(\tilde{E}_{ij}-F_i'w_j)|\ge t\big\}\\
\le&2(h_n s_n+k_n)\{(p_n-h_n)s_n+(d_n-k_n)\}\exp\{-n\min(c_5^{-2}t^2,c_5^{-1}t)\}\\
\le&2(p_n s_n+d_n)^2\exp\{-n\min(c_5^{-2}t^2,c_5^{-1}t)\},
\end{aligned}
\tag{S2.16}
$$

for some universal constant $c_5 > 0$, where the first inequality is by union bound inequality and the second inequality holds from (A1.1) and Bernstein's inequality. Plugging $t = 2c_5[\log\{n(p_n s_n + d_n)\}/n]^{1/2}$ into (S2.16) yields

$$
\begin{aligned}
&P\big(\max_{j\le h_n s_n+k_n}\|n^{-1}\sum_{i=1}^{n}\tilde{F}_i(\tilde{E}_{ij}-F_i'w_j)\|_\infty \le 2c_5[\log\{n(p_n s_n+d_n)\}/n]^{1/2}\big)\\
\ge&1-2(p_n s_n+d_n)^{-2}n^{-4}\to 1,
\end{aligned}
$$

which completes the proof of part 12). To show part 13), note that

$$\begin{aligned}
&\max_{j\le h_n s_n+k_n}\|n^{-1}\sum_{i=1}^n \tilde{F}_i(\breve{E}_{ij}-F_i'w_j)\|_\infty\\
\le&\max_{j\le h_n s_n+k_n}\|n^{-1}\sum_{i=1}^n \tilde{F}_i(\tilde{E}_{ij}-F_i'w_j)\|_\infty+\max_{j\le h_n s_n+k_n}\|n^{-1}\sum_{i=1}^n \tilde{F}_i(\breve{E}_{ij}-\tilde{E}_{ij})\|_\infty\\
\le&\max_{j\le h_n s_n+k_n}\|n^{-1}\sum_{i=1}^n \tilde{F}_i(\tilde{E}_{ij}-F_i'w_j)\|_\infty+\|\hat{\Lambda}^{-1/2}\Lambda^{1/2}-I\|_\infty\max_{j\le h_n s_n+k_n}\|n^{-1}\sum_{i=1}^n \tilde{F}_i\tilde{E}_{ij}\|_\infty\\
\le&\max_{j\le h_n s_n+k_n}\|n^{-1}\sum_{i=1}^n \tilde{F}_i(\tilde{E}_{ij}-F_i'w_j)\|_\infty+\|\hat{\Lambda}^{-1/2}\Lambda^{1/2}-I\|_\infty\cdot\\
&\{\|n^{-1}\sum_{i=1}^n \tilde{G}_i\tilde{G}_i'-E(\tilde{G}_i\tilde{G}_i')\|_\infty+\|E(\tilde{G}_i\tilde{G}_i')\|_\infty\}.
\end{aligned}\tag{S2.17}$$

By combining parts 1–12), (A1.1), and (A3.1) with (S2.17), we have that with probability tending to 1:

$$\max_{j\le h_n s_n+k_n}\|n^{-1}\sum_{i=1}^n \tilde{F}_i(\breve{E}_{ij}-F_i'w_j)\|_\infty\le c_6[\log\{n(p_ns_n+d_n)\}/n]^{1/2},$$

for some universal constant $c_6>0$, which completes the proof of part 13). To show part 14), note that

$$\begin{aligned}
&\max_{j\le h_n s_n+k_n}\|n^{-1}\sum_{i=1}^n \breve{F}_i(\breve{E}_{ij}-F_i'w_j)\|_\infty\\
\le&(2+\|\hat{\Lambda}^{-1/2}\Lambda^{1/2}-I\|_\infty)\max_{j\le h_n s_n+k_n}\|n^{-1}\sum_{i=1}^n \tilde{F}_i(\tilde{E}_{ij}-F_i'w_j)\|_\infty.
\end{aligned}\tag{S2.18}$$

By combining parts 1–13), (A3.1) with (S2.18), we have that with probability tending to 1:

$$\max_{j\le h_n s_n+k_n}\|n^{-1}\sum_{i=1}^n \breve{F}_i(\breve{E}_{ij}-F_i'w_j)\|_\infty\le c_7[\log\{n(p_ns_n+d_n)\}/n]^{1/2},$$

for some universal constant $c_7 > 0$, which completes the proof of part 14). To show part 15), note that for any $t > 0$,

$$P\big(\|n^{-1}\sum_{i=1}^{n}\tilde{G}_i\|_\infty \geq t\big) = P\big(\max_{l\leq p_n s_n + d_n}|n^{-1}\sum_{i=1}^{n}\tilde{G}_{il}| \geq t\big)$$
$$\leq \sum_{l=1}^{p_n s_n + d_n} P\big(|n^{-1}\sum_{i=1}^{n}\tilde{G}_{il}| \geq t\big) \leq 2(p_n s_n + d_n)\exp\{-n(t/c_8)^2\}, \quad \text{(S2.19)}$$

for some universal constant $c_8 > 0$, where the last inequality follows from (A1.1) and Hoeffding's inequality. Plugging $t = c_8[\log\{n(p_n s_n + d_n)\}/n]^{1/2}$ into (S2.19) yields

$$P\Big(\|n^{-1}\sum_{i=1}^{n}\tilde{G}_i\|_\infty \leq c_8[\log\{n(p_n s_n + d_n)\}/n]^{1/2}\Big)$$
$$\geq 1 - 2n^{-1} \to 1, \quad \text{(S2.20)}$$

which completes the proof of part 15). To show part 16), note that

$$\|n^{-1}\sum_{i=1}^{n}G_i\|_\infty \leq \{1 + \lambda_{\max}(\Lambda^{1/2})\}\|n^{-1}\sum_{i=1}^{n}\tilde{G}_i\|_\infty \leq c_9\|n^{-1}\sum_{i=1}^{n}\tilde{G}_i\|_\infty,$$

for some universal constant $c_9 > 0$, where the last inequality is based on (A2.1). Together with (S2.20), it is obvious that

$$P\Big(\|n^{-1}\sum_{i=1}^{n}G_i\|_\infty \leq c_8 c_9[\log\{n(p_n s_n + d_n)\}/n]^{1/2}\Big)$$
$$\geq 1 - 2n^{-1} \to 1, \quad \text{(S2.21)}$$

which completes the proof of part 16). To show part 17), note that

$$\|n^{-1}\sum_{i=1}^{n}\breve{G}_i\|_\infty \leq (2 + \|\hat{\Lambda}^{-1/2}\Lambda^{1/2} - I\|_\infty)\|n^{-1}\sum_{i=1}^{n}\tilde{G}_i\|_\infty.$$

Together with parts 5) and 15), the assertion in part 17) holds obviously. To show part 18), note that for any $t > 0$,

$$P\big(\max_{j\le p_ns_n+d_n}\max_{i\le n}|\tilde{G}_{ij}|\ge t\big)\le\sum_{j=1}^{p_ns_n+d_n}\sum_{i=1}^{n}P\big(|\tilde{G}_{ij}|\ge t\big)$$

$$\le 2n(p_ns_n+d_n)\exp\{-(t/c_{10})^2\}, \tag{S2.22}$$

for some universal constant $c_{10} > 0$, where the last inequality is based on (A1.1). Plugging $t = 2c_{10}[\log\{n(p_ns_n + d_n)\}]^{1/2}$ into (S2.22) yields

$$P\Big(\max_{j\le p_ns_n+d_n}\max_{i\le n}|\tilde{G}_{ij}|\le 2c_{10}[\log\{n(p_ns_n+d_n)\}]^{1/2}\Big)$$

$$\ge 1-2n^{-3}(p_ns_n+d_n)^{-3}\to 1,$$

which completes the proof of part 18). In a similar fashion to part 18), one can show parts 19) and 20). □

**Lemma 3.** *Under conditions (A1.1), (A2.1), (A2.2), (A2.4), (A3.1) and (A3.3), we have that with probability tending to 1:*

*1)* $\|n^{-1}(\Theta\Lambda^{-1/2},Z)'(Y-\Theta\eta-Z\gamma)\|_\infty\le c_1\big(q_ns_n^{-\delta}+[\log\{n(p_ns_n+d_n)\}/n]^{1/2}\big)$, *for some universal constant* $c_1 > 0$.

*2)* $\|n^{-1}(\Theta,Z)'(Y-\Theta\eta-Z\gamma)\|_\infty\le c_2\big(q_ns_n^{-\delta}+[\log\{n(p_ns_n+d_n)\}/n]^{1/2}\big)$, *for some universal constant* $c_2 > 0$.

*Note that* $Y = (Y_1,\ldots,Y_n)'$ *denotes the* $n\times 1$ *response vector.*

*Proof.* First of all, note that

$$\begin{aligned}&\|n^{-1}(\Theta\Lambda^{-1/2}, Z)'(Y-\Theta\eta - Z\gamma)\|_\infty\\ =&\|n^{-1}\sum_{i=1}^n \tilde{G}_i(Y_i - \sum_{j=1}^{p_n}\sum_{k=1}^{s_n}\theta_{ijk}\eta_{jk} - Z_i'\gamma)\|_\infty \le \Delta_1 + \Delta_2, \end{aligned} \tag{S2.23}$$

with $\Delta_1 = \|n^{-1}\sum_{i=1}^n \tilde{G}_i\epsilon_i\|_\infty$ and $\Delta_2 = \|n^{-1}\sum_{i=1}^n \tilde{G}_i \sum_{j=1}^{p_n}\sum_{k=s_n+1}^\infty \theta_{ijk}\eta_{jk}\|_\infty$. To bound the term $\Delta_1$, it follows directly from part 8) of Lemma 2 that with probability tending to 1,

$$\Delta_1 \le c_1[\log\{n(p_n s_n + d_n)\}/n]^{1/2}, \tag{S2.24}$$

for some universal constant $c_1 > 0$. To bound the term $\Delta_2$, first note that

$$\Delta_2 = \max_{l\le p_n s_n + d_n} |n^{-1}\sum_{i=1}^n \tilde{G}_{il}\sum_{j=1}^{q_n}\sum_{k=s_n+1}^\infty \theta_{ijk}\eta_{jk}| \le \Psi_1^{1/2}\Psi_2^{1/2}, \tag{S2.25}$$

with $\Psi_1 = \max_{l\le p_n s_n+d_n} n^{-1}\sum_{i=1}^n \tilde{G}_{il}^2$ and $\Psi_2 = n^{-1}\sum_{i=1}^n(\sum_{j=1}^{q_n}\sum_{k=s_n+1}^\infty \theta_{ijk}\eta_{jk})^2$. For the term $\Psi_1$, we have

$$\begin{aligned}\Psi_1 =& \max_{l\le p_n s_n+d_n}[n^{-1}\sum_{i=1}^n\{\tilde{G}_{il}^2 - E(\tilde{G}_{il}^2)\} + E(\tilde{G}_{1l}^2)]\\ \le&\|n^{-1}(\Theta\Lambda^{-1/2}, Z)'(\Theta\Lambda^{-1/2}, Z) - E(\tilde{G}_i\tilde{G}_i')\|_\infty + \lambda_{\max}(\tilde{\mathcal{I}}_n) \le c_2, \end{aligned} \tag{S2.26}$$

for some universal constant $c_2 > 0$, with probability tending to 1, where the last inequality is based on Lemma 2, (A2.4) and (A3.1). For the term $\Psi_2$, note that

$$\begin{aligned}E(\Psi_2) \le& n^{-1}q_n\sum_{i=1}^n\sum_{j=1}^{q_n}E(\sum_{k=s_n+1}^\infty \theta_{ijk}\eta_{jk})^2 \le q_n\sum_{j=1}^{q_n}\big(\sum_{k=s_n+1}^\infty \omega_{jk}\big)\big(\sum_{k=s_n+1}^\infty \eta_{jk}^2 k^{2\delta}k^{-2\delta}\big)\\ \lesssim& q_n^2 s_n^{-2\delta}\big(\sup_{j\le q_n}\sum_{k=s_n+1}^\infty \omega_{jk}\big)\big(\sup_{j\le q_n}\sum_{k=s_n+1}^\infty \eta_{jk}^2 k^{2\delta}\big) \lesssim o(q_n^2 s_n^{-2\delta}),\end{aligned}$$

where the last inequality follows from (A2.1), (A2.2) and (A3.3). This further implies that with probability tending to 1,

$$\Psi_2 \leq c_3 q_n^2 s_n^{-2\delta}, \tag{S2.27}$$

for some universal constant $c_3 > 0$. By combining (S2.26), (S2.27) with (S2.25), it can be seen that that with probability tending to 1,

$$\Delta_2 \leq c_4 q_n s_n^{-\delta}, \tag{S2.28}$$

for some universal constant $c_4 > 0$. To this end, by combining (S2.24), (S2.28) with (S2.23), it is obvious that with probability tending to 1,

$$\|n^{-1}(\Theta\Lambda^{-1/2}, Z)'(Y - \Theta\eta - Z\gamma)\|_\infty \leq c_5\big(q_n s_n^{-\delta} + [\log\{n(p_n s_n + d_n)\}/n]^{1/2}\big),$$

for some universal constant $c_5 > 0$, which completes the proof of part 1). Together with the fact that $\|n^{-1}(\Theta, Z)'(Y - \Theta\eta - Z\gamma)\|_\infty \lesssim \|n^{-1}(\Theta\Lambda^{-1/2}, Z)'(Y - \Theta\eta - Z\gamma)\|_\infty$, the assertion in part 2) holds trivially. □

**Lemma 4.** *Under conditions (A1.1), (A2.1), (A2.4), (A3.1)-(A3.2) and (B1)-(B5), we have that with probability tending to 1:*

*1)* $\sum_{j=1}^{p_n} \|\Theta_j(\eta_j - \eta_j^*)\|_2^2 + \sum_{l=1}^{d_n}(\sum_{i=1}^n Z_{il}^2)|\gamma_l - \gamma_l^*|^2 \leq c_1 n\|\tilde{\nu}\|_2^2$,

*for some universal constant* $c_1 > 0$.

*2)* $\|\tilde{\nu}\|_1 \leq c_2 n^{-1/2}\{s_n^{1/2}\sum_{j=1}^{p_n}\|\Theta_j(\eta_j - \eta_j^*)\|_2 + \sum_{l=1}^{d_n}(\sum_{i=1}^n Z_{il}^2)^{1/2}|\gamma_l - \gamma_l^*|\}$,

*for some universal constant* $c_2 > 0$.

*3)* $\lambda_n\|\tilde{\nu}\|_1 \le c_3 s_n^{1/2} n^{1/18}\{P_{\lambda_n}(\eta^*,\gamma^*) + P_{\lambda_n}(\eta,\gamma) + n^{-1/9}\|\tilde{\nu}\|_2^2\}$,

*for some universal constant* $c_3 > 0$.

*Recall that* $(\eta^*,\gamma^*)$ *is the true* $(\eta,\gamma)$*, and* $\tilde{\nu} = [(\Lambda^{1/2}\eta - \Lambda^{1/2}\eta^*)', (\gamma-\gamma^*)']'$.

*Proof.* First of all, note that

$$
\begin{aligned}
&\sum_{j=1}^{p_n}\|\Theta_j(\eta_j-\eta_j^*)\|_2^2 + \sum_{l=1}^{d_n}(\sum_{i=1}^{n} Z_{il}^2)|\gamma_l-\gamma_l^*|^2\\
=&\sum_{j=1}^{p_n}\{\Lambda_j^{1/2}(\eta_j-\eta_j^*)\}'\{\Lambda_j^{-1/2}\Theta_j{}'\Theta_j\Lambda_j^{-1/2} - E(\Lambda_j^{-1/2}\Theta_j{}'\Theta_j\Lambda_j^{-1/2})\}\cdot\\
&\{\Lambda_j^{1/2}(\eta_j-\eta_j^*)\} + \sum_{j=1}^{p_n}\{\Lambda_j^{1/2}(\eta_j-\eta_j^*)\}'E(\Lambda_j^{-1/2}\Theta_j{}'\Theta_j\Lambda_j^{-1/2})\{\Lambda_j^{1/2}(\eta_j-\eta_j^*)\}\\
&+\sum_{l=1}^{d_n}\{\sum_{i=1}^{n} Z_{il}^2 - nE(Z_{il}^2)\}|\gamma_l-\gamma_l^*|^2 + n\sum_{l=1}^{d_n}E(Z_{il}^2)|\gamma_l-\gamma_l^*|^2\\
\le& n\{\lambda_{\max}(E(\tilde{G}_i\tilde{G}_i')) + s_n\|n^{-1}(\Theta\Lambda^{-1/2},Z)'(\Theta\Lambda^{-1/2},Z) - E(\tilde{G}_i\tilde{G}_i')\|_\infty\}\cdot\\
&\{\sum_{j=1}^{p_n}\|\Lambda_j^{1/2}(\eta_j-\eta_j^*)\|_2^2 + \sum_{l=1}^{d_n}|\gamma_l-\gamma_l^*|^2\}\\
\le& c_1 n\big(1 + s_n[\log\{n(p_n s_n + d_n)\}/n]^{1/2}\big)\|\tilde{\nu}\|_2^2 \le c_2 n\|\tilde{\nu}\|_2^2, \qquad \text{(S2.29)}
\end{aligned}
$$

for some universal constants $c_1, c_2 > 0$, with probability tending to 1, where the second last inequality follows from (A2.4) and Lemma 2, and the last inequality is based on (A3.1) and (A3.2). This completes the proof of part 1). To show part

2), note that

$$
\begin{aligned}
&\|\tilde{\nu}\|_1 = \sum_{j=1}^{p_n} \|\Lambda_j^{1/2}(\eta_j - \eta_j^*)\|_1 + \sum_{l=1}^{d_n} |\gamma_l - \gamma_l^*| \\
\leq & s_n^{1/2} \sum_{j=1}^{p_n} \left[\{\Lambda_j^{1/2}(\eta_j - \eta_j^*)\}'\{\Lambda_j^{1/2}(\eta_j - \eta_j^*)\}\right]^{1/2} + \sum_{l=1}^{d_n} |\gamma_l - \gamma_l^*| \\
\leq & s_n^{1/2}\{\lambda_{\min}(E(\tilde{G}_i\tilde{G}_i'))\}^{-1/2} \sum_{j=1}^{p_n} \left[\{\Lambda_j^{1/2}(\eta_j - \eta_j^*)\}'\{n^{-1}E(\Lambda_j^{-1/2}\Theta_j\Theta_j\Lambda_j^{-1/2})\}\{\Lambda_j^{1/2}(\eta_j - \eta_j^*)\}\right]^{1/2} \\
& + \{\lambda_{\min}(E(\tilde{G}_i\tilde{G}_i'))\}^{-1/2} \sum_{l=1}^{d_n} \{E(Z_{il}^2)\}^{1/2}|\gamma_l - \gamma_l^*| \\
\leq & c_3 s_n^{1/2} \sum_{j=1}^{p_n} \left[\{\Lambda_j^{1/2}(\eta_j - \eta_j^*)\}'\{n^{-1}E(\Lambda_j^{-1/2}\Theta_j\Theta_j\Lambda_j^{-1/2})\}\{\Lambda_j^{1/2}(\eta_j - \eta_j^*)\}\right]^{1/2} \\
& + c_3 \sum_{l=1}^{d_n} \{E(Z_{il}^2)\}^{1/2}|\gamma_l - \gamma_l^*| \\
\leq & c_3 n^{-1/2}\{s_n^{1/2} \sum_{j=1}^{p_n} \|\Theta_j(\eta_j - \eta_j^*)\|_2 + \sum_{l=1}^{d_n} (\sum_{i=1}^{n} Z_{il}^2)^{1/2}|\gamma_l - \gamma_l^*|\} \\
& + c_3 s_n^{1/2}\|n^{-1}(\Theta\Lambda^{-1/2}, Z)'(\Theta\Lambda^{-1/2}, Z) - E(\tilde{G}_i\tilde{G}_i')\|_\infty \|\tilde{\nu}\|_1 \\
\leq & c_3 n^{-1/2}\{s_n^{1/2} \sum_{j=1}^{p_n} \|\Theta_j(\eta_j - \eta_j^*)\|_2 + \sum_{l=1}^{d_n} (\sum_{i=1}^{n} Z_{il}^2)^{1/2}|\gamma_l - \gamma_l^*|\} \\
& + c_4[s_n \log\{n(p_n s_n + d_n)\}/n]^{1/2}\|\tilde{\nu}\|_1,
\end{aligned}
$$

for some universal constants $c_3, c_4 > 0$, with probability tending to 1, where the third inequality follows from (A2.4) and Lemma 2, and the last inequality is based on Lemma 2. Together with the fact that $s_n \log\{n(p_n s_n + d_n)\}/n \to 0$

under (A3.1) and (A3.2), it is seen that

$$\|\tilde{\nu}\|_1 \lesssim n^{-1/2}\{s_n^{1/2}\sum_{j=1}^{p_n}\|\Theta_j(\eta_j-\eta_j^*)\|_2+\sum_{l=1}^{d_n}(\sum_{i=1}^{n}Z_{il}^2)^{1/2}|\gamma_l-\gamma_l^*|\}, \quad \text{(S2.30)}$$

with probability tending to 1, which completes the proof of part 2). To show part 3), note that

$$\begin{aligned}
&\lambda_n\|\tilde{\nu}\|_1\\
\lesssim& s_n^{1/2}n^{1/18}\lambda_n L\{\sum_{j=1}^{p_n}n^{-5/9}\|\Theta_j(\eta_j-\eta_j^*)\|_2+\sum_{l=1}^{d_n}n^{-5/9}(\sum_{i=1}^{n}Z_{il}^2)^{1/2}|\gamma_l-\gamma_l^*|\}\\
\lesssim& s_n^{1/2}n^{1/18}\Big[\sum_{j=1}^{p_n}\rho_{\lambda_n}(n^{-5/9}\|\Theta_j(\eta_j-\eta_j^*)\|_2)+\sum_{j=1}^{d_n}\rho_{\lambda_n}(n^{-5/9}(\sum_{i=1}^{n}Z_{il}^2)^{1/2}|\gamma_l-\gamma_l^*|)\\
&+2^{-1}\mu n^{-10/9}\{\sum_{j=1}^{p_n}\|\Theta_j(\eta_j-\eta_j^*)\|_2^2+\sum_{l=1}^{d_n}(\sum_{i=1}^{n}Z_{il}^2)|\gamma_l-\gamma_l^*|^2\}\Big]\\
\lesssim& s_n^{1/2}n^{1/18}\{P_{\lambda_n}(\eta^*,\gamma^*)+P_{\lambda_n}(\eta,\gamma)+n^{-1/9}\|\tilde{\nu}\|_2^2\},
\end{aligned}$$

with probability tending to 1, where the first inequality holds from (S2.30), the second inequality is based on part 3) of Lemma 1, and the last inequality follows from the subadditivity of $\rho_\lambda(\cdot)$ and (S2.29). This completes the proof of part 3). □

**Lemma 5.** *Let $X_1, X_2, \ldots, X_n$ be centered independent random vectors, with each $X_i = (X_{i1}, \ldots, X_{ip})' \in \mathbb{R}^p$. Assume the following conditions (a)–(c):*

*(a)* $\min_{j\le p} n^{-1}\sum_{i=1}^n E(X_{ij}^2) \ge c_1$, *for some universal constant* $c_1 > 0$.

*(b)* $X_{ij}$ *are sub-exponential with* $\sup_{k\geq 1} k^{-1}(E|X_{ij}|^k)^{1/k} \leq c_2$*, for some universal constant* $c_2 > 0$*.*

*(c)* $\log^9(np)/n \to 0$*.*

*Then, under (a)–(c), we have:*

$$\lim_{n\to\infty} \sup_{A\in\mathcal{A}^{Re}} |P(V \in A) - P_e(V_e \in A)| = 0,$$

*with* $V = n^{-1/2}\sum_{i=1}^n X_i$ *and* $V_e = n^{-1/2}\sum_{i=1}^n e_i X_i$ *where* $e = \{e_1, \ldots, e_n\}$ *is a set of i.i.d standard normals independent of the data, and* $P_e(\cdot)$ *means the probability with respect to* $e$ *only. The set* $\mathcal{A}^{Re}$ *consists of all hyperrectangles* $A$ *of the form* $A = \{\omega \in \mathbb{R}^p : a_j \leq \omega_j \leq b_j, j \leq p\}$ *for some* $-\infty \leq a_j \leq b_j \leq \infty$ *for all* $j \leq p$*. Further assume that there exist statistics* $\hat{V}$ *and* $\hat{V}_e$ *in* $\mathbb{R}^p$ *satisfying conditions (d)–(e):*

*(d) There exists a sequence of constants* $a_n > 0$ *such that*

$$P(\|\hat{V} - V\|_\infty \geq a_n) \to 0, \qquad P_e(\|\hat{V}_e - V_e\|_\infty \geq a_n) \xrightarrow{p} 0.$$

*(e) The sequence* $a_n$ *in (d) also satisfies*

$$a_n^2 \max\{1, \log(p/a_n)\} \to 0.$$

*Then, under (a)–(e), we have:*

$$\lim_{n\to\infty} \sup_{A\in\mathcal{A}^{Re}} |P(\hat{V} \in A) - P_e(\hat{V}_e \in A)| = 0.$$

*Proof.* Similar to the Lemma H.7 in Ning and Liu (2017), this is adapted from Chernozhukov et al. (2013) and Chernozhukov et al. (2017). □

Next, we state Lemma 6, which is on the property of $\hat{w}$.

**Lemma 6.** *Under conditions (A1.1), (A2.1), (A2.4), (A3.1), (A4.3) and (A4.4), we have*

1) *There is a universal constant* $c_1 > 0$ *such that:*

$$P\big(\max_{j\le h_n s_n+k_n} \|n^{-1}\sum_{i=1}^{n} \breve{F}_i F_i'(\hat{w}_j - w_j)\|_\infty \le c_1[\log\{n(p_n s_n + d_n)\}/n]^{1/2}\big) \to 1,$$
$$P\big(\max_{j\le h_n s_n+k_n} \|n^{-1}\sum_{i=1}^{n} \tilde{F}_i F_i'(\hat{w}_j - w_j)\|_\infty \le c_1[\log\{n(p_n s_n + d_n)\}/n]^{1/2}\big) \to 1.$$

2) *There is a universal constant* $c_2 > 0$ *such that:*

$$P\big(\bigcap_{j=1}^{h_n s_n+k_n} \big\{\|diag\{\hat{\Lambda}^{1/2}_{\mathcal{H}_n^c}, I_{(d_n-k_n)}\}(\hat{w}_j - w_j)\|_2 \le c_2[\rho_{nj}\log\{n(p_n s_n + d_n)\}/n]^{1/2}\big\}\big) \to 1,$$

$$P\big(\max_{j\le h_n s_n+k_n} \|diag\{\hat{\Lambda}^{1/2}_{\mathcal{H}_n^c}, I_{(d_n-k_n)}\}(\hat{w}_j - w_j)\|_2 \le c_2[\rho_n\log\{n(p_n s_n + d_n)\}/n]^{1/2}\big) \to 1.$$

3) *There is a universal constant* $c_3 > 0$ *such that:*

$$P\big(\bigcap_{j=1}^{h_n s_n+k_n} \big\{\|diag\{\hat{\Lambda}^{1/2}_{\mathcal{H}_n^c}, I_{(d_n-k_n)}\}(\hat{w}_j - w_j)\|_1 \le c_3\rho_{nj}[\log\{n(p_n s_n + d_n)\}/n]^{1/2}\big\}\big) \to 1,$$

$$P\big(\max_{j\le h_n s_n+k_n} \|diag\{\hat{\Lambda}^{1/2}_{\mathcal{H}_n^c}, I_{(d_n-k_n)}\}(\hat{w}_j - w_j)\|_1 \le c_3\rho_n[\log\{n(p_n s_n + d_n)\}/n]^{1/2}\big) \to 1.$$

*Proof.* By the definition of $\hat{w}$ in (3.5) of the main paper, it holds true for all

$j \leq h_n s_n + k_n$ that

$$
\begin{aligned}
&(2n)^{-1}\sum_{i=1}^{n}(\breve{E}_{ij} - F_i'\hat{w}_j)^2 + \lambda_n^* \|diag\{\hat{\Lambda}_{\mathcal{H}_n^c}^{1/2}, I_{(d_n-k_n)}\}\hat{w}_j\|_1 \\
\leq &(2n)^{-1}\sum_{i=1}^{n}(\breve{E}_{ij} - F_i'w_j)^2 + \lambda_n^* \|diag\{\hat{\Lambda}_{\mathcal{H}_n^c}^{1/2}, I_{(d_n-k_n)}\}w_j\|_1,
\end{aligned}
$$

which implies that for all $j \leq h_n s_n + k_n$,

$$
\begin{aligned}
0 \leq &(\hat{w}_j - w_j)'\{(2n)^{-1}\sum_{i=1}^{n} F_i F_i'\}(\hat{w}_j - w_j) \\
\leq &\|n^{-1}\sum_{i=1}^{n}\breve{F}_i(\breve{E}_{ij} - F_i'w_j)\|_\infty \cdot \|diag\{\hat{\Lambda}_{\mathcal{H}_n^c}^{1/2}, I_{(d_n-k_n)}\}(\hat{w}_j - w_j)\|_1 \\
&+ \lambda_n^* \|diag\{\hat{\Lambda}_{\mathcal{H}_n^c}^{1/2}, I_{(d_n-k_n)}\}w_j\|_1 - \lambda_n^* \|diag\{\hat{\Lambda}_{\mathcal{H}_n^c}^{1/2}, I_{(d_n-k_n)}\}\hat{w}_j\|_1. \quad \text{(S2.31)}
\end{aligned}
$$

Now, we denote $\mathcal{S}_j = \{l : w_{jl} \neq 0\}$ as the support set of $w_j$, whose complement is $\mathcal{S}_j^c = \{1, \ldots, (p_n - h_n)s_n + d_n - k_n\}/\mathcal{S}_j$. For any vector $v = (v_1, \ldots, v_{(p_n-h_n)s_n+d_n-k_n})'$, we write the vector $v_{\mathcal{S}_j}$ as restricting $v$ to $\mathcal{S}_j$. Then, it follows from triangle inequality that for all $j \leq h_n s_n + k_n$,

$$
\begin{aligned}
&\|diag\{\hat{\Lambda}_{\mathcal{H}_n^c}^{1/2}, I_{(d_n-k_n)}\}\hat{w}_j\|_1 \\
=&\|\big(diag\{\hat{\Lambda}_{\mathcal{H}_n^c}^{1/2}, I_{(d_n-k_n)}\}\hat{w}_j\big)_{\mathcal{S}_j}\|_1 + \|\big(diag\{\hat{\Lambda}_{\mathcal{H}_n^c}^{1/2}, I_{(d_n-k_n)}\}\hat{w}_j\big)_{\mathcal{S}_j^c}\|_1 \\
\geq &\|diag\{\hat{\Lambda}_{\mathcal{H}_n^c}^{1/2}, I_{(d_n-k_n)}\}w_j\|_1 - \|\big(diag\{\hat{\Lambda}_{\mathcal{H}_n^c}^{1/2}, I_{(d_n-k_n)}\}(\hat{w}_j - w_j)\big)_{\mathcal{S}_j}\|_1 \\
&+ \|\big(diag\{\hat{\Lambda}_{\mathcal{H}_n^c}^{1/2}, I_{(d_n-k_n)}\}\hat{w}_j\big)_{\mathcal{S}_j^c}\|_1.
\end{aligned}
$$

Together with (S2.31) yields that for all $j \leq h_n s_n + k_n$,

$$
\begin{aligned}
0 \leq & (\hat{w}_j - w_j)'\{(2n)^{-1}\sum_{i=1}^{n} F_i F_i'\}(\hat{w}_j - w_j) \qquad \text{(S2.32)} \\
\leq & \{\lambda_n^* + \|n^{-1}\sum_{i=1}^{n} \breve{F}_i(\breve{E}_{ij} - F_i' w_j)\|_\infty\} \cdot \|\big(diag\{\hat{\Lambda}_{\mathcal{H}_n^c}^{1/2}, I_{(d_n-k_n)}\}(\hat{w}_j - w_j)\big)_{\mathcal{S}_j}\|_1 \\
& - \{\lambda_n^* - \|n^{-1}\sum_{i=1}^{n} \breve{F}_i(\breve{E}_{ij} - F_i' w_j)\|_\infty\} \cdot \|\big(diag\{\hat{\Lambda}_{\mathcal{H}_n^c}^{1/2}, I_{(d_n-k_n)}\}(\hat{w}_j - w_j)\big)_{\mathcal{S}_j^c}\|_1 \\
\leq & \{\lambda_n^* + \max_{l \leq h_n s_n + k_n} \|n^{-1}\sum_{i=1}^{n} \breve{F}_i(\breve{E}_{il} - F_i' w_l)\|_\infty\} \cdot \|\big(diag\{\hat{\Lambda}_{\mathcal{H}_n^c}^{1/2}, I_{(d_n-k_n)}\}(\hat{w}_j - w_j)\big)_{\mathcal{S}_j}\|_1 \\
& - \{\lambda_n^* - \max_{l \leq h_n s_n + k_n} \|n^{-1}\sum_{i=1}^{n} \breve{F}_i(\breve{E}_{il} - F_i' w_l)\|_\infty\} \cdot \|\big(diag\{\hat{\Lambda}_{\mathcal{H}_n^c}^{1/2}, I_{(d_n-k_n)}\}(\hat{w}_j - w_j)\big)_{\mathcal{S}_j^c}\|_1.
\end{aligned}
$$

Based on Lemma 2, there is a universal constant $c_1 > 0$ such that with probability tending to 1:

$$
\max_{j \leq h_n s_n + k_n} \|n^{-1}\sum_{i=1}^{n} \breve{F}_i(\breve{E}_{ij} - F_i' w_j)\|_\infty \leq c_1[\log\{n(p_n s_n + d_n)\}/n]^{1/2}.
$$

By choosing $K_1 \geq 2c_1$ in (A4.3), we have that with probability tending to 1:

$$
\max_{j \leq h_n s_n + k_n} \|n^{-1}\sum_{i=1}^{n} \breve{F}_i(\breve{E}_{ij} - F_i' w_j)\|_\infty \leq 2^{-1}\lambda_n^*. \qquad \text{(S2.33)}
$$

It follows from (S2.32) and (S2.33) that

$$
\begin{aligned}
P\Big(\bigcap_{j=1}^{h_n s_n + k_n} & \{\|\big(diag\{\hat{\Lambda}_{\mathcal{H}_n^c}^{1/2}, I_{(d_n-k_n)}\}(\hat{w}_j - w_j)\big)_{\mathcal{S}_j^c}\|_1 \leq \\
& 3\|\big(diag\{\hat{\Lambda}_{\mathcal{H}_n^c}^{1/2}, I_{(d_n-k_n)}\}(\hat{w}_j - w_j)\big)_{\mathcal{S}_j}\|_1\}\Big) \to 1,
\end{aligned}
$$

which further implies that

$$P\Big(\bigcap_{j=1}^{h_n s_n+k_n}\big\{\|diag\{\hat{\Lambda}_{\mathcal{H}_n^c}^{1/2}, I_{(d_n-k_n)}\}(\hat{w}_j-w_j)\|_1 \leq$$

$$4\rho_{nj}^{1/2}\|diag\{\hat{\Lambda}_{\mathcal{H}_n^c}^{1/2}, I_{(d_n-k_n)}\}(\hat{w}_j-w_j)\|_2\big\}\Big) \to 1. \tag{S2.34}$$

Based on (3.5) of the main paper and the Karush-Kuhn-Tucker condition, it is seen that

$$\max_{j\leq h_n s_n+k_n}\|n^{-1}\sum_{i=1}^n \breve{F}_i(\breve{E}_{ij}-F_i'\hat{w}_j)\|_\infty \leq \lambda_n^*. \tag{S2.35}$$

To bound the term $\max_{j\leq h_n s_n+k_n}\|n^{-1}\sum_{i=1}^n \breve{F}_i F_i'(\hat{w}_j-w_j)\|_\infty$, note that

$$\begin{aligned}
&\max_{j\leq h_n s_n+k_n}\|n^{-1}\sum_{i=1}^n \breve{F}_i F_i'(\hat{w}_j-w_j)\|_\infty \\
\leq &\max_{j\leq h_n s_n+k_n}\|n^{-1}\sum_{i=1}^n \breve{F}_i(\breve{E}_{ij}-F_i'\hat{w}_j)\|_\infty + \max_{j\leq h_n s_n+k_n}\|n^{-1}\sum_{i=1}^n \breve{F}_i(\breve{E}_{ij}-F_i'w_j)\|_\infty \\
\leq &3\lambda_n^*/2 \leq (3K_2/2)\cdot[\log\{n(p_n s_n+d_n)\}/n]^{1/2},
\end{aligned} \tag{S2.36}$$

with probability tending to 1, where the second inequality follows from (S2.33) and (S2.35), and the last inequality holds from (A4.3). Together with Lemma 2, it can be deduced that with probability tending to 1:

$$\max_{j\leq h_n s_n+k_n}\|n^{-1}\sum_{i=1}^n \tilde{F}_i F_i'(\hat{w}_j-w_j)\|_\infty \leq (3K_2)\cdot[\log\{n(p_n s_n+d_n)\}/n]^{1/2}.$$

This finishes the proof of part 1). To show part 2), first note that for all $j \leq$

$h_n s_n + k_n$,

$$
\begin{aligned}
0 \leq & (\hat{w}_j - w_j)'(n^{-1}\sum_{i=1}^{n} F_i F_i')(\hat{w}_j - w_j) \\
= & [diag\{\hat{\Lambda}^{1/2}_{\mathcal{H}_n^c}, I_{(d_n-k_n)}\}(\hat{w}_j - w_j)]' n^{-1}\sum_{i=1}^{n} \breve{F}_i F_i'(\hat{w}_j - w_j) \\
\leq & \|n^{-1}\sum_{i=1}^{n} \breve{F}_i F_i'(\hat{w}_j - w_j)\|_\infty \cdot \|diag\{\hat{\Lambda}^{1/2}_{\mathcal{H}_n^c}, I_{(d_n-k_n)}\}(\hat{w}_j - w_j)\|_1.
\end{aligned}
$$

Together with (S2.34) and (S2.36) yields

$$
\begin{aligned}
P\Big( \bigcap_{j=1}^{h_n s_n + k_n} & \big\{(\hat{w}_j - w_j)'(n^{-1}\sum_{i=1}^{n} F_i F_i')(\hat{w}_j - w_j) \leq 6K_2[\rho_{nj}\log\{n(p_n s_n + d_n)\}/n]^{1/2} \\
& \|diag\{\hat{\Lambda}^{1/2}_{\mathcal{H}_n^c}, I_{(d_n-k_n)}\}(\hat{w}_j - w_j)\|_2\big\}\Big) \to 1. \qquad \text{(S2.37)}
\end{aligned}
$$

Also note that for all $j \leq h_n s_n + k_n$,

$$
\begin{aligned}
& (\hat{w}_j - w_j)'(n^{-1}\sum_{i=1}^{n} F_i F_i')(\hat{w}_j - w_j) \\
= & [diag\{\hat{\Lambda}^{1/2}_{\mathcal{H}_n^c}, I_{(d_n-k_n)}\}(\hat{w}_j - w_j)]' E(\tilde{F}_i \tilde{F}_i')[diag\{\hat{\Lambda}^{1/2}_{\mathcal{H}_n^c} I_{(d_n-k_n)}\}(\hat{w}_j - w_j)] - \\
& [diag\{\hat{\Lambda}^{1/2}_{\mathcal{H}_n^c}, I_{(d_n-k_n)}\}(\hat{w}_j - w_j)]' \{E(\tilde{F}_i \tilde{F}_i') - n^{-1}\sum_{i=1}^{n} \breve{F}_i \breve{F}_i'\} \cdot \\
& [diag\{\hat{\Lambda}^{1/2}_{\mathcal{H}_n^c} I_{(d_n-k_n)}\}(\hat{w}_j - w_j)] \\
\geq & \lambda_{\min}(E(\tilde{F}_i \tilde{F}_i'))\|diag\{\hat{\Lambda}^{1/2}_{\mathcal{H}_n^c}, I_{(d_n-k_n)}\}(\hat{w}_j - w_j)\|_2^2 - \|n^{-1}\sum_{i=1}^{n} \breve{F}_i \breve{F}_i' - E(\tilde{F}_i \tilde{F}_i')\|_\infty \cdot \\
& \|diag\{\hat{\Lambda}^{1/2}_{\mathcal{H}_n^c}, I_{(d_n-k_n)}\}(\hat{w}_j - w_j)\|_1^2 \\
\geq & c_2\|diag\{\hat{\Lambda}^{1/2}_{\mathcal{H}_n^c}, I_{(d_n-k_n)}\}(\hat{w}_j - w_j)\|_2^2 - \|n^{-1}\sum_{i=1}^{n} \breve{F}_i \breve{F}_i' - E(\tilde{F}_i \tilde{F}_i')\|_\infty \cdot \\
& \|diag\{\hat{\Lambda}^{1/2}_{\mathcal{H}_n^c}, I_{(d_n-k_n)}\}(\hat{w}_j - w_j)\|_1^2,
\end{aligned}
$$

for some universal constant $c_2 > 0$, where the last inequality is by (A2.4). Together with Lemma 2, (S2.34), and (A4.4), it can be deduced that

$$P\Big(\bigcap_{j=1}^{h_n s_n + k_n} \big\{(\hat{w}_j - w_j)'(n^{-1}\sum_{i=1}^{n} F_i F_i')(\hat{w}_j - w_j) \geq 2^{-1} c_2 \cdot \\ \|diag\{\hat{\Lambda}_{\mathcal{H}_n^c}^{1/2}, I_{(d_n - k_n)}\}(\hat{w}_j - w_j)\|_2^2\big\}\Big) \to 1.$$

Together with (S2.37) yields

$$P\Big(\bigcap_{j=1}^{h_n s_n + k_n} \big\{\|diag\{\hat{\Lambda}_{\mathcal{H}_n^c}^{1/2}, I_{(d_n - k_n)}\}(\hat{w}_j - w_j)\|_2 \leq 12 c_2^{-1} K_2 \cdot \\ [\rho_{nj} \log\{n(p_n s_n + d_n)\}/n]^{1/2}\big\}\Big) \to 1, \tag{S2.38}$$

which further implies that

$$P\Big(\max_{j \leq h_n s_n + k_n} \|diag\{\hat{\Lambda}_{\mathcal{H}_n^c}^{1/2}, I_{(d_n - k_n)}\}(\hat{w}_j - w_j)\|_2 \leq 12 c_2^{-1} K_2 \cdot \\ [\rho_n \log\{n(p_n s_n + d_n)\}/n]^{1/2}\Big) \to 1.$$

This completes the proof of part 2). By combining (S2.38) with (S2.34), it can be deduced that

$$P\Big(\bigcap_{j=1}^{h_n s_n + k_n} \big\{\|diag\{\hat{\Lambda}_{\mathcal{H}_n^c}^{1/2}, I_{(d_n - k_n)}\}(\hat{w}_j - w_j)\|_1 \leq 48 c_2^{-1} K_2 \cdot \\ [\rho_{nj}^2 \log\{n(p_n s_n + d_n)\}/n]^{1/2}\big\}\Big) \to 1,$$

which further implies that

$$P\big(\max_{j\le h_n s_n+k_n}\|diag\{\hat{\Lambda}^{1/2}_{\mathcal{H}_n^c}, I_{(d_n-k_n)}\}(\hat{w}_j-w_j)\|_1 \le 48c_2^{-1}K_2\cdot$$
$$[\rho_n^2\log\{n(p_ns_n+d_n)\}/n]^{1/2}\big)\to 1.$$

This completes the proof of part 3). □

Next, we state Lemma 7 as follows.

**Lemma 7.** *Under conditions (A1.1), (A2)–(A4) and (B1)–(B5), we have*

*1) There is a universal constant $c_1>0$ that with probability tending to 1:*

$$\max_{l\le h_ns_n+k_n}|n^{-1/2}\sum_{i=1}^n(S_{il}-S^*_{il})|\le c_1\big[\lambda_n s_n^{1/2}(q_n+r_n)n^{-1/18}\log^{1/2}\{n(p_ns_n+d_n)\}$$
$$+n^{1/2}q_ns_n^{-\delta}\log^{1/2}\{n(p_ns_n+d_n)\}+n^{-1/2}\rho_n\log\{n(p_ns_n+d_n)\}\big].$$

*2) There is a universal constant $c_2>0$ that with probability tending to 1:*

$$\max_{l\le h_ns_n+k_n}\{n^{-1}\sum_{i=1}^n(\hat{S}_{il}-S^*_{il})^2\}^{1/2}\le c_2\big[\lambda_n s_n^{1/2}(q_n+r_n)n^{-1/18}\log\{n(p_ns_n+d_n)\}$$
$$+n^{-1/2}\rho_n\log^{3/2}\{n(p_ns_n+d_n)\}\big].$$

*Proof.* First of all, it follows from triangle inequality that

$$\max_{l\le h_ns_n+k_n}|n^{-1/2}\sum_{i=1}^n(S_{il}-S^*_{il})|$$
$$\le\Delta_1+\Delta_2+\Delta_3+\Delta_4+\Delta_5+\Delta_6+\Delta_7+\Delta_8, \qquad \text{(S2.39)}$$

where

$$\Delta_1 = n^{-1/2} \max_{l \le h_n s_n + k_n} |(\hat{w}_l - w_l)' \sum_{i=1}^n F_i \epsilon_i|,$$

$$\Delta_2 = n^{-1/2} \max_{l \le h_n s_n + k_n} |\sum_{i=1}^n (\tilde{E}_{il} - \breve{E}_{il}) \epsilon_i|,$$

$$\Delta_3 = n^{-1/2} \max_{l \le h_n s_n + k_n} |\sum_{i=1}^n (w_l' F_i - \tilde{E}_{il}) (\sum_{j=1}^{q_n} \sum_{k=s_n+1}^{\infty} \theta_{ijk} \eta_{jk})|,$$

$$\Delta_4 = n^{-1/2} \max_{l \le h_n s_n + k_n} |(\hat{w}_l - w_l)' \sum_{i=1}^n F_i (\sum_{j=1}^{q_n} \sum_{k=s_n+1}^{\infty} \theta_{ijk} \eta_{jk})|,$$

$$\Delta_5 = n^{-1/2} \max_{l \le h_n s_n + k_n} |\sum_{i=1}^n (\tilde{E}_{il} - \breve{E}_{il}) (\sum_{j=1}^{q_n} \sum_{k=s_n+1}^{\infty} \theta_{ijk} \eta_{jk})|,$$

$$\Delta_6 = n^{-1/2} \max_{l \le h_n s_n + k_n} |\sum_{i=1}^n (w_l' F_i - \tilde{E}_{il}) F_i' \left( \begin{smallmatrix} \hat{\eta}_{\mathcal{H}_n^c} - \eta_{\mathcal{H}_n^c} \\ \hat{\gamma}_{\mathcal{K}_n^c} - \gamma_{\mathcal{K}_n^c} \end{smallmatrix} \right)|,$$

$$\Delta_7 = n^{-1/2} \max_{l \le h_n s_n + k_n} |(\hat{w}_l - w_l)' \sum_{i=1}^n F_i F_i' \left( \begin{smallmatrix} \hat{\eta}_{\mathcal{H}_n^c} - \eta_{\mathcal{H}_n^c} \\ \hat{\gamma}_{\mathcal{K}_n^c} - \gamma_{\mathcal{K}_n^c} \end{smallmatrix} \right)|,$$

$$\Delta_8 = n^{-1/2} \max_{l \le h_n s_n + k_n} |\sum_{i=1}^n (\tilde{E}_{il} - \breve{E}_{il}) F_i' \left( \begin{smallmatrix} \hat{\eta}_{\mathcal{H}_n^c} - \eta_{\mathcal{H}_n^c} \\ \hat{\gamma}_{\mathcal{K}_n^c} - \gamma_{\mathcal{K}_n^c} \end{smallmatrix} \right)|.$$

To bound $\Delta_1$, note that

$$\begin{aligned} \Delta_1 =& n^{1/2} \max_{l \le h_n s_n + k_n} |[diag\{\hat{\Lambda}_{\mathcal{H}_n^c}^{1/2}, I_{(d_n - k_n)}\}(\hat{w}_l - w_l)]' n^{-1} \sum_{i=1}^n \breve{F}_i \epsilon_i| \\ \le& n^{1/2} \|n^{-1} \sum_{i=1}^n \breve{F}_i \epsilon_i\|_\infty \cdot \max_{l \le h_n s_n + k_n} \|diag\{\hat{\Lambda}_{\mathcal{H}_n^c}^{1/2}, I_{(d_n - k_n)}\}(\hat{w}_l - w_l)\|_1 \\ \le& c_1 n^{-1/2} \rho_n \log\{n(p_n s_n + d_n)\}, \end{aligned} \tag{S2.40}$$

for some universal constant $c_1 > 0$, with probability tending to 1, where the first inequality is by Holder's inequality, and the last inequality follows from

Lemma 2 and Lemma 6. To bound $\Delta_2$, note that

$$\Delta_2 \leq n^{1/2}\|\Lambda^{1/2}\hat{\Lambda}^{-1/2} - I\|_\infty \cdot \|n^{-1}\sum_{i=1}^{n}\tilde{G}_i\epsilon_i\|_\infty$$
$$\leq c_2 n^{-1/2}\log\{n(p_n s_n + d_n)\}, \tag{S2.41}$$

for some universal constant $c_2 > 0$, with probability tending to 1, where the last inequality is based on Lemma 2. To bound $\Delta_3$, note that

$$\Delta_3 \leq n^{-1/2}(\max_{l\leq h_n s_n + k_n}\max_{i\leq n}|w_l' F_i - \tilde{E}_{il}|)\cdot\sum_{i=1}^{n}|\sum_{j=1}^{q_n}\sum_{k=s_n+1}^{\infty}\theta_{ijk}\eta_{jk}|$$
$$\leq (\max_{l\leq h_n s_n + k_n}\max_{i\leq n}|w_l' F_i - \tilde{E}_{il}|)\cdot\{\sum_{i=1}^{n}(\sum_{j=1}^{q_n}\sum_{k=s_n+1}^{\infty}\theta_{ijk}\eta_{jk})^2\}^{1/2}$$
$$\leq c_3[\log\{n(p_n s_n + d_n)\}]^{1/2}\cdot\{\sum_{i=1}^{n}(\sum_{j=1}^{q_n}\sum_{k=s_n+1}^{\infty}\theta_{ijk}\eta_{jk})^2\}^{1/2}, \tag{S2.42}$$

for some universal constant $c_3 > 0$, with probability tending to 1, where the first inequality is by Holder's inequality, the second inequality holds from Cauchy Schwarz inequality, and the last inequality follows from Lemma 2. To bound the term $\sum_{i=1}^{n}(\sum_{j=1}^{q_n}\sum_{k=s_n+1}^{\infty}\theta_{ijk}\eta_{jk})^2$, first note that

$$E\{\sum_{i=1}^{n}(\sum_{j=1}^{q_n}\sum_{k=s_n+1}^{\infty}\theta_{ijk}\eta_{jk})^2\} \leq q_n\sum_{i=1}^{n}\sum_{j=1}^{q_n}E(\sum_{k=s_n+1}^{\infty}\theta_{ijk}\eta_{jk})^2$$
$$\lesssim nq_n^2 s_n^{-2\delta}\big(\sup_{j\leq q_n}\sum_{k=s_n+1}^{\infty}\omega_{jk}\big)\big(\sup_{j\leq q_n}\sum_{k=s_n+1}^{\infty}\eta_{jk}^2 k^{2\delta}\big) \lesssim o(nq_n^2 s_n^{-2\delta}),$$

where the last equality follows from (A2.1) and (A2.2). Together with Markov's

inequality yields

$$P\Big\{\sum_{i=1}^{n}(\sum_{j=1}^{q_n}\sum_{k=s_n+1}^{\infty}\theta_{ijk}\eta_{jk})^2 \le nq_n^2 s_n^{-2\delta}\Big\} \to 1. \tag{S2.43}$$

Together with (S2.42) yields that with probability tending to 1:

$$\Delta_3 \le c_3 n^{1/2} q_n s_n^{-\delta} \log^{1/2}\{n(p_n s_n + d_n)\}. \tag{S2.44}$$

To bound $\Delta_4$, note that

$$\begin{aligned}
\Delta_4 =& n^{-1/2} \max_{l\le h_n s_n + k_n} |[diag\{\hat{\Lambda}^{1/2}_{\mathcal{H}_n^c}, I_{(d_n-k_n)}\}(\hat{w}_l - w_l)]' \sum_{i=1}^{n} \breve{F}_i(\sum_{j=1}^{q_n}\sum_{k=s_n+1}^{\infty}\theta_{ijk}\eta_{jk})| \\
\le& n^{-1/2} \max_{l\le h_n s_n + k_n} \|diag\{\hat{\Lambda}^{1/2}_{\mathcal{H}_n^c}, I_{(d_n-k_n)}\}(\hat{w}_l - w_l)\|_1 \cdot \|\sum_{i=1}^{n} \breve{F}_i(\sum_{j=1}^{q_n}\sum_{k=s_n+1}^{\infty}\theta_{ijk}\eta_{jk})\|_\infty \\
\le& (2 + \|\Lambda^{1/2}\hat{\Lambda}^{-1/2} - I\|_\infty) \cdot \max_{l\le h_n s_n + k_n} \|diag\{\hat{\Lambda}^{1/2}_{\mathcal{H}_n^c}, I_{(d_n-k_n)}\}(\hat{w}_l - w_l)\|_1 \cdot \\
& \max_{j\le p_n s_n + d_n} \max_{i\le n} |\tilde{G}_{ij}| \cdot \{\sum_{i=1}^{n}(\sum_{j=1}^{q_n}\sum_{k=s_n+1}^{\infty}\theta_{ijk}\eta_{jk})^2\}^{1/2} \\
\le& c_4 \rho_n q_n s_n^{-\delta} \log\{n(p_n s_n + d_n)\},
\end{aligned} \tag{S2.45}$$

for some universal constant $c_4 > 0$, with probability tending to 1, where the last inequality is based on Lemma 2, Lemma 6, and (S2.43). To bound $\Delta_5$, note that

$$\begin{aligned}
\Delta_5 \le& n^{-1/2} \max_{l\le h_n s_n + k_n} \max_{i\le n} |\tilde{E}_{il} - \breve{E}_{il}| \cdot \sum_{i=1}^{n} |\sum_{j=1}^{q_n}\sum_{k=s_n+1}^{\infty}\theta_{ijk}\eta_{jk}| \\
\le& \|\Lambda^{1/2}\hat{\Lambda}^{-1/2} - I\|_\infty \cdot \max_{j\le p_n s_n + d_n} \max_{i\le n} |\tilde{G}_{ij}| \cdot \{\sum_{i=1}^{n}(\sum_{j=1}^{q_n}\sum_{k=s_n+1}^{\infty}\theta_{ijk}\eta_{jk})^2\}^{1/2} \\
\le& c_5 q_n s_n^{-\delta} \log\{n(p_n s_n + d_n)\},
\end{aligned} \tag{S2.46}$$

for some universal constant $c_5 > 0$, with probability tending to 1, where the last inequality is based on Lemma 2, and (S2.43). To bound $\Delta_6$, note that

$$\begin{aligned}\Delta_6 =& n^{1/2} \max_{l \le h_n s_n + k_n} |\{n^{-1} \sum_{i=1}^{n} \tilde{F}_i (w_l' F_i - \tilde{E}_{il})\}' diag\{\Lambda_{\mathcal{H}_n^c}^{1/2}, I_{(d_n - k_n)}\} \left( \begin{smallmatrix} \hat{\eta}_{\mathcal{H}_n^c} - \eta_{\mathcal{H}_n^c} \\ \hat{\gamma}_{\mathcal{K}_n^c} - \gamma_{\mathcal{K}_n^c} \end{smallmatrix} \right)| \\ \le& 2n^{1/2} \max_{l \le h_n s_n + k_n} \|n^{-1} \sum_{i=1}^{n} \tilde{F}_i (w_l' F_i - \tilde{E}_{il})\|_\infty \cdot \max\left\{ \|\Lambda^{1/2}(\hat{\eta} - \eta)\|_1, \|\hat{\gamma} - \gamma\|_1 \right\} \\ \le& c_6 \lambda_n s_n^{1/2} (q_n + r_n) n^{-1/18} \log^{1/2}\{n(p_n s_n + d_n)\}, \end{aligned} \tag{S2.47}$$

for some universal constant $c_6 > 0$, with probability tending to 1, where the last inequality is based on Lemma 2, and Theorem 1. To bound $\Delta_7$, note that

$$\begin{aligned}\Delta_7 =& n^{1/2} \max_{l \le h_n s_n + k_n} |\{n^{-1} \sum_{i=1}^{n} \tilde{F}_i F_i' (\hat{w}_l - w_l)\}' diag\{\Lambda_{\mathcal{H}_n^c}^{1/2}, I_{(d_n - k_n)}\} \left( \begin{smallmatrix} \hat{\eta}_{\mathcal{H}_n^c} - \eta_{\mathcal{H}_n^c} \\ \hat{\gamma}_{\mathcal{K}_n^c} - \gamma_{\mathcal{K}_n^c} \end{smallmatrix} \right)| \\ \le& 2n^{1/2} \max_{l \le h_n s_n + k_n} \|n^{-1} \sum_{i=1}^{n} \tilde{F}_i F_i' (\hat{w}_l - w_l)\|_\infty \cdot \max\left\{ \|\Lambda^{1/2}(\hat{\eta} - \eta)\|_1, \|\hat{\gamma} - \gamma\|_1 \right\} \\ \le& c_7 \lambda_n s_n^{1/2} (q_n + r_n) n^{-1/18} \log^{1/2}\{n(p_n s_n + d_n)\}, \end{aligned} \tag{S2.48}$$

for some universal constant $c_7 > 0$, with probability tending to 1, where the last

inequality is based on Lemma 6, and Theorem 1. To bound $\Delta_8$, note that

$$\begin{aligned}\Delta_8 =& n^{-1/2}\max_{l\le h_ns_n+k_n}|\{\sum_{i=1}^n(\tilde{E}_{il}-\breve{E}_{il})\tilde{F}_i\}'diag\{\Lambda^{1/2}_{\mathcal{H}_n^c},I_{(d_n-k_n)}\}\left(\begin{smallmatrix}\hat{\eta}_{\mathcal{H}_n^c}-\eta_{\mathcal{H}_n^c}\\ \hat{\gamma}_{\mathcal{K}_n^c}-\gamma_{\mathcal{K}_n^c}\end{smallmatrix}\right)|\\ \le& 2n^{1/2}\max_{l\le h_ns_n+k_n}\|n^{-1}\sum_{i=1}^n(\tilde{E}_{il}-\breve{E}_{il})\tilde{F}_i\|_\infty\cdot\max\left\{\|\Lambda^{1/2}(\hat{\eta}-\eta)\|_1,\|\hat{\gamma}-\gamma\|_1\right\}\\ \le& 2n^{1/2}\|\Lambda^{1/2}\hat{\Lambda}^{-1/2}-I\|_\infty\cdot\{\|n^{-1}\sum_{i=1}^n\tilde{G}_i\tilde{G}_i'-E(\tilde{G}_i\tilde{G}_i')\|_\infty+\|E(\tilde{G}_i\tilde{G}_i')\|_\infty\}\\ &\cdot\max\left\{\|\Lambda^{1/2}(\hat{\eta}-\eta)\|_1,\|\hat{\gamma}-\gamma\|_1\right\}\\ \le& c_8\lambda_ns_n^{1/2}(q_n+r_n)n^{-1/18}\log^{1/2}\{n(p_ns_n+d_n)\},\end{aligned}\tag{S2.49}$$

for some universal constant $c_8>0$, with probability tending to 1, where the last inequality is based on Lemma 2, (A1.1), and Theorem 1. By combining (S2.40), (S2.41), (S2.44), (S2.45), (S2.46), (S2.47), (S2.48), (S2.49) with (S2.39), it can be deduced that there is a universal constant $c_9>0$ such that with probability tending to 1:

$$\begin{aligned}\max_{l\le h_ns_n+k_n}|n^{-1/2}\sum_{i=1}^n(S_{il}-S_{il}^*)|\le c_9\big[&\lambda_ns_n^{1/2}(q_n+r_n)n^{-1/18}\log^{1/2}\{n(p_ns_n+d_n)\}\\ &+n^{1/2}q_ns_n^{-\delta}\log^{1/2}\{n(p_ns_n+d_n)\}+n^{-1/2}\rho_n\log\{n(p_ns_n+d_n)\}\big].\end{aligned}$$

This completes the proof of part 1). To show part 2), it follows from triangle inequality that

$$\begin{aligned}&\max_{l\le h_ns_n+k_n}\{n^{-1}\sum_{i=1}^n(\hat{S}_{il}-S_{il}^*)^2\}^{1/2}\\ \lesssim&\Delta_1^*+\Delta_2^*+\Delta_3^*+\Delta_4^*+\Delta_5^*+\Delta_6^*+\Delta_7^*+\Delta_8^*,\end{aligned}\tag{S2.50}$$

where

$$\Delta_1^* = \big[\max_{l\le h_n s_n+k_n} n^{-1}\sum_{i=1}^n\{(\hat{w}_l - w_l)'F_i\epsilon_i\}^2\big]^{1/2},$$

$$\Delta_2^* = \big[\max_{l\le h_n s_n+k_n} n^{-1}\sum_{i=1}^n\{(\tilde{E}_{il} - \breve{E}_{il})\epsilon_i\}^2\big]^{1/2},$$

$$\Delta_3^* = \big\{\max_{l\le h_n s_n+k_n} n^{-1}\sum_{i=1}^n(w_l'F_i - \tilde{E}_{il})^2(\sum_{j=1}^{q_n}\sum_{k=s_n+1}^{\infty}\theta_{ijk}\eta_{jk})^2\big\}^{1/2},$$

$$\Delta_4^* = \big[\max_{l\le h_n s_n+k_n} n^{-1}\sum_{i=1}^n\{(\hat{w}_l - w_l)'F_i(\sum_{j=1}^{q_n}\sum_{k=s_n+1}^{\infty}\theta_{ijk}\eta_{jk})\}^2\big]^{1/2},$$

$$\Delta_5^* = \big\{\max_{l\le h_n s_n+k_n} n^{-1}\sum_{i=1}^n(\tilde{E}_{il} - \breve{E}_{il})^2(\sum_{j=1}^{q_n}\sum_{k=s_n+1}^{\infty}\theta_{ijk}\eta_{jk})^2\big\}^{1/2},$$

$$\Delta_6^* = \big[\max_{l\le h_n s_n+k_n} n^{-1}\sum_{i=1}^n\{(w_l'F_i - \tilde{E}_{il})G_i'\big(\begin{smallmatrix}\hat{\eta}-\eta\\ \hat{\gamma}-\gamma\end{smallmatrix}\big)\}^2\big]^{1/2},$$

$$\Delta_7^* = \big[\max_{l\le h_n s_n+k_n} n^{-1}\sum_{i=1}^n\{(\hat{w}_l - w_l)'F_iG_i'\big(\begin{smallmatrix}\hat{\eta}-\eta\\ \hat{\gamma}-\gamma\end{smallmatrix}\big)\}^2\big]^{1/2},$$

$$\Delta_8^* = \big[\max_{l\le h_n s_n+k_n} n^{-1}\sum_{i=1}^n\{(\tilde{E}_{il} - \breve{E}_{il})G_i'\big(\begin{smallmatrix}\hat{\eta}-\eta\\ \hat{\gamma}-\gamma\end{smallmatrix}\big)\}^2\big]^{1/2}.$$

To bound $\Delta_1^*$, note that

$$\begin{aligned}
\Delta_1^* \le& \max_{l\le h_n s_n+k_n}\max_{i\le n}|[diag\{\hat{\Lambda}^{1/2}_{\mathcal{H}_n^c}, I_{(d_n-k_n)}\}(\hat{w}_l - w_l)]'\breve{F}_i\epsilon_i| \\
\le& \max_{i\le n}\|\breve{F}_i\epsilon_i\|_\infty \cdot \max_{l\le h_n s_n+k_n}\|diag\{\hat{\Lambda}^{1/2}_{\mathcal{H}_n^c}, I_{(d_n-k_n)}\}(\hat{w}_l - w_l)\|_1 \\
\le& (2+\|\Lambda^{1/2}\hat{\Lambda}^{-1/2} - I\|_\infty)(\max_{j\le p_n s_n+d_n}\max_{i\le n}|\tilde{G}_{ij}|)(\max_{i\le n}|\epsilon_i|) \\
&\cdot \max_{l\le h_n s_n+k_n}\|diag\{\hat{\Lambda}^{1/2}_{\mathcal{H}_n^c}, I_{(d_n-k_n)}\}(\hat{w}_l - w_l)\|_1 \\
\le& c_{10}n^{-1/2}\rho_n\log^{3/2}\{n(p_ns_n+d_n)\},
\end{aligned} \tag{S2.51}$$

for some universal constant $c_{10} > 0$, with probability tending to 1, where the last

inequality is based on Lemma 2, and Lemma 6. To bound $\Delta_2^*$, note that

$$\begin{aligned}\Delta_2^* \leq & \max_{l\leq h_n s_n+k_n}\max_{i\leq n}|(\tilde{E}_{il}-\breve{E}_{il})\epsilon_i| \\ \leq & \|\Lambda^{1/2}\hat{\Lambda}^{-1/2}-I\|_\infty(\max_{j\leq p_n s_n+d_n}\max_{i\leq n}|\tilde{G}_{ij}|)(\max_{i\leq n}|\epsilon_i|) \\ \leq & c_{11}n^{-1/2}\log^{3/2}\{n(p_n s_n+d_n)\}, \end{aligned} \tag{S2.52}$$

for some universal constant $c_{11}>0$, with probability tending to 1, where the last inequality is based on Lemma 2. To bound $\Delta_3^*$, note that

$$\begin{aligned}\Delta_3^* \leq & n^{-1/2}\max_{l\leq h_n s_n+k_n}\max_{i\leq n}|w_l'F_i-\tilde{E}_{il}|\cdot\{\sum_{i=1}^n(\sum_{j=1}^{q_n}\sum_{k=s_n+1}^{\infty}\theta_{ijk}\eta_{jk})^2\}^{1/2} \\ \leq & c_{12}q_n s_n^{-\delta}\log^{1/2}\{n(p_n s_n+d_n)\}, \end{aligned} \tag{S2.53}$$

for some universal constant $c_{12}>0$, with probability tending to 1, where the last inequality is based on Lemma 2, and (S2.43). To bound $\Delta_4^*$, note that

$$\begin{aligned}\Delta_4^* \leq & n^{-1/2}\max_{l\leq h_n s_n+k_n}\max_{i\leq n}|(\hat{w}_l-w_l)'F_i|\cdot\{\sum_{i=1}^n(\sum_{j=1}^{q_n}\sum_{k=s_n+1}^{\infty}\theta_{ijk}\eta_{jk})^2\}^{1/2} \\ \leq & n^{-1/2}(2+\|\Lambda^{1/2}\hat{\Lambda}^{-1/2}-I\|_\infty)(\max_{j\leq p_n s_n+d_n}\max_{i\leq n}|\tilde{G}_{ij}|)\cdot \\ & \max_{l\leq h_n s_n+k_n}\|diag\{\hat{\Lambda}^{1/2}_{\mathcal{H}_n^c},I_{(d_n-k_n)}\}(\hat{w}_l-w_l)\|_1\cdot\{\sum_{i=1}^n(\sum_{j=1}^{q_n}\sum_{k=s_n+1}^{\infty}\theta_{ijk}\eta_{jk})^2\}^{1/2} \\ \leq & c_{13}n^{-1/2}\rho_n q_n s_n^{-\delta}\log\{n(p_n s_n+d_n)\}, \end{aligned} \tag{S2.54}$$

for some universal constant $c_{13}>0$, with probability tending to 1, where the last

inequality is by Lemma 2, Lemma 6, and (S2.43). To bound $\Delta_5^*$, note that

$$\begin{aligned}\Delta_5^* \leq & n^{-1/2} \max_{l \leq h_n s_n + k_n} \max_{i \leq n} |\tilde{E}_{il} - \breve{E}_{il}| \cdot \{\sum_{i=1}^{n} (\sum_{j=1}^{q_n} \sum_{k=s_n+1}^{\infty} \theta_{ijk} \eta_{jk})^2\}^{1/2} \\ \leq & n^{-1/2} \|\Lambda^{1/2} \hat{\Lambda}^{-1/2} - I\|_\infty (\max_{j \leq p_n s_n + d_n} \max_{i \leq n} |\tilde{G}_{ij}|) \cdot \{\sum_{i=1}^{n} (\sum_{j=1}^{q_n} \sum_{k=s_n+1}^{\infty} \theta_{ijk} \eta_{jk})^2\}^{1/2} \\ \leq & c_{14} n^{-1/2} q_n s_n^{-\delta} \log\{n(p_n s_n + d_n)\}, \end{aligned} \tag{S2.55}$$

for some universal constant $c_{14} > 0$, with probability tending to 1, where the last inequality is by Lemma 2, and (S2.43). To bound $\Delta_6^*$, note that

$$\begin{aligned}\Delta_6^* \leq & n^{-1/2} \max_{l \leq h_n s_n + k_n} \max_{i \leq n} |w_l' F_i - \tilde{E}_{il}| \cdot [\sum_{i=1}^{n} \{G_i' \left(\begin{smallmatrix} \hat{\eta} - \eta \\ \hat{\gamma} - \gamma \end{smallmatrix}\right)\}^2]^{1/2} \\ \leq & 2 \max_{l \leq h_n s_n + k_n} \max_{i \leq n} |w_l' F_i - \tilde{E}_{il}| \cdot (\max_{j \leq p_n s_n + d_n} \max_{i \leq n} |\tilde{G}_{ij}|) \cdot \max\left\{\|\Lambda^{1/2}(\hat{\eta} - \eta)\|_1, \|\hat{\gamma} - \gamma\|_1\right\} \\ \leq & c_{15} \lambda_n s_n^{1/2} (q_n + r_n) n^{-1/18} \log\{n(p_n s_n + d_n)\}, \end{aligned} \tag{S2.56}$$

for some universal constant $c_{15} > 0$, with probability tending to 1, where the last inequality is by Lemma 2, and Theorem 1. To bound $\Delta_7^*$, note that

$$\begin{aligned}\Delta_7^* \leq & n^{-1/2} \max_{l \leq h_n s_n + k_n} \max_{i \leq n} |(\hat{w}_l - w_l)' F_i| \cdot [\sum_{i=1}^{n} \{G_i' \left(\begin{smallmatrix} \hat{\eta} - \eta \\ \hat{\gamma} - \gamma \end{smallmatrix}\right)\}^2]^{1/2} \\ \leq & 2(2 + \|\Lambda^{1/2} \hat{\Lambda}^{-1/2} - I\|_\infty) \cdot \max_{l \leq h_n s_n + k_n} \|diag\{\hat{\Lambda}_{\mathcal{H}_n^c}^{1/2}, I_{(d_n - k_n)}\}(\hat{w}_l - w_l)\|_1 \cdot \\ & (\max_{j \leq p_n s_n + d_n} \max_{i \leq n} |\tilde{G}_{ij}|^2) \cdot \max\left\{\|\Lambda^{1/2}(\hat{\eta} - \eta)\|_1, \|\hat{\gamma} - \gamma\|_1\right\} \\ \leq & c_{16} \lambda_n \rho_n s_n^{1/2} (q_n + r_n) n^{-5/9} \log^{3/2}\{n(p_n s_n + d_n)\}, \end{aligned} \tag{S2.57}$$

for some universal constant $c_{16} > 0$, with probability tending to 1, where the last

inequality is by Lemma 2, Lemma 6, and Theorem 1. To bound $\Delta_8^*$, note that

$$\begin{aligned}\Delta_8^* \leq & n^{-1/2} \max_{l \leq h_n s_n + k_n} \max_{i \leq n} |\tilde{E}_{il} - \breve{E}_{il}| \cdot [\sum_{i=1}^{n} \{G_i'\big(\tbinom{\hat{\eta}-\eta}{\hat{\gamma}-\gamma}\big)\}^2]^{1/2} \\ \leq & 2\|\Lambda^{1/2}\hat{\Lambda}^{-1/2} - I\|_\infty \cdot (\max_{j \leq p_n s_n + d_n} \max_{i \leq n} |\tilde{G}_{ij}|^2) \cdot \max\big\{\|\Lambda^{1/2}(\hat{\eta}-\eta)\|_1, \|\hat{\gamma}-\gamma\|_1\big\} \\ \leq & c_{17}\lambda_n s_n^{1/2}(q_n + r_n) n^{-5/9} \log^{3/2}\{n(p_n s_n + d_n)\}, \end{aligned} \tag{S2.58}$$

for some universal constant $c_{17} > 0$, with probability tending to 1, where the last inequality is by Lemma 2, and Theorem 1. By combining (S2.51), (S2.52), (S2.53), (S2.54), (S2.55), (S2.56), (S2.57), (S2.58) with (S2.50), it is seen that there is a universal constant $c_{18} > 0$ such that with probability tending to 1:

$$\begin{aligned}\max_{l \leq h_n s_n + k_n} \{n^{-1} \sum_{i=1}^{n} (\hat{S}_{il} - S_{il}^*)^2\}^{1/2} \leq & c_{18}\big[\lambda_n s_n^{1/2}(q_n + r_n) n^{-1/18} \log\{n(p_n s_n + d_n)\} \\ & + n^{-1/2}\rho_n \log^{3/2}\{n(p_n s_n + d_n)\}\big].\end{aligned}$$

This completes the proof of part 2). □

## S3 Proofs of Main Theorems

*Proof of Theorem 1.* First note that with some abuse of notation, we write $(\eta^*, \gamma^*)$ as the true version of some estimator $(\eta, \gamma)$, and denotes their differences as $\nu = [(\eta - \eta^*)', (\gamma - \gamma^*)']'$ and $\tilde{\nu} = diag\{\Lambda^{1/2}, I_{d_n}\}\nu$. Based on the first order necessary condition of the optimization theory, any local minima $(\hat{\eta}, \hat{\gamma})$ of $Q_n(\eta, \gamma)$ from (2.3) in the main article must satisfy $(\hat{\eta}, \hat{\gamma}) \in \{(\eta, \gamma) : \langle \nabla L_n(\eta, \gamma) +$

$\nabla P_{\lambda_n}(\eta,\gamma),\nu\rangle \le 0,\ \ \|\eta\|_1+\|\gamma\|_1 \le B_n\}$. Accordingly, to show Theorem 1, it suffices to justify that any estimator $(\eta,\gamma)\in\{(\eta,\gamma):\langle\nabla L_n(\eta,\gamma)+\nabla P_{\lambda_n}(\eta,\gamma),\nu\rangle \le 0,\ \ \|\eta\|_1+\|\gamma\|_1 \le B_n\}$ satisfies parts 1) and 2) of Theorem 1. Therefore, we start the proof with an arbitrary estimator $(\eta,\gamma)$ satisfying

$$(\eta,\gamma)\in\{(\eta,\gamma):\langle\nabla L_n(\eta,\gamma)+\nabla P_{\lambda_n}(\eta,\gamma),\nu\rangle \le 0,\ \ \|\eta\|_1+\|\gamma\|_1 \le B_n\}. \tag{S3.1}$$

In addition, it can be verified that

$$\begin{aligned}
&\langle\nabla L_n(\eta,\gamma)-\nabla L_n(\eta^*,\gamma^*),\nu\rangle = \tilde\nu'\{n^{-1}(\Theta\Lambda^{-1/2},Z)'(\Theta\Lambda^{-1/2},Z)\}\tilde\nu\\
=&\tilde\nu' E(\tilde G_i\tilde G_i')\tilde\nu+\tilde\nu'\{n^{-1}(\Theta\Lambda^{-1/2},Z)'(\Theta\Lambda^{-1/2},Z)-E(\tilde G_i\tilde G_i')\}\tilde\nu\\
\ge&\lambda_{\min}(E(\tilde G_i\tilde G_i'))\|\tilde\nu\|_2^2-\|n^{-1}(\Theta\Lambda^{-1/2},Z)'(\Theta\Lambda^{-1/2},Z)-E(\tilde G_i\tilde G_i')\|_\infty\|\tilde\nu\|_1^2\\
\ge&c_1\|\tilde\nu\|_2^2-c_2[\log\{n(p_ns_n+d_n)\}/n]^{1/2}\|\tilde\nu\|_1^2,
\end{aligned} \tag{S3.2}$$

for some universal constants $c_1,c_2>0$, with probability tending to 1, where the last inequality follows from (A2.4) and Lemma 2. Denoting the function $P_{\lambda_n,\mu}(\eta,\gamma)=P_{\lambda_n}(\eta,\gamma)+2^{-1}\mu n^{-10/9}\{\sum_{j=1}^{p_n}\|\Theta_j\eta_j\|_2^2+\sum_{l=1}^{d_n}\sum_{i=1}^{n}Z_{il}^2\gamma_l^2\}$, it then follows from condition (B5) that $P_{\lambda_n,\mu}(\eta,\gamma)$ is convex in $(\eta,\gamma)$, which entails that $P_{\lambda_n,\mu}(\eta^*,\gamma^*)-P_{\lambda_n,\mu}(\eta,\gamma)\ge-\langle\nabla P_{\lambda_n,\mu}(\eta,\gamma),\nu\rangle$. This further implies

$$\begin{aligned}
-\langle\nabla P_{\lambda_n}(\eta,\gamma),\nu\rangle \le& 2^{-1}\mu n^{-10/9}\{\sum_{j=1}^{p_n}\|\Theta_j(\eta_j-\eta_j^*)\|_2^2+\sum_{l=1}^{d_n}\sum_{i=1}^{n}Z_{il}^2(\gamma_l-\gamma_l^*)^2\}\\
&+P_{\lambda_n}(\eta^*,\gamma^*)-P_{\lambda_n}(\eta,\gamma).
\end{aligned} \tag{S3.3}$$

Based on the above discussion, we have

$$
\begin{aligned}
&c_1\|\tilde{\nu}\|_2^2 - c_2[\log\{n(p_n s_n + d_n)\}/n]^{1/2}\|\tilde{\nu}\|_1^2 \le \langle \nabla L_n(\eta,\gamma) - \nabla L_n(\eta^*,\gamma^*), \nu\rangle \\
\le& - \langle \nabla P_{\lambda_n}(\eta,\gamma), \nu\rangle - \langle \nabla L_n(\eta^*,\gamma^*), \nu\rangle \\
\le& 2^{-1}\mu n^{-10/9}\{\sum_{j=1}^{p_n}\|\Theta_j(\eta_j-\eta_j^*)\|_2^2 + \sum_{l=1}^{d_n}\sum_{i=1}^{n} Z_{il}^2(\gamma_l-\gamma_l^*)^2\} \\
&+ P_{\lambda_n}(\eta^*,\gamma^*) - P_{\lambda_n}(\eta,\gamma) - \langle \nabla L_n(\eta^*,\gamma^*), \nu\rangle \\
\le& 2^{-1}\mu n^{-10/9}\{\sum_{j=1}^{p_n}\|\Theta_j(\eta_j-\eta_j^*)\|_2^2 + \sum_{l=1}^{d_n}\sum_{i=1}^{n} Z_{il}^2(\gamma_l-\gamma_l^*)^2\} \\
&+ P_{\lambda_n}(\eta^*,\gamma^*) - P_{\lambda_n}(\eta,\gamma) + \|n^{-1}(\Theta\Lambda^{-1/2}, Z)'(Y-\Theta\eta^*-Z\gamma^*)\|_\infty\|\tilde{\nu}\|_1 \\
\le& c_3 n^{-1/9}\|\tilde{\nu}\|_2^2 + c_3\big(q_n s_n^{-\delta} + [\log\{n(p_n s_n + d_n)\}/n]^{1/2}\big)\|\tilde{\nu}\|_1 \\
&+ P_{\lambda_n}(\eta^*,\gamma^*) - P_{\lambda_n}(\eta,\gamma),
\end{aligned}
\tag{S3.4}
$$

for some universal constants $c_3 > 0$, with probability tending to 1, where the first inequality follows from (S3.2), the second inequality holds from (S3.1), the third inequality is based on (S3.3), and the last inequality follows from Lemma 3 and Lemma 4. Some rearrangement of (S3.4) leads to

$$
\begin{aligned}
0 \le& (c_1 - c_3 n^{-1/9})\|\tilde{\nu}\|_2^2 \le c_4\big\{q_n s_n^{-\delta} + [\log\{n(p_n s_n + d_n)\}/n]^{1/2}(1+\|\tilde{\nu}\|_1)\big\}\|\tilde{\nu}\|_1 \\
&+ P_{\lambda_n}(\eta^*,\gamma^*) - P_{\lambda_n}(\eta,\gamma) \\
\le& c_5\big\{q_n s_n^{-\delta} + [B_n^2\log\{n(p_n s_n + d_n)\}/n]^{1/2}\big\}\|\tilde{\nu}\|_1 \\
&+ P_{\lambda_n}(\eta^*,\gamma^*) - P_{\lambda_n}(\eta,\gamma),
\end{aligned}
\tag{S3.5}
$$

for some universal constants $c_4, c_5 > 0$, with probability tending to 1, where the last inequality follows from the fact that $\|\tilde{\nu}\|_1 \lesssim B_n$. By combining (S3.5) with Lemma 4, (A3.1) and (A4.2), it can be deduced that $0 \leq \|\tilde{\nu}\|_2^2 \lesssim P_{\lambda_n}(\eta^*, \gamma^*) - P_{\lambda_n}(\eta, \gamma)$, with probability tending to 1. Together with part 4) of Lemma 1, we have that

$$\|\tilde{\nu}\|_2^2 \lesssim \lambda_n \Big\{ \sum_{l \in \mathcal{B}_n} n^{-5/9} (\sum_{i=1}^n Z_{il}^2)^{1/2} |\gamma_l - \gamma_l^*| - \sum_{l \in \mathcal{B}_n^c} n^{-5/9} (\sum_{i=1}^n Z_{il}^2)^{1/2} |\gamma_l - \gamma_l^*| \Big\} + \lambda_n \Big\{ \sum_{j \in \mathcal{A}_n} n^{-5/9} \|\Theta_j(\eta_j - \eta_j^*)\|_2 - \sum_{j \in \mathcal{A}_n^c} n^{-5/9} \|\Theta_j(\eta_j - \eta_j^*)\|_2 \Big\}, \quad \text{(S3.6)}$$

with probability tending to 1. On one hand, (S3.6) implies that

$$\begin{aligned} \|\tilde{\nu}\|_2^2 &\lesssim \lambda_n \Big\{ \sum_{j \in \mathcal{A}_n} n^{-5/9} \|\Theta_j(\eta_j - \eta_j^*)\|_2 + \sum_{l \in \mathcal{B}_n} n^{-5/9} (\sum_{i=1}^n Z_{il}^2)^{1/2} |\gamma_l - \gamma_l^*| \Big\} \\ &\lesssim \lambda_n n^{-5/9} (q_n + r_n)^{1/2} \Big\{ \sum_{j \in \mathcal{A}_n} \|\Theta_j(\eta_j - \eta_j^*)\|_2^2 + \sum_{l \in \mathcal{B}_n} (\sum_{i=1}^n Z_{il}^2) |\gamma_l - \gamma_l^*|^2 \Big\}^{1/2} \\ &\lesssim \lambda_n n^{-1/18} (q_n + r_n)^{1/2} \|\tilde{\nu}\|_2, \end{aligned}$$

with probability tending to 1, where the last inequality holds from part 1) of Lemma 4. This further entails that

$$\|\tilde{\nu}\|_2 \lesssim \lambda_n (q_n + r_n)^{1/2} n^{-1/18}, \quad \text{(S3.7)}$$

with probability tending to 1, which completes the proof of part 1). On the other hand, (S3.6) also implies that with probability tending to 1,

$$\begin{aligned}&\sum_{j\in\mathcal{A}_n^c}\|\Theta_j(\eta_j-\eta_j^*)\|_2+\sum_{l\in\mathcal{B}_n^c}(\sum_{i=1}^n Z_{il}^2)^{1/2}|\gamma_l-\gamma_l^*|\\ \le&\sum_{j\in\mathcal{A}_n}\|\Theta_j(\eta_j-\eta_j^*)\|_2+\sum_{l\in\mathcal{B}_n}(\sum_{i=1}^n Z_{il}^2)^{1/2}|\gamma_l-\gamma_l^*|. \qquad \text{(S3.8)}\end{aligned}$$

Finally, we have

$$\begin{aligned}\|\tilde{\nu}\|_1 &\lesssim n^{-1/2}\{s_n^{1/2}\sum_{j=1}^{p_n}\|\Theta_j(\eta_j-\eta_j^*)\|_2+\sum_{l=1}^{d_n}(\sum_{i=1}^n Z_{il}^2)^{1/2}|\gamma_l-\gamma_l^*|\}\\ &\lesssim n^{-1/2}s_n^{1/2}\{\sum_{j\in\mathcal{A}_n}\|\Theta_j(\eta_j-\eta_j^*)\|_2+\sum_{l\in\mathcal{B}_n}(\sum_{i=1}^n Z_{il}^2)^{1/2}|\gamma_l-\gamma_l^*|\}\\ &\lesssim n^{-1/2}s_n^{1/2}(q_n+r_n)^{1/2}\{\sum_{j\in\mathcal{A}_n}\|\Theta_j(\eta_j-\eta_j^*)\|_2^2+\sum_{l\in\mathcal{B}_n}(\sum_{i=1}^n Z_{il}^2)|\gamma_l-\gamma_l^*|^2\}^{1/2}\\ &\lesssim s_n^{1/2}(q_n+r_n)^{1/2}\|\tilde{\nu}\|_2\lesssim\lambda_n s_n^{1/2}(q_n+r_n)n^{-1/18},\end{aligned}$$

with probability tending to 1, where the first inequality holds from part 2) of Lemma 4, the second inequality is based on (S3.8), the fourth inequality follows from part 1) of Lemma 4, and the last inequality is due to (S3.7). This completes the proof of part 2). □

*Proof of Theorem 2.* Recall the four quantities

$$\hat{T}(\beta_{\mathcal{H}_n},\gamma_{\mathcal{K}_n})=n^{-1/2}\sum_{i=1}^n S_i,\qquad \hat{T}_e=n^{-1/2}\sum_{i=1}^n e_i\hat{S}_i,$$

$$T^*=n^{-1/2}\sum_{i=1}^n S_i^*,\qquad T_e^*=n^{-1/2}\sum_{i=1}^n e_iS_i^*,$$

where $S_i^* = (w'F_i - \tilde{E}_i)\epsilon_i$. Note that $\{S_i^* : i \leq n\}$ are centered independent random vectors such that

$$\begin{aligned}\min_{l \leq h_n s_n + k_n} n^{-1} \sum_{i=1}^n E(S_{il}^{*2}) &= \{\min_{l \leq h_n s_n + k_n} E(w_l' F_i - \tilde{E}_{il})^2\} \cdot n^{-1} \sum_{i=1}^n E(\epsilon_i^2) \\ &\geq \lambda_{\min}(E(\tilde{E}_i \tilde{E}_i') - w' E(F_i \tilde{E}_i')) \cdot n^{-1} \sum_{i=1}^n E(\epsilon_i^2) \\ &\geq \lambda_{\min}(E(\tilde{G}_i \tilde{G}_i')) \cdot n^{-1} \sum_{i=1}^n E(\epsilon_i^2) \geq c_1, \qquad \text{(S3.9)}\end{aligned}$$

for some universal constant $c_1 > 0$, where the last inequality is based on (A1.2) and (A2.4). In addition, it follows from (A1.1) that $\{S_{il}^* : i \leq n, l \leq h_n s_n + k_n\}$ are sub-exponential with

$$\sup_{k \geq 1} k^{-1} (E|S_{il}^*|^k)^{1/k} \leq c_2, \qquad \text{(S3.10)}$$

for some universal constant $c_2 > 0$. By combining (S3.9), (S3.10), (A3.1) with Lemma 5, we have

$$\lim_{n \to \infty} \sup_{A \in \mathcal{A}^{Re}} \left| P(T^* \in A) - P_e(T_e^* \in A) \right| = 0, \qquad \text{(S3.11)}$$

where the set $\mathcal{A}^{Re}$ is defined in Lemma 5. Since $\|\hat{T}(\beta_{\mathcal{H}_n}, \gamma_{\mathcal{K}_n}) - T^*\|_\infty = \|n^{-1/2} \sum_{i=1}^n (S_i - S_i^*)\|_\infty$, it follows from Lemma 7 that there exists a universal constant $c_3 > 0$ such that

$$P(\|\hat{T}(\beta_{\mathcal{H}_n}, \gamma_{\mathcal{K}_n}) - T^*\|_\infty \geq c_3 f_n) \to 0, \qquad \text{(S3.12)}$$

where $f_n = \lambda_n s_n^{1/2}(q_n + r_n) n^{-1/18} \log^{1/2}\{n(p_n s_n + d_n)\} + n^{1/2} q_n s_n^{-\delta} \log^{1/2}\{n(p_n s_n +$

$d_n)\}+n^{-1/2}\rho_n \log\{n(p_n s_n+d_n)\}$. To bound $\|\hat{T}_e-T_e^*\|_\infty$, note that for any $t>0$,

$$
\begin{aligned}
&P_e(\|\hat{T}_e - T_e^*\|_\infty \geq t) = P_e(\max_{l\leq h_n s_n+k_n} |n^{-1/2}\sum_{i=1}^n e_i(\hat{S}_{il}-S_{il}^*)| \geq t)\\
\leq& \sum_{l=1}^{h_n s_n+k_n} P_e(|n^{-1/2}\sum_{i=1}^n e_i(\hat{S}_{il}-S_{il}^*)| \geq t)\\
\leq& 2\sum_{l=1}^{h_n s_n+k_n} \exp\big[-\{2n^{-1}\sum_{i=1}^n(\hat{S}_{il}-S_{il}^*)^2\}^{-1}t^2\big]\\
\leq& 2(p_n s_n+d_n)\exp\big[-\{2\max_{l\leq h_n s_n+k_n} n^{-1}\sum_{i=1}^n(\hat{S}_{il}-S_{il}^*)^2\}^{-1}t^2\big]
\end{aligned}
$$

where the first inequality is by union bound inequality, and the second inequality is based on Hoeffding inequality. Plugging $t=\{2\max_{l\leq h_n s_n+k_n} n^{-1}\sum_{i=1}^n(\hat{S}_{il}-S_{il}^*)^2\}^{1/2}\log^{1/2}\{n(p_n s_n+d_n)\}$ into the above inequality yields

$$
P_e(\|\hat{T}_e - T_e^*\|_\infty \geq \{2\max_{l\leq h_n s_n+k_n} n^{-1}\sum_{i=1}^n(\hat{S}_{il}-S_{il}^*)^2\}^{1/2}\log^{1/2}\{n(p_n s_n+d_n)\})
$$

$\leq 2n^{-1}\to 0$.

Together with Lemma 7, there exists a universal constant $c_4>0$ such that

$$
P_e(\|\hat{T}_e - T_e^*\|_\infty \geq c_4 g_n) \xrightarrow{p} 0,
$$

where $g_n=\lambda_n s_n^{1/2}(q_n+r_n)n^{-1/18}\log^{3/2}\{n(p_n s_n+d_n)\}+n^{-1/2}\rho_n\log^2\{n(p_n s_n+d_n)\}$. Together with (S3.12), there exists a universal constant $c_5>0$ such that

$$
\begin{aligned}
&P(\|\hat{T}(\beta_{\mathcal{H}_n},\gamma_{\mathcal{K}_n})-T^*\|_\infty \geq c_5 a_n)\to 0,\\
&P_e(\|\hat{T}_e - T_e^*\|_\infty \geq c_5 a_n) \xrightarrow{p} 0, \qquad \text{(S3.13)}
\end{aligned}
$$

where $a_n = \lambda_n s_n^{1/2}(q_n+r_n)n^{-1/18}\log^{3/2}\{n(p_n s_n+d_n)\}+n^{-1/2}\rho_n \log^2\{n(p_n s_n+d_n)\} + n^{1/2}q_n s_n^{-\delta}\log^{1/2}\{n(p_n s_n + d_n)\}$. Under (A3.3), (A4.1) and (A4.4), we have

$$a_n^2\{1+\log(h_n s_n + k_n) - \log a_n\} \to 0. \tag{S3.14}$$

By combining (S3.9), (S3.10), (S3.11), (S3.13), (S3.14) with Lemma 5, it can be concluded that

$$\begin{aligned}
&\lim_{n\to\infty}\sup_{A\in\mathcal{A}^{Re}}\left|P(\hat{T}(\beta_{\mathcal{H}_n},\gamma_{\mathcal{K}_n})\in A) - P_e(\hat{T}_e\in A)\right| \\
&= \lim_{n\to\infty}\sup_{t\geq 0}\left|P(\|\hat{T}(\beta_{\mathcal{H}_n},\gamma_{\mathcal{K}_n})\|_\infty \leq t) - P_e(\|\hat{T}_e\|_\infty \leq t)\right| = 0,
\end{aligned} \tag{S3.15}$$

which completes the proof. □

*Proof of Theorem 3.* Given the true $(\beta_{\mathcal{H}_n},\gamma_{\mathcal{K}_n})$, we have

$$\begin{aligned}
&\lim_{n\to\infty}\left|Power(\beta_{\mathcal{H}_n},\gamma_{\mathcal{K}_n}) - Power^*(\beta_{\mathcal{H}_n},\gamma_{\mathcal{K}_n})\right| \\
&= \lim_{n\to\infty}\Big|P\{\|\hat{T}(\beta_{\mathcal{H}_n},\gamma_{\mathcal{K}_n}) + n^{-1/2}\sum_{i=1}^n\big(\hat{w}'F_i - \breve{E}_i\big)E_i'\big(^{F_{\{b_k:k\leq s_n\}}(\beta_{\mathcal{H}_n})}_{\gamma_{\mathcal{K}_n}}\big)\|_\infty \leq c_B(\alpha)\} - \\
&\quad P_{e^*}\{\|\hat{T}_{e^*} + n^{-1/2}\sum_{i=1}^n\big(\hat{w}'F_i - \breve{E}_i\big)E_i'\big(^{F_{\{b_k:k\leq s_n\}}(\beta_{\mathcal{H}_n})}_{\gamma_{\mathcal{K}_n}}\big)\|_\infty \leq c_B(\alpha)\}\Big| \\
&\leq \lim_{n\to\infty}\sup_{A\in\mathcal{A}^{Re}}\left|P(\hat{T}(\beta_{\mathcal{H}_n},\gamma_{\mathcal{K}_n})\in A) - P_{e^*}(\hat{T}_{e^*}\in A)\right| = 0,
\end{aligned}$$

where the last equality is by (S3.15). This completes the proof. □

*Proof of Theorem 4.* First of all, it follows from triangle inequality that

$$Power^*(\beta_{\mathcal{H}_n}, \gamma_{\mathcal{K}_n}) = P_{e^*}\{\|\hat{T}_{e^*} + n^{-1/2}\sum_{i=1}^{n}\big(\hat{w}'F_i - \breve{E}_i\big)E_i'\big(\tbinom{\eta_{\mathcal{H}_n}}{\gamma_{\mathcal{K}_n}}\big)\|_\infty > c_B(\alpha)\}$$
$$\geq 1 - P_{e^*}\{\|\hat{T}_{e^*}\|_\infty \geq \|n^{-1/2}\sum_{i=1}^{n}\big(\hat{w}'F_i - \breve{E}_i\big)E_i'\big(\tbinom{\eta_{\mathcal{H}_n}}{\gamma_{\mathcal{K}_n}}\big)\|_\infty - c_B(\alpha)\}. \quad \text{(S3.16)}$$

Moreover, we have that for any $t > 0$,

$$P_{e^*}(\|\hat{T}_{e^*}\|_\infty \geq t) = P_{e^*}(\max_{l\leq h_n s_n + k_n}|n^{-1/2}\sum_{i=1}^{n} e_i^*\hat{S}_{il}| \geq t)$$
$$\leq \sum_{l=1}^{h_n s_n + k_n} P_{e^*}(|n^{-1/2}\sum_{i=1}^{n} e_i^*\hat{S}_{il}| \geq t) \leq 2\sum_{l=1}^{h_n s_n + k_n}\exp[-\{2n^{-1}\sum_{i=1}^{n}\hat{S}_{il}^2\}^{-1}t^2]$$
$$\leq 2(p_n s_n + d_n)\exp[-\{2\max_{l\leq h_n s_n + k_n} n^{-1}\sum_{i=1}^{n}\hat{S}_{il}^2\}^{-1}t^2], \quad \text{(S3.17)}$$

where the second inequality is by Hoeffding inequality. To bound the term $\max_{l\leq h_n s_n + k_n} n^{-1}\sum_{i=1}^{n}\hat{S}_{il}^2$, first note that

$$\max_{l\leq h_n s_n + k_n} n^{-1}\sum_{i=1}^{n}\hat{S}_{il}^2 \leq 2\big\{\max_{l\leq h_n s_n + k_n} n^{-1}\sum_{i=1}^{n}(\hat{S}_{il} - S_{il}^*)^2 + \max_{l\leq h_n s_n + k_n} n^{-1}\sum_{i=1}^{n} S_{il}^{*2}\big\}. \quad \text{(S3.18)}$$

To bound the term $\max_{l\leq h_n s_n + k_n} n^{-1}\sum_{i=1}^{n} S_{il}^{*2}$, note that

$$\max_{l\leq h_n s_n + k_n} n^{-1}\sum_{i=1}^{n} S_{il}^{*2} \leq \Delta_1 + \Delta_2 + \Delta_3, \quad \text{(S3.19)}$$

where

$$\Delta_1 = \max_{l \le h_n s_n + k_n} |n^{-1} \sum_{i=1}^{n} (w_l' F_i - \tilde{E}_{il})^2 \{\epsilon_i^2 - E(\epsilon_i^2)\}|,$$

$$\Delta_2 = \max_{l \le h_n s_n + k_n} |n^{-1} \sum_{i=1}^{n} E(\epsilon_i^2) \{(w_l' F_i - \tilde{E}_{il})^2 - E(w_l' F_i - \tilde{E}_{il})^2\}|,$$

$$\Delta_3 = \max_{l \le h_n s_n + k_n} n^{-1} \sum_{i=1}^{n} E(\epsilon_i^2) E(w_l' F_i - \tilde{E}_{il})^2.$$

For $\Delta_3$, it follows from (A1.1) that

$$\Delta_3 \le c_1, \tag{S3.20}$$

for some universal constant $c_1 > 0$. For $\Delta_2$, it follows from union bound inequality and Bernstein inequality that there exists a universal constant $c_2 > 0$ such that with probability tending to 1:

$$\Delta_2 \le c_2 [\log\{n(p_n s_n + d_n)\}/n]^{1/2}. \tag{S3.21}$$

For $\Delta_1$, conditional on $\{G_i : i \le n\}$, we have that for any $t > 0$,

$$\begin{aligned}
&P(\Delta_1 \ge t | \{G_i : i \le n\}) \\
\le& \sum_{l=1}^{h_n s_n + k_n} P(|n^{-1} \sum_{i=1}^{n} (w_l' F_i - \tilde{E}_{il})^2 \{\epsilon_i^2 - E(\epsilon_i^2)\}| \ge t | \{G_i : i \le n\}) \\
\le& 2(p_n s_n + d_n) \exp\Big[ - c_3 n \min\Big\{ \frac{t^2}{\max_{l \le h_n s_n + k_n} \max_{i \le n} (w_l' F_i - \tilde{E}_{il})^4}, \\
&\qquad \frac{t}{\max_{l \le h_n s_n + k_n} \max_{i \le n} (w_l' F_i - \tilde{E}_{il})^2} \Big\} \Big],
\end{aligned}$$

for some universal constant $c_3 > 0$, where the last inequality is by Bernstein inequality. Plugging $t = c_3^{-1/2} [\log\{n(p_n s_n + d_n)\}/n]^{1/2} \max_{l \le h_n s_n + k_n} \max_{i \le n} (w_l' F_i -$

$\tilde{E}_{il})^2$ into the above inequality yields

$$P(\Delta_1 \geq c_3^{-1/2}[\log\{n(p_n s_n + d_n)\}/n]^{1/2} \max_{l \leq h_n s_n + k_n} \max_{i \leq n} (w_l' F_i - \tilde{E}_{il})^2 | \{G_i : i \leq n\}) \leq 2n^{-1}.$$

Taking expectation on both sides of the above inequality yields that with probability tending to 1:

$$\Delta_1 \leq c_3^{-1/2}[\log\{n(p_n s_n + d_n)\}/n]^{1/2} \max_{l \leq h_n s_n + k_n} \max_{i \leq n} (w_l' F_i - \tilde{E}_{il})^2.$$

Together with Lemma 2, there exists a universal constant $c_4 > 0$ such that with probability tending to 1:

$$\Delta_1 \leq c_4[\log^3\{n(p_n s_n + d_n)\}/n]^{1/2}. \tag{S3.22}$$

By combining (S3.22), (S3.21), (S3.20) with (S3.19), we have that with probability tending to 1:

$$\max_{l \leq h_n s_n + k_n} n^{-1} \sum_{i=1}^{n} S_{il}^{*2} \leq 2c_1.$$

Together with (S3.18) and Lemma 7, there exists a universal constant $c_5 > 0$ such that with probability tending to 1:

$$\max_{l \leq h_n s_n + k_n} n^{-1} \sum_{i=1}^{n} \hat{S}_{il}^2 \leq c_5. \tag{S3.23}$$

Plugging $t = c_B(\alpha)$ into (S3.17) yields

$$c_B(\alpha) \leq \{4 \log(p_n s_n + d_n) \cdot \max_{l \leq h_n s_n + k_n} n^{-1} \sum_{i=1}^{n} \hat{S}_{il}^2\}^{1/2}.$$

Together with (S3.23), there exists a universal constant $c_6 > 0$ such that with probability tending to 1:

$$c_B(\alpha) \leq c_6 \log^{1/2}(p_n s_n + d_n). \tag{S3.24}$$

To bound the term $\|n^{-1/2}\sum_{i=1}^n \big(\hat{w}'F_i - \breve{E}_i\big)E_i'\big(\begin{smallmatrix}\eta_{\mathcal{H}_n}\\ \gamma_{\mathcal{K}_n}\end{smallmatrix}\big)\|_\infty$, first note that

$$\|n^{-1/2}\sum_{i=1}^n \big(\hat{w}'F_i - \breve{E}_i\big)E_i'\big(\begin{smallmatrix}\eta_{\mathcal{H}_n}\\ \gamma_{\mathcal{K}_n}\end{smallmatrix}\big)\|_\infty \geq \Pi_1 - \Pi_2, \tag{S3.25}$$

where

$$\Pi_1 = n^{1/2}\|\{E(\tilde{E}_i\tilde{E}_i') - w'E(F_i\tilde{E}_i')\}\big(\begin{smallmatrix}\Lambda_{\mathcal{H}_n}^{1/2}\eta_{\mathcal{H}_n}\\ \gamma_{\mathcal{K}_n}\end{smallmatrix}\big)\|_\infty,$$

$$\Pi_2 = n^{1/2}\|\big[n^{-1}\sum_{i=1}^n \big(\hat{w}'F_i - \breve{E}_i\big)\tilde{E}_i' - \{w'E(F_i\tilde{E}_i') - E(\tilde{E}_i\tilde{E}_i')\}\big]\cdot\big(\begin{smallmatrix}\Lambda_{\mathcal{H}_n}^{1/2}\eta_{\mathcal{H}_n}\\ \gamma_{\mathcal{K}_n}\end{smallmatrix}\big)\|_\infty.$$

For $\Pi_1$, it follows from the definition of $\mathcal{F}_n$ that

$$\Pi_1 \geq K\rho_n(q_n + r_n)\log^{1/2}\{n(p_n s_n + d_n)\}. \tag{S3.26}$$

For $\Pi_2$, it is not hard to verify that

$$\Pi_2 \leq n^{1/2}(\Omega_1 + \Omega_2)\Omega_3, \tag{S3.27}$$

where

$$\Omega_1 = \|n^{-1}\sum_{i=1}^n \breve{E}_i\tilde{E}_i' - E(\tilde{E}_i\tilde{E}_i')\|_\infty,$$

$$\Omega_2 = \|n^{-1}\sum_{i=1}^n \hat{w}'F_i\tilde{E}_i' - w'E(F_i\tilde{E}_i')\|_\infty,$$

$$\Omega_3 = \|\Lambda_{\mathcal{H}_n}^{1/2}\eta_{\mathcal{H}_n}\|_1 + \|\gamma_{\mathcal{K}_n}\|_1.$$

To bound $\Omega_1$, note that

$$\begin{aligned}\Omega_1 =&\|diag\{\hat{\Lambda}_{\mathcal{H}_n}^{-1/2}\Lambda_{\mathcal{H}_n}^{1/2}, I_{k_n}\}\{n^{-1}\sum_{i=1}^n \tilde{E}_i\tilde{E}_i' - E(\tilde{E}_i\tilde{E}_i')\}+\\ &diag\{\hat{\Lambda}_{\mathcal{H}_n}^{-1/2}\Lambda_{\mathcal{H}_n}^{1/2} - I, 0_{k_n\times k_n}\}E(\tilde{E}_i\tilde{E}_i')\|_\infty\\ \leq&(2+\|\Lambda^{1/2}\hat{\Lambda}^{-1/2} - I\|_\infty)\cdot\|n^{-1}\sum_{i=1}^n \tilde{E}_i\tilde{E}_i' - E(\tilde{E}_i\tilde{E}_i')\|_\infty+\\ &\|\Lambda^{1/2}\hat{\Lambda}^{-1/2} - I\|_\infty\cdot\|E(\tilde{E}_i\tilde{E}_i')\|_\infty\\ \leq&c_7[\log\{n(p_ns_n+d_n)\}/n]^{1/2},\end{aligned} \tag{S3.28}$$

for some universal constant $c_7 > 0$, with probability tending to 1, where the last inequality is based on Lemma 2 and (A1.1). To bound $\Omega_2$, note that

$$\begin{aligned}\Omega_2 =&\|[diag\{\hat{\Lambda}_{\mathcal{H}_n^c}^{1/2}, I_{(d_n-k_n)}\}(\hat{w}-w)]'diag\{\hat{\Lambda}_{\mathcal{H}_n^c}^{-1/2}\Lambda_{\mathcal{H}_n^c}^{1/2}, I_{(d_n-k_n)}\}n^{-1}\sum_{i=1}^n \tilde{F}_i\tilde{E}_i'\\ &+n^{-1}\sum_{i=1}^n w'F_i\tilde{E}_i' - E(w'F_i\tilde{E}_i')\|_\infty\\ \leq&\max_{l\leq h_ns_n+k_n}\|diag\{\hat{\Lambda}_{\mathcal{H}_n^c}^{1/2}, I_{(d_n-k_n)}\}(\hat{w}_l-w_l)\|_1\|diag\{\hat{\Lambda}_{\mathcal{H}_n^c}^{-1/2}\Lambda_{\mathcal{H}_n^c}^{1/2}, I_{(d_n-k_n)}\}\|_\infty\|n^{-1}\sum_{i=1}^n \tilde{F}_i\tilde{E}_i'\|_\infty\\ &+\|n^{-1}\sum_{i=1}^n w'F_i\tilde{E}_i' - E(w'F_i\tilde{E}_i')\|_\infty\\ \leq&(2+\|\Lambda^{1/2}\hat{\Lambda}^{-1/2} - I\|_\infty)\cdot\{\|n^{-1}\sum_{i=1}^n \tilde{G}_i\tilde{G}_i' - E(\tilde{G}_i\tilde{G}_i')\|_\infty + \|E(\tilde{G}_i\tilde{G}_i')\|_\infty\}\cdot\\ &\max_{l\leq h_ns_n+k_n}\|diag\{\hat{\Lambda}_{\mathcal{H}_n^c}^{1/2}, I_{(d_n-k_n)}\}(\hat{w}_l-w_l)\|_1 + \|n^{-1}\sum_{i=1}^n w'F_i\tilde{E}_i' - E(w'F_i\tilde{E}_i')\|_\infty\\ \leq&c_8\rho_n[\log\{n(p_ns_n+d_n)\}/n]^{1/2},\end{aligned} \tag{S3.29}$$

for some universal constant $c_8 > 0$, with probability tending to 1, where the last

inequality is by Lemma 2, Lemma 6, and (A1.1). To bound $\Omega_3$, note that

$$\begin{aligned}\Omega_3 \leq & \sum_{j=1}^{q_n}\sum_{k=1}^{s_n}\omega_{jk}^{1/2}k^{-\delta}|\eta_{jk}|k^{\delta}+\sum_{l=1}^{r_n}|\gamma_l| \\ \leq & \sum_{j=1}^{q_n}(\sum_{k=1}^{s_n}\omega_{jk}k^{-2\delta})^{1/2}(\sum_{k=1}^{s_n}\eta_{jk}^2k^{2\delta})^{1/2}+r_n\|\gamma\|_\infty \\ \leq & c_9(q_n+r_n), \end{aligned} \tag{S3.30}$$

for some universal constant $c_9 > 0$, where the last inequality is by (A2.1), (A2.2), and (A2.3). By combining (S3.28), (S3.29), (S3.30) with (S3.27), there exists a universal constant $c_{10} > 0$ such that with probability tending to 1:

$$\Pi_2 \leq c_{10}\rho_n(q_n+r_n)\log^{1/2}\{n(p_ns_n+d_n)\}. \tag{S3.31}$$

By choosing $K \geq 2c_{10}$ in $\mathcal{F}_n$, it follows from (S3.26), (S3.31), and (S3.25) that with probability tending to 1:

$$\|n^{-1/2}\sum_{i=1}^{n}\big(\hat{w}'F_i-\breve{E}_i\big)E_i'\big(\begin{smallmatrix}\eta_{\mathcal{H}_n}\\ \gamma_{\mathcal{K}_n}\end{smallmatrix}\big)\|_\infty \geq c_{10}\rho_n(q_n+r_n)\log^{1/2}\{n(p_ns_n+d_n)\}.$$

Together with (S3.24) yields that with probability tending to 1:

$$\begin{aligned}&\|n^{-1/2}\sum_{i=1}^{n}\big(\hat{w}'F_i-\breve{E}_i\big)E_i'\big(\begin{smallmatrix}\eta_{\mathcal{H}_n}\\ \gamma_{\mathcal{K}_n}\end{smallmatrix}\big)\|_\infty - c_B(\alpha) \\ \geq & 2^{-1}c_{10}\rho_n(q_n+r_n)\log^{1/2}\{n(p_ns_n+d_n)\}. \end{aligned} \tag{S3.32}$$

Plugging $t = \|n^{-1/2}\sum_{i=1}^{n}\big(\hat{w}'F_i-\breve{E}_i\big)E_i'\big(\begin{smallmatrix}\eta_{\mathcal{H}_n}\\ \gamma_{\mathcal{K}_n}\end{smallmatrix}\big)\|_\infty - c_B(\alpha)$ into (S3.17) yields

that with probability tending to 1:

$$P_{e^*}\{\|\hat{T}_{e^*}\|_\infty \geq \|n^{-1/2}\sum_{i=1}^{n}\big(\hat{w}'F_i - \breve{E}_i\big)E_i'\big(\tbinom{\eta_{\mathcal{H}_n}}{\gamma_{\mathcal{K}_n}}\big)\|_\infty - c_B(\alpha)\}$$

$$\leq 2(p_n s_n + d_n)\exp\big[-\{2\max_{l\leq h_n s_n + k_n} n^{-1}\sum_{i=1}^{n}\hat{S}_{il}^2\}^{-1}\cdot$$

$$\{\|n^{-1/2}\sum_{i=1}^{n}\big(\hat{w}'F_i - \breve{E}_i\big)E_i'\big(\tbinom{\eta_{\mathcal{H}_n}}{\gamma_{\mathcal{K}_n}}\big)\|_\infty - c_B(\alpha)\}^2\big].$$

Together with (S3.32), (S3.23), and (S3.16), it can be deduced that with probability tending to 1:

$$Power^*(\beta_{\mathcal{H}_n}, \gamma_{\mathcal{K}_n}) \geq 1 - 2(p_n s_n + d_n)\{n(p_n s_n + d_n)\}^{-8^{-1}c_5^{-1}c_{10}^2\rho_n^2(q_n+r_n)^2}.$$

This completes the proof. □

## Bibliography

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

Chernozhukov, V., D. Chetverikov, and K. Kato (2017). Central limit theorems and bootstrap in high dimensions. *The Annals of Probability 45*, 2309–2352.

Loh, P.-L. and M. J. Wainwright (2015). Regularized m-estimators with nonconvexity: statistical and algorithmic theory for local optima. *Journal of Machine Learning Research 71*, 559–616.

Ning, Y. and H. Liu (2017). A general theory of hypothesis tests and confidence regions for sparse high dimensional models. *The Annals of Statistics 45*, 158–195.