Supplementary Material to "Inference on Large-scale Generalized Functional Linear Model"

*Kaijie Xue and Riquan Zhang**

Shanghai University of International Business and Economics

**corresponding author: zhangriquan@163.com*

S1 Notations

First of all, we present the notations for the entire paper and supplement. We write the inner product as $\langle u, v\rangle = u'v$ for any two vectors $u, v \in \mathbb{R}^p$. For a generic vector $u = (u_1, \ldots, u_p)' \in \mathbb{R}^p$, we denote its ℓ_q-norms by $\|u\|_q = (\sum_{l=1}^p |u_l|^q)^{1/q}$ for $1 \leq q < \infty$, $\|u\|_0 = card\{l : u_l \neq 0\}$, and $\|u\|_\infty = \max_{l\leq p} |u_l|$. Given any vector $u = (u_1, \ldots, u_p)' \in \mathbb{R}^p$ and subset $Q \subseteq \{1, \ldots, p\}$, we write $u_Q \in \mathbb{R}^{\text{card}(Q)}$ to denote the sub-vector as restricting u to Q. For a generic matrix $B = [b_{ij}]_{p\times q}$, we define its matrix element-wise max-norm by $\|B\|_\infty = \max_{i,j} |b_{ij}|$. If B is symmetric, we further let $\lambda_{\max}(B)$ and $\lambda_{\min}(B)$ to represent the maximum and minimum eigenvalues. For two sequences μ_n and v_n, we write $\mu_n \lesssim v_n$ if $\mu_n \leq k_1 v_n$ for a universal constant $k_1 > 0$, and similarly write $\mu_n \gtrsim v_n$ provided that $\mu_n \geq k_2 v_n$ for a universal constant $k_2 > 0$. To this end, we write $\mu_n \asymp v_n$ as long as $|\mu_n| \lesssim |v_n|$ and $|\mu_n| \gtrsim |v_n|$.

Notice that we let the index set $\mathbb{P}_n = \{1, \ldots, p_n\}$ to denote all predictors, and write an arbitrary nonzero subset as $\mathcal{H}_n \subseteq \mathbb{P}_n$ containing $|\mathcal{H}_n| = h_n > 0$ elements, whose complement is $\mathcal{H}_n^c = \mathbb{P}_n \setminus \mathcal{H}_n$. We write $Y = (Y_1, \ldots, Y_n)'$ as the response vector. We let $\eta_{\mathcal{H}_n}$ to denote the vector of attaching $\{\eta_j : j \in \mathcal{H}_n\}$ vertically in a column, whose estimator $\hat{\eta}_{\mathcal{H}_n}$ is defined analogously. Also, we write $\beta_{\mathcal{H}_n} = \{\beta_j : j \in \mathcal{H}_n\}$ to represent the collection of regression curves. We further denote $F_{\{b_k : k \leq s_n\}}(\beta_{\mathcal{H}_n}) = \eta_{\mathcal{H}_n}$ as the function of mapping $\beta_{\mathcal{H}_n}$ onto $\eta_{\mathcal{H}_n}$. We write the matrix $\Theta_{\mathcal{H}_n}$ by attaching $\{\Theta_j : j \in \mathcal{H}_n\}$ in a line, and let $\Theta = \Theta_{\mathbb{P}_n}$ for brevity. For each $j \leq p_n$, we write the moment estimator of the diagonal matrix $\Lambda_j = diag\{\omega_{j1}, \ldots, \omega_{js_n}\}$ by $\hat{\Lambda}_j = diag\{\hat{\omega}_{j1}, \ldots, \hat{\omega}_{js_n}\}$, with each $\hat{\omega}_{jk} = n^{-1} \sum_{i=1}^n \theta_{ijk}^2$. We then formulate the two diagonal matrices $\Lambda_{\mathcal{H}_n} = diag\{\Lambda_j : j \in \mathcal{H}_n\}$ and $\Lambda = diag\{\Lambda_j : j \in \mathbb{P}_n\}$, whose moment estimates are given by $\hat{\Lambda}_{\mathcal{H}_n} = diag\{\hat{\Lambda}_j : j \in \mathcal{H}_n\}$ and $\hat{\Lambda} = diag\{\hat{\Lambda}_j : j \in \mathbb{P}_n\}$. To this end, we present a series of matrices (expressed in row vectors) as

$$\Theta = (G_1, \ldots, G_n)', \Theta_{\mathcal{H}_n} = (E_1, \ldots, E_n)', \Theta_{\mathcal{H}_n^c} = (F_1, \ldots, F_n)', \tilde{\Theta} = \Theta\Lambda^{-1/2} = (\tilde{G}_1, \ldots, \tilde{G}_n)',$$

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

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

Several scaled-forms of the vector η are abbreviated by $\tilde{\eta} = \Lambda^{1/2}\eta$ and $\breve{\eta} = \hat{\Lambda}^{1/2}\eta$, and similarly for $\tilde{\eta}_{\mathcal{H}_n}$ and $\breve{\eta}_{\mathcal{H}_n}$. With some abuse of notation, we sometimes write η^* as the genuine version of η, and denotes their differences by $\nu = \eta - \eta^*$ and $\tilde{\nu} = \Lambda^{1/2}\nu$. Also recall that we denote an unknown matrix

w by

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

with each $w_j = (w_{j1}, \ldots, w_{j,(p_n-h_n)s_n})'$. We use the value $\rho_n = \sup_{j \le h_n s_n} \rho_{nj}$ ($\rho_{nj} = card\{l : w_{jl} \neq 0\}$) to stand for the degree of sparsity of w. Based on the penalized estimators $(\hat{\eta}, \hat{\alpha}_0)$ and $\hat{w}$ from (2.5) and (3.8) of the main article, we write a series of random vectors for $i = 1, \ldots, n$ by

$$\tilde{S}_i = (\tilde{S}_{i1}, \ldots, \tilde{S}_{i,h_n s_n})' = (w' F_i - \tilde{E}_i)\{Y_i - b'(\alpha_0 + G_i'\eta)\},$$

$$S_i = (S_{i1}, \ldots, S_{i,h_n s_n})' = (\hat{w}' F_i - \breve{E}_i)\{Y_i - b'(\hat{\alpha}_0 + E_i'\eta_{\mathcal{H}_n} + F_i'\hat{\eta}_{\mathcal{H}_n^c})\},$$

$$\hat{S}_i = (\hat{S}_{i1}, \ldots, \hat{S}_{i,h_n s_n})' = (\hat{w}' F_i - \breve{E}_i)\{Y_i - b'(\hat{\alpha}_0 + G_i'\hat{\eta})\},$$

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

$$S(\tilde{\eta}_{\mathcal{H}_n}; \eta_{\mathcal{H}_n^c}, \alpha_0) = n^{-1} \sum_{i=1}^n \tilde{S}_i, \quad \hat{S}(\hat{\Lambda}_{\mathcal{H}_n}^{1/2} \eta_{\mathcal{H}_n}; \hat{\eta}_{\mathcal{H}_n^c}, \hat{\alpha}_0) = n^{-1} \sum_{i=1}^n S_i,$$

$$\hat{S}(\hat{\Lambda}_{\mathcal{H}_n}^{1/2} \hat{\eta}_{\mathcal{H}_n}; \hat{\eta}_{\mathcal{H}_n^c}, \hat{\alpha}_0) = n^{-1} \sum_{i=1}^n \hat{S}_i, \quad \hat{T}(\beta_{\mathcal{H}_n}) = n^{-1/2} \sum_{i=1}^n S_i,$$

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

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

where $e = \{e_1, \ldots, e_n\}$ represents a collection of i.i.d. $N(0, 1)$, that are independent of the data. $P_e(\cdot)$ represents the conditional probability that only treats e as random. $c_B(\alpha)$ is defined as the $(1 - \alpha)$th quantile of $\|\hat{T}_e\|_\infty$. The next section contains the auxiliary lemmas with 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 $|\rho_\lambda'(t)| \le \lambda L$, *for any* $t \ne 0$.

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

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

4) Under conditions (B1)–(B5), if $P_{\lambda_n}(\eta^*) - P_{\lambda_n}(\eta) \ge 0$, *where* η^* *stands for the true version of* η *and* $P_{\lambda_n}(\eta) = \sum_{j=1}^{p_n} \rho_{\lambda_n}\big(n^{-5/9}\|\Theta_j\eta_j\|_2\big)$, *then one has*

$$0 \le P_{\lambda_n}(\eta^*) - P_{\lambda_n}(\eta) \le \lambda_n L\big\{\sum_{j\in\mathcal{A}_n} n^{-5/9}\|\Theta_j(\eta_j - \eta_j^*)\|_2 - \sum_{j\in\mathcal{A}_n^c} n^{-5/9}\|\Theta_j(\eta_j - \eta_j^*)\|_2\big\},$$

where the subset $\mathcal{A}_n \subseteq \mathcal{P}_n$ *denotes the index set corresponding to the largest* q_n *elements of* $\{n^{-5/9}\|\Theta_j(\eta_j - \eta_j^*)\|_2 : j \le p_n\}$ *in magnitude, with* $\mathcal{A}_n^c = \mathcal{P}_n \setminus \mathcal{A}_n$.

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

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

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

and

$$\sum_{j\in\mathcal{A}_n} n^{-5/9}\|\Theta_j(\eta_j-\eta_j^*)\|_2 = \sum_{j\in\mathcal{A}_n} \rho_{\lambda_n}(n^{-5/9}\|\Theta_j(\eta_j-\eta_j^*)\|_2) f_n(n^{-5/9}\|\Theta_j(\eta_j-\eta_j^*)\|_2)$$

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

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

$$0 \leq P_{\lambda_n}(\eta^*) - P_{\lambda_n}(\eta) = \sum_{j=1}^{q_n} \rho_{\lambda_n}(n^{-5/9}\|\Theta_j\eta_j^*\|_2) - \sum_{j=1}^{p_n} \rho_{\lambda_n}(n^{-5/9}\|\Theta_j\eta_j\|_2)$$

$$\leq \sum_{j=1}^{q_n} \rho_{\lambda_n}(n^{-5/9}\|\Theta_j(\eta_j-\eta_j^*)\|_2) - \sum_{j=q_n+1}^{p_n} \rho_{\lambda_n}(n^{-5/9}\|\Theta_j\eta_j\|_2)$$

$$\leq \sum_{j\in\mathcal{A}_n} \rho_{\lambda_n}(n^{-5/9}\|\Theta_j(\eta_j-\eta_j^*)\|_2) - \sum_{j\in\mathcal{A}_n^c} \rho_{\lambda_n}(n^{-5/9}\|\Theta_j(\eta_j-\eta_j^*)\|_2)$$

$$\leq \big\{f_n(\max_{j\in\mathcal{A}_n^c} n^{-5/9}\|\Theta_j(\eta_j-\eta_j^*)\|_2)\big\}^{-1}\big\{\sum_{j\in\mathcal{A}_n} n^{-5/9}\|\Theta_j(\eta_j-\eta_j^*)\|_2 - \sum_{j\in\mathcal{A}_n^c} n^{-5/9}\|\Theta_j(\eta_j-\eta_j^*)\|_2\big\}$$

$$\leq \lambda_n L\big\{\sum_{j\in\mathcal{A}_n} n^{-5/9}\|\Theta_j(\eta_j-\eta_j^*)\|_2 - \sum_{j\in\mathcal{A}_n^c} n^{-5/9}\|\Theta_j(\eta_j-\eta_j^*)\|_2\big\},$$

which completes the proof. □

Lemma 2. *Under conditions (A2.1), (A3.1) and (A4.1), denoting I as the identity matrix, 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) $\|n^{-1}\sum_{i=1}^n \tilde{G}_i\tilde{G}_i' - E(\tilde{G}_i\tilde{G}_i')\|_\infty \leq c_5\{\log(np_ns_n)/n\}^{1/2}$*, for some universal constant* $c_5 > 0$.

7) $\|n^{-1}\sum_{i=1}^n G_iG_i{}' - E(G_iG_i{}')\|_\infty \leq c_6\{\log(np_ns_n)/n\}^{1/2}$*, for some universal constant* $c_6 > 0$.

8) $\|n^{-1}\sum_{i=1}^n \breve{G}_i\breve{G}_i' - E(\tilde{G}_i\tilde{G}_i')\|_\infty \leq c_7\{\log(np_ns_n)/n\}^{1/2}$*, for some universal constant* $c_7 > 0$.

9) $\max_{j\leq h_ns_n} \|n^{-1}\sum_{i=1}^n \tilde{F}_i(\tilde{E}_{ij} - F_i'w_j)\|_\infty \leq c_8\{\log(np_ns_n)/n\}^{1/2}$*, for some universal constant* $c_8 > 0$.

10) $\max_{j\leq h_ns_n} \|n^{-1}\sum_{i=1}^n \tilde{F}_i(\breve{E}_{ij} - F_i'w_j)\|_\infty \leq c_9\{\log(np_ns_n)/n\}^{1/2}$*, for some universal constant* $c_9 > 0$.

11) $\max_{j\le h_n s_n}\|n^{-1}\sum_{i=1}^{n}\breve{F}_i(\breve{E}_{ij}-F_i'w_j)\|_\infty \le c_{10}\{\log(np_ns_n)/n\}^{1/2}$*, for some universal constant* $c_{10}>0$*.*

12) $\|n^{-1}\sum_{i=1}^{n}\tilde{G}_i\|_\infty \le c_{11}\{\log(np_ns_n)/n\}^{1/2}$*, for some universal constant* $c_{11}>0$*.*

13) $\|n^{-1}\sum_{i=1}^{n}G_i\|_\infty \le c_{12}\{\log(np_ns_n)/n\}^{1/2}$*, for some universal constant* $c_{12}>0$*.*

14) $\|n^{-1}\sum_{i=1}^{n}\breve{G}_i\|_\infty \le c_{13}\{\log(np_ns_n)/n\}^{1/2}$*, for some universal constant* $c_{13}>0$*.*

15) $\max_{j\le p_ns_n}\max_{i\le n}|\tilde{G}_{ij}| \le c_{14}\{\log(np_ns_n)\}^{1/2}$*, for some universal constant* $c_{14}>0$*.*

16) $\max_{j\le h_ns_n}\max_{i\le n}|{w_j}'F_i| \le c_{15}\{\log(np_ns_n)\}^{1/2}$*, for some universal constant* $c_{15}>0$*.*

17) $\max_{j\le h_ns_n}|n^{-1}\sum_{i=1}^{n}\{(\tilde{E}_{ij}-F_i'w_j)^2-E(\tilde{E}_{ij}-F_i'w_j)^2\}| \le c_{17}\{\log(np_ns_n)/n\}^{1/2}$*, for some universal constant* $c_{17}>0$*.*

18) $\max_{i\le n}|E_i'\eta_{\mathcal{H}_n}| \le c_{17}\{\log(nq_ns_n)\}^{1/2}\|\Lambda^{1/2}\eta\|_1$*, for some universal constant* $c_{17}>0$*. Note that* $\mathcal{H}_n$ *can be arbitrary subset of* $\{1,\ldots,p_n\}$*.*

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

$$P\big(\|\hat{\Lambda}\Lambda^{-1} - I\|_\infty \geq t\big) = P\big\{\max_{j\leq p_n}\max_{k\leq s_n}|n^{-1}\sum_{i=1}^{n}(\omega_{jk}^{-1}\theta_{ijk}^2 - 1)| \geq t\big\}$$
$$\leq \sum_{j=1}^{p_n}\sum_{k=1}^{s_n} P\big\{|n^{-1}\sum_{i=1}^{n}(\omega_{jk}^{-1}\theta_{ijk}^2 - 1)| \geq t\big\}$$
$$\leq 2p_n s_n \exp\{-n\min(c_1^{-2}t^2, c_1^{-1}t)\}, \tag{S2.3}$$

for some universal constant $c_1 > 0$, where the first inequality is by union bound inequality and the second inequality holds from (A2.1) and Bernstein's inequality. Plugging $t = c_1\{\log(np_n s_n)/n\}^{1/2}$ into (S2.3) yields

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

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

$$\lambda_{\min}(\hat{\Lambda}) = \lambda_{\min}(\Lambda\hat{\Lambda}\Lambda^{-1}) \geq \lambda_{\min}(\Lambda)\lambda_{\min}(\hat{\Lambda}\Lambda^{-1}) = \lambda_{\min}(\Lambda)\lambda_{\min}(\hat{\Lambda}\Lambda^{-1} - I + I)$$
$$\geq (1 - \|\hat{\Lambda}\Lambda^{-1} - I\|_\infty)\lambda_{\min}(\Lambda). \tag{S2.5}$$

By combining (S2.4), (S2.5), (A4.1) with $\lambda_{\min}(\Lambda) > 0$, it can be deduced that

$$P\{\lambda_{\min}(\hat{\Lambda}) > 0\} \geq 1 - 2n^{-1} \to 1, \tag{S2.6}$$

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

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

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

Together with (S2.4) and (S2.6), it is apparent that

$$P\big[\|\Lambda\hat{\Lambda}^{-1} - I\|_\infty \le 2c_1\{\log(np_ns_n)/n\}^{1/2}\big] \ge 1 - 4n^{-1} \to 1, \qquad \text{(S2.7)}$$

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

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

Together with (S2.4), it is clear that

$$P\big[\|\hat{\Lambda}^{1/2}\Lambda^{-1/2} - I\|_\infty \le c_1\{\log(np_ns_n)/n\}^{1/2}\big] \ge 1 - 2n^{-1} \to 1, \qquad \text{(S2.8)}$$

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

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

Together with (S2.6) and (S2.7), it is obvious that

$$P\big[\|\Lambda^{1/2}\hat{\Lambda}^{-1/2} - I\|_\infty \le 2c_1\{\log(np_ns_n)/n\}^{1/2}\big] \ge 1 - 6n^{-1} \to 1, \qquad \text{(S2.9)}$$

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

$$\begin{aligned}
&P\big(\|n^{-1}\sum_{i=1}^{n}\tilde{G}_i\tilde{G}_i' - E(\tilde{G}_i\tilde{G}_i')\|_\infty \geq t\big)\\
=&P\big[\max_{l_1\leq p_ns_n}\max_{l_2\leq p_ns_n}|n^{-1}\sum_{i=1}^{n}\{\tilde{G}_{il_1}\tilde{G}_{il_2} - E(\tilde{G}_{il_1}\tilde{G}_{il_2})\}| \geq t\big]\\
\leq&\sum_{l_1=1}^{p_ns_n}\sum_{l_2=1}^{p_ns_n}P\big[|n^{-1}\sum_{i=1}^{n}\{\tilde{G}_{il_1}\tilde{G}_{il_2} - E(\tilde{G}_{il_1}\tilde{G}_{il_2})\}| \geq t\big]\\
\leq&2(p_ns_n)^2\exp\{-n\min(c_2^{-2}t^2, c_2^{-1}t)\}, \qquad \text{(S2.10)}
\end{aligned}$$

for some universal constant $c_2 > 0$, where the last inequality follows from (A2.1) and Bernstein's inequality. Plugging $t = 2c_2\{\log(np_ns_n)/n\}^{1/2}$ into (S2.10) yields

$$\begin{aligned}
&P\Big(\|n^{-1}\sum_{i=1}^{n}\tilde{G}_i\tilde{G}_i' - E(\tilde{G}_i\tilde{G}_i')\|_\infty \leq 2c_2\{\log(np_ns_n)/n\}^{1/2}\Big)\\
\geq&1 - 2(p_ns_n)^{-2}n^{-4} \to 1, \qquad \text{(S2.11)}
\end{aligned}$$

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

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

for some universal constant $c_3 > 0$, where the last inequality is based on (A3.1).

Together with (S2.11), it is obvious that

$$
\begin{aligned}
&P\Big(\|n^{-1}\sum_{i=1}^{n} G_iG_i{}' - E(G_iG_i{}')\|_\infty \le 2c_2c_3\{\log(np_ns_n)/n\}^{1/2}\Big)\\
\ge& 1 - 2(p_ns_n)^{-2}n^{-4} \to 1, \qquad \text{(S2.12)}
\end{aligned}
$$

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

$$
\begin{aligned}
&\|n^{-1}\sum_{i=1}^{n} \breve{G}_i\breve{G}_i' - E(\tilde{G}_i\tilde{G}_i')\|_\infty\\
=&\|(\hat{\Lambda}^{-1/2}\Lambda^{1/2})\{n^{-1}\sum_{i=1}^{n} \tilde{G}_i\tilde{G}_i' - E(\tilde{G}_i\tilde{G}_i')\}(\hat{\Lambda}^{-1/2}\Lambda^{1/2})\\
&+ (\hat{\Lambda}^{-1/2}\Lambda^{1/2})\{E(\tilde{G}_i\tilde{G}_i')\}(\hat{\Lambda}^{-1/2}\Lambda^{1/2} - I) + (\hat{\Lambda}^{-1/2}\Lambda^{1/2} - I)\{E(\tilde{G}_i\tilde{G}_i')\}\|_\infty\\
\le&\|(\hat{\Lambda}^{-1/2}\Lambda^{1/2})\{n^{-1}\sum_{i=1}^{n} \tilde{G}_i\tilde{G}_i' - E(\tilde{G}_i\tilde{G}_i')\}(\hat{\Lambda}^{-1/2}\Lambda^{1/2})\|_\infty\\
&+ \|(\hat{\Lambda}^{-1/2}\Lambda^{1/2})\{E(\tilde{G}_i\tilde{G}_i')\}(\hat{\Lambda}^{-1/2}\Lambda^{1/2} - I)\|_\infty + \|(\hat{\Lambda}^{-1/2}\Lambda^{1/2} - I)\{E(\tilde{G}_i\tilde{G}_i')\}\|_\infty\\
\le&(2 + \|\hat{\Lambda}^{-1}\Lambda - I\|_\infty)\|n^{-1}\sum_{i=1}^{n} \tilde{G}_i\tilde{G}_i' - E(\tilde{G}_i\tilde{G}_i')\|_\infty + \|\hat{\Lambda}^{-1/2}\Lambda^{1/2} - I\|_\infty\|E(\tilde{G}_i\tilde{G}_i')\|_\infty\\
&+ (2 + \|\hat{\Lambda}^{-1/2}\Lambda^{1/2} - I\|_\infty)\|\hat{\Lambda}^{-1/2}\Lambda^{1/2} - I\|_\infty\|E(\tilde{G}_i\tilde{G}_i')\|_\infty. \qquad \text{(S2.13)}
\end{aligned}
$$

By combining parts 1–6), (A2.1), and (A4.1) with (S2.13), we have that with probability tending to 1:

$$
\|n^{-1}\sum_{i=1}^{n} \breve{G}_i\breve{G}_i' - E(\tilde{G}_i\tilde{G}_i')\|_\infty \le c_4\{\log(np_ns_n)/n\}^{1/2},
$$

for some universal constant $c_4 > 0$, which completes the proof of part 8). To show part 9), note that for any $t > 0$,

$$
\begin{aligned}
&P\big\{\max_{j\le h_n s_n}\|n^{-1}\sum_{i=1}^{n}\tilde{F}_i(\tilde{E}_{ij}-F_i'w_j)\|_\infty \ge t\big\}\\
\le&\sum_{j=1}^{h_n s_n}\sum_{l=1}^{(p_n-h_n)s_n} P\big\{|n^{-1}\sum_{i=1}^{n}\tilde{F}_{il}(\tilde{E}_{ij}-F_i'w_j)|\ge t\big\}\\
\le&2h_n s_n(p_n-h_n)s_n\exp\{-n\min(c_5^{-2}t^2, c_5^{-1}t)\}\\
\le&2(p_n s_n)^2\exp\{-n\min(c_5^{-2}t^2, c_5^{-1}t)\},
\end{aligned}
\tag{S2.14}
$$

for some universal constant $c_5 > 0$, where the first inequality is by union bound inequality and the second inequality holds from (A2.1) and Bernstein's inequality. Plugging $t = 2c_5\{\log(np_n s_n)/n\}^{1/2}$ into (S2.14) yields

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

which completes the proof of part 9). To show part 10), note that

$$
\begin{aligned}
&\max_{j\le h_n s_n}\|n^{-1}\sum_{i=1}^{n}\tilde{F}_i(\breve{E}_{ij}-F_i'w_j)\|_\infty\\
\le&\max_{j\le h_n s_n}\|n^{-1}\sum_{i=1}^{n}\tilde{F}_i(\tilde{E}_{ij}-F_i'w_j)\|_\infty+\max_{j\le h_n s_n}\|n^{-1}\sum_{i=1}^{n}\tilde{F}_i(\breve{E}_{ij}-\tilde{E}_{ij})\|_\infty\\
\le&\max_{j\le h_n s_n}\|n^{-1}\sum_{i=1}^{n}\tilde{F}_i(\tilde{E}_{ij}-F_i'w_j)\|_\infty+\|\hat{\Lambda}^{-1/2}\Lambda^{1/2}-I\|_\infty\max_{j\le h_n s_n}\|n^{-1}\sum_{i=1}^{n}\tilde{F}_i\tilde{E}_{ij}\|_\infty\\
\le&\max_{j\le h_n s_n}\|n^{-1}\sum_{i=1}^{n}\tilde{F}_i(\tilde{E}_{ij}-F_i'w_j)\|_\infty+\|\hat{\Lambda}^{-1/2}\Lambda^{1/2}-I\|_\infty\cdot\\
&\{\|n^{-1}\sum_{i=1}^{n}\tilde{G}_i\tilde{G}_i'-E(\tilde{G}_i\tilde{G}_i')\|_\infty+\|E(\tilde{G}_i\tilde{G}_i')\|_\infty\}.
\end{aligned}
\tag{S2.15}
$$

By combining parts 1–9), (A2.1), and (A4.1) with (S2.15), we have that with probability tending to 1:

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

for some universal constant $c_6 > 0$, which completes the proof of part 10). To show part 11), note that

$$\begin{aligned}&\max_{j\le h_n s_n} \|n^{-1}\sum_{i=1}^n \breve{F}_i(\breve{E}_{ij} - F_i' w_j)\|_\infty \\ \le&(2 + \|\hat{\Lambda}^{-1/2}\Lambda^{1/2} - I\|_\infty) \max_{j\le h_n s_n} \|n^{-1}\sum_{i=1}^n \tilde{F}_i(\tilde{E}_{ij} - F_i' w_j)\|_\infty. \qquad \text{(S2.16)}\end{aligned}$$

By combining parts 1–10), (A4.1) with (S2.16), we have that with probability tending to 1:

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

for some universal constant $c_7 > 0$, which completes the proof of part 11). To show part 12), note that for any $t > 0$,

$$\begin{aligned}&P\big(\|n^{-1}\sum_{i=1}^n \tilde{G}_i\|_\infty \ge t\big) = P\big(\max_{l\le p_n s_n} |n^{-1}\sum_{i=1}^n \tilde{G}_{il}| \ge t\big) \\ \le&\sum_{l=1}^{p_n s_n} P\big(|n^{-1}\sum_{i=1}^n \tilde{G}_{il}| \ge t\big) \le 2p_n s_n \exp\{-n(t/c_8)^2\}, \qquad \text{(S2.17)}\end{aligned}$$

for some universal constant $c_8 > 0$, where the last inequality follows from (A2.1) and Hoeffding's inequality. Plugging $t = c_8\{\log(np_n s_n)/n\}^{1/2}$ into (S2.17) yields

$$P\Big(\|n^{-1}\sum_{i=1}^{n}\tilde{G}_i\|_\infty \le c_8\{\log(np_ns_n)/n\}^{1/2}\Big) \ge 1-2n^{-1}\to 1, \qquad \text{(S2.18)}$$

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

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

for some universal constant $c_9>0$, where the last inequality is based on (A3.1).

Together with (S2.18), it is obvious that

$$P\Big(\|n^{-1}\sum_{i=1}^{n}G_i\|_\infty \le c_8c_9\{\log(np_ns_n)/n\}^{1/2}\Big) \ge 1-2n^{-1}\to 1, \qquad \text{(S2.19)}$$

which completes the proof of part 13). To show part 14), note that

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

Together with parts 5) and 12), the assertion in part 14) holds obviously. To show part 15), note that for any $t>0$,

$$\begin{aligned}&P\big(\max_{j\le p_ns_n}\max_{i\le n}|\tilde{G}_{ij}|\ge t\big)\\ \le&\sum_{j=1}^{p_ns_n}\sum_{i=1}^{n}P\big(|\tilde{G}_{ij}|\ge t\big)\le 2np_ns_n\exp\{-(t/c_{10})^2\},\end{aligned} \qquad \text{(S2.20)}$$

for some universal constant $c_{10}>0$, where the last inequality is based on (A2.1).

Plugging $t=2c_{10}\{\log(np_ns_n)/n\}^{1/2}$ into (S2.20) yields

$$P\big(\max_{j\le p_ns_n}\max_{i\le n}|\tilde{G}_{ij}|\le 2c_{10}\{\log(np_ns_n)/n\}^{1/2}\Big)\ge 1-2n^{-3}(p_ns_n)^{-3}\to 1,$$

which completes the proof of part 15). In a similar fashion to part 15), one can show parts 16). Similar reasoning as part 6) leads to part 17). Part 26) follows from Holder's inequality and (A2.1). □

Lemma 3. *Under conditions (A1), (A2.1), (A2.3), (A3), (A4.1), (A4.3), we have that with probability tending to 1:*

1) $\max_{i\leq n} |\delta_i^*| \leq c_1 q_n \log^{1/2}(nq_n s_n)$, *for some universal constant* $c_1 > 0$.

2) $\max_{i\leq n} |\delta_i| \leq c_2 q_n \log^{1/2}(nq_n s_n)$, *for some universal constant* $c_2 > 0$.

3) $\sum_{i=1}^n (\sum_{j=1}^{q_n} \sum_{k=s_n+1}^{\infty} \theta_{ijk}\eta_{jk})^2 \leq c_3 n q_n^2 s_n^{-2\delta+1}$, *for some universal constant* $c_3 > 0$.

4) $\max_{i\leq n} |\sum_{j=1}^{q_n} \sum_{k=s_n+1}^{\infty} \theta_{ijk}\eta_{jk}|^2 \leq c_4 n q_n^2 s_n^{-2\delta+1}$, *for some universal constant* $c_4 > 0$.

5) $\max_{i\leq n} var(\epsilon_i|X_i) \leq c_5 \exp\{c_5 q_n \log^{1/2}(nq_n s_n)\}$, *for some universal constant* $c_5 > 0$.

6) $\max_{i\leq n} |\epsilon_i| \leq c_6 \log(n) \exp\{c_6 q_n \log^{1/2}(nq_n s_n)\}$, *for some universal constant* $c_6 > 0$.

7) $|n^{-1}\sum_{i=1}^n \epsilon_i| \leq c_7 \{\log(n)/n\}^{1/2} \exp\{c_7 q_n \log^{1/2}(nq_n s_n)\}$, *for some universal constant* $c_7 > 0$.

8) $\|n^{-1}\sum_{i=1}^{n}\tilde{G}_i\epsilon_i\|_\infty \le c_8\{\log^2(np_ns_n)/n\}^{1/2}\exp\{c_8q_n\log^{1/2}(nq_ns_n)\}$, *for some universal constant* $c_8 > 0$.

9) $\|n^{-1}\sum_{i=1}^{n}G_i\epsilon_i\|_\infty \le c_9\{\log^2(np_ns_n)/n\}^{1/2}\exp\{c_9q_n\log^{1/2}(nq_ns_n)\}$, *for some universal constant* $c_9 > 0$.

10) $\|n^{-1}\sum_{i=1}^{n}\breve{G}_i\epsilon_i\|_\infty \le c_{10}\{\log^2(np_ns_n)/n\}^{1/2}\exp\{c_{10}q_n\log^{1/2}(nq_ns_n)\}$, *for some universal constant* $c_{10} > 0$.

11) $|n^{-1}\sum_{i=1}^{n}\{Y_i - b'(\delta_i^*)\}| \le c_{11}[q_ns_n^{-\delta+1/2} + \{\log(n)/n\}^{1/2}]\exp\{c_{11}q_n\log^{1/2}(nq_ns_n)\}$, *for some universal constant* $c_{11} > 0$.

12) $\|n^{-1}\sum_{i=1}^{n}\tilde{G}_i\{Y_i - b'(\delta_i^*)\}\|_\infty \le c_{12}q_ns_n^{-\delta+1/2}\exp\{c_{12}q_n\log^{1/2}(nq_ns_n)\} + c_{12}\{\log^2(np_ns_n)/n\}^{1/2}\exp\{c_{12}q_n\log^{1/2}(nq_ns_n)\}$, *for some universal constant* $c_{12} > 0$.

13) $\|n^{-1}\sum_{i=1}^{n}G_i\{Y_i - b'(\delta_i^*)\}\|_\infty \le c_{13}q_ns_n^{-\delta+1/2}\exp\{c_{13}q_n\log^{1/2}(nq_ns_n)\} + c_{13}\{\log^2(np_ns_n)/n\}^{1/2}\exp\{c_{13}q_n\log^{1/2}(nq_ns_n)\}$, *for some universal constant* $c_{13} > 0$.

14) $\|n^{-1}\sum_{i=1}^{n}\breve{G}_i\{Y_i - b'(\delta_i^*)\}\|_\infty \le c_{14}q_ns_n^{-\delta+1/2}\exp\{c_{14}q_n\log^{1/2}(nq_ns_n)\} + c_{14}\{\log^2(np_ns_n)/n\}^{1/2}\exp\{c_{14}q_n\log^{1/2}(nq_ns_n)\}$, *for some universal constant* $c_{14} > 0$.

Recall that $\delta_i^* = \alpha_0 + \sum_{j=1}^{p_n}\sum_{k=1}^{s_n}\theta_{ijk}\eta_{jk}$ *and* $\delta_i = \alpha_0 + \sum_{j=1}^{p_n}\sum_{k=1}^{\infty}\theta_{ijk}\eta_{jk}$.

Proof. To show part 1), note that with probability tending to 1,

$$
\begin{aligned}
&\max_{i\le n}|\delta_i^*| \le |\alpha_0| + c_1\{\log(nq_ns_n)\}^{1/2}\|\Lambda^{1/2}\eta\|_1\\
\le&|\alpha_0| + c_1\{\log(nq_ns_n)\}^{1/2}\{q_n(\sup_{j\le q_n}\sum_{k=1}^{s_n}\omega_{jk}k^{-2\delta})^{1/2}(\sup_{j\le q_n}\sum_{k=1}^{s_n}\eta_{jk}^2k^{2\delta})^{1/2}\}\\
\le&c_2q_n\log^{1/2}(nq_ns_n), \qquad\qquad \text{(S2.21)}
\end{aligned}
$$

for some universal constants $c_1, c_2 > 0$, where the first inequality holds from part 18) of Lemma 2, and the last inequality is based on (A3). This completes the proof of part 1). To show part 3), first note that

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

where the last inequality follows from (A3.1), (A3.2) and (A4.3). Together with Markov inequality yields that with probability tending to 1,

$$
\sum_{i=1}^{n}(\sum_{j=1}^{q_n}\sum_{k=s_n+1}^{\infty}\theta_{ijk}\eta_{jk})^2 \le c_3nq_n^2s_n^{-2\delta+1}, \qquad \text{(S2.22)}
$$

for some universal constant $c_3 > 0$, which completes the proof of part 3). Thus, we have that with probability tending to 1,

$$
\max_{i\le n}|\sum_{j=1}^{q_n}\sum_{k=s_n+1}^{\infty}\theta_{ijk}\eta_{jk}|^2 \le \sum_{i=1}^{n}(\sum_{j=1}^{q_n}\sum_{k=s_n+1}^{\infty}\theta_{ijk}\eta_{jk})^2 \le c_3nq_n^2s_n^{-2\delta+1}, \quad \text{(S2.23)}
$$

which completes the proof of part 4). To show part 2), note that

$$\max_{i\le n}|\delta_i| = \max_{i\le n}|\delta_i^* + \sum_{j=1}^{q_n}\sum_{k=s_n+1}^{\infty}\theta_{ijk}\eta_{jk}| \le \max_{i\le n}|\delta_i^*| + \max_{i\le n}|\sum_{j=1}^{q_n}\sum_{k=s_n+1}^{\infty}\theta_{ijk}\eta_{jk}|$$
$$\le c_4 q_n \log^{1/2}(nq_ns_n), \tag{S2.24}$$

for some universal constant $c_4 > 0$, with probability tending to 1, where the last inequality is based on part 1), part 4), and (A4.3). This completes the proof of part 2). To show part 5), note that

$$\max_{i\le n}\text{var}(\epsilon_i|X_i) = \max_{i\le n} a_i(\phi_i)b''(\delta_i) \le \{\max_{i\le n} a_i(\phi_i)\}\cdot\{\max_{i\le n} b''(\delta_i)\}$$
$$\le\{\max_{i\le n} a_i(\phi_i)\}\cdot\exp\{\max_{i\le n}|\delta_i|\} \le c_5\exp\{c_5 q_n\log^{1/2}(nq_ns_n)\}, \tag{S2.25}$$

for some universal constant $c_5 > 0$, with probability tending to 1, where the second last inequality is by (A1), and the last inequality is by $\max_{i\le n} a_i(\phi_i) < \infty$, and part 2). This completes the proof of part 5). To show part 6), note that conditional on the data $\{X_i\}_{i=1}^n$, we have that for any $t > 0$,

$$P[\max_{i\le n}|\epsilon_i| \ge t|\{X_i\}_{i=1}^n]$$
$$\le\sum_{i=1}^n P[|\epsilon_i|\ge t|\{X_i\}_{i=1}^n] = \sum_{i=1}^n P(|\epsilon_i|\ge t|X_i)$$
$$\le 2\sum_{i=1}^n\exp(-\min[\frac{c_6^{-2}t^2}{\{1+\text{var}(\epsilon_i|X_i)\}^2}, \frac{c_6^{-1}t}{1+\text{var}(\epsilon_i|X_i)}]) + 2\sum_{i=1}^n\exp\{-\frac{c_6^{-2}t^2}{1+\text{var}(\epsilon_i|X_i)}\}$$
$$\le 2n\exp(-\min[\frac{c_6^{-2}t^2}{\{1+\max_{i\le n}\text{var}(\epsilon_i|X_i)\}^2}, \frac{c_6^{-1}t}{1+\max_{i\le n}\text{var}(\epsilon_i|X_i)}])+$$
$$2n\exp\{-\frac{c_6^{-2}t^2}{1+\max_{i\le n}\text{var}(\epsilon_i|X_i)}\}, \tag{S2.26}$$

for some universal constant $c_6 > 0$, where the first inequality is by union bound inequality, and the second inequality holds from (A2.3). Plugging $t = 2c_6\{\log(n)\}\{1+\max_{i\le n}\text{var}(\epsilon_i|X_i)\}$ into (S2.26) yields

$$P[\max_{i\le n}|\epsilon_i| \le 2c_6\{\log(n)\}\{1+\max_{i\le n}\text{var}(\epsilon_i|X_i)\}|\{X_i\}_{i=1}^n] \\ \ge 1-2n^{-3}-2n^{-1}. \quad \text{(S2.27)}$$

Taking expectation on both sides of (S2.27), we obtain

$$P[\max_{i\le n}|\epsilon_i| \le 2c_6\{\log(n)\}\{1+\max_{i\le n}\text{var}(\epsilon_i|X_i)\}] \ge 1-2n^{-3}-2n^{-1}.$$

Together with (S2.25) yields that with probability tending to 1,

$$\max_{i\le n}|\epsilon_i| \le c_7\log(n)\exp\{c_7 q_n \log^{1/2}(nq_ns_n)\}, \quad \text{(S2.28)}$$

for some universal constant $c_7 > 0$, which finishes the proof of part 6). To show part 7), note that conditional on the data $\{X_i\}_{i=1}^n$, we have that for any $t > 0$,

$$P[|n^{-1}\sum_{i=1}^n \epsilon_i| \ge t|\{X_i\}_{i=1}^n] \\ \le 2\exp(-n\min[\frac{c_8^{-2}t^2}{\{1+\max_{i\le n}\text{var}(\epsilon_i|X_i)\}^2}, \frac{c_8^{-1}t}{1+\max_{i\le n}\text{var}(\epsilon_i|X_i)}])+ \\ 2\exp\{-\frac{c_8^{-2}nt^2}{1+\max_{i\le n}\text{var}(\epsilon_i|X_i)}\}, \quad \text{(S2.29)}$$

for some universal constant $c_8 > 0$, where the first inequality is based on (A2.3), Hoeffding's inequality, and Bernstein's inequality. Plugging $t = c_8\{\log(n)/n\}^{1/2}\{1+$

$\max_{i \le n} \text{var}(\epsilon_i|X_i)\}$ into (S2.29) yields

$$P[|n^{-1}\sum_{i=1}^{n}\epsilon_i| \le c_8\{\log(n)/n\}^{1/2}\{1+\max_{i\le n}\text{var}(\epsilon_i|X_i)\}|\{X_i\}_{i=1}^n] \\ \ge 1-4n^{-1}. \tag{S2.30}$$

Taking expectation on both sides of (S2.30), we obtain

$$P[|n^{-1}\sum_{i=1}^{n}\epsilon_i| \le c_8\{\log(n)/n\}^{1/2}\{1+\max_{i\le n}\text{var}(\epsilon_i|X_i)\}] \ge 1-4n^{-1}.$$

Together with (S2.25) yields that with probability tending to 1,

$$|n^{-1}\sum_{i=1}^{n}\epsilon_i| \le c_9\{\log(n)/n\}^{1/2}\exp\{c_9 q_n \log^{1/2}(nq_n s_n)\}, \tag{S2.31}$$

for some universal constant $c_9 > 0$, which finishes the proof of part 7). To show part 8), note that conditional on the data $\{X_i\}_{i=1}^n$, we have that for any $t>0$,

$$
\begin{aligned}
&P[\|n^{-1}\sum_{i=1}^n \tilde{G}_i\epsilon_i\|_\infty \geq t|\{X_i\}_{i=1}^n] = P[\max_{l\leq p_ns_n}|n^{-1}\sum_{i=1}^n \tilde{G}_{il}\epsilon_i| \geq t|\{X_i\}_{i=1}^n]\\
\leq& \sum_{l=1}^{p_ns_n} P[|n^{-1}\sum_{i=1}^n \tilde{G}_{il}\epsilon_i| \geq t|\{X_i\}_{i=1}^n]\\
\leq& \sum_{l=1}^{p_ns_n} 2\exp[-\frac{c_{10}^{-2}nt^2}{\max_{i\leq n}|\tilde{G}_{il}|^2\{1+\max_{i\leq n}\text{var}(\epsilon_i|X_i)\}}]+\\
&\sum_{l=1}^{p_ns_n} 2\exp(-n\min[\frac{c_{10}^{-2}t^2}{\max_{i\leq n}|\tilde{G}_{il}|^2\{1+\max_{i\leq n}\text{var}(\epsilon_i|X_i)\}^2},\\
&\frac{c_{10}^{-1}t}{\max_{i\leq n}|\tilde{G}_{il}|\{1+\max_{i\leq n}\text{var}(\epsilon_i|X_i)\}}])\\
\leq& 2p_ns_n\exp[-\frac{c_{10}^{-2}nt^2}{\max_{l\leq p_ns_n}\max_{i\leq n}|\tilde{G}_{il}|^2\{1+\max_{i\leq n}\text{var}(\epsilon_i|X_i)\}}]+\\
&2p_ns_n\exp(-n\min[\frac{c_{10}^{-2}t^2}{\max_{l\leq p_ns_n}\max_{i\leq n}|\tilde{G}_{il}|^2\{1+\max_{i\leq n}\text{var}(\epsilon_i|X_i)\}^2},\\
&\frac{c_{10}^{-1}t}{\max_{l\leq p_ns_n}\max_{i\leq n}|\tilde{G}_{il}|\{1+\max_{i\leq n}\text{var}(\epsilon_i|X_i)\}}]),
\end{aligned}
\tag{S2.32}
$$

for some universal constant $c_{10}>0$, where the first inequality is by union bound inequality, and the second inequality is based on (A2.3), Hoeffding's inequality, and Bernstein's inequality. Plugging $t = c_{10}\{\log(np_ns_n)/n\}^{1/2}\max_{l\leq p_ns_n+d_n}\max_{i\leq n}|\tilde{G}_{il}|\{1+\max_{i\leq n}\text{var}(\epsilon_i|X_i)\}$ into (S2.32) yields

$$
\begin{aligned}
&P\big(\|n^{-1}\sum_{i=1}^n \tilde{G}_i\epsilon_i\|_\infty \leq c_{10}\{\log(np_ns_n)/n\}^{1/2}\max_{l\leq p_ns_n+d_n}\max_{i\leq n}|\tilde{G}_{il}|\{1+\max_{i\leq n}\text{var}(\epsilon_i|X_i)\}|\\
&\quad\{X_i\}_{i=1}^n\big) \geq 1-4n^{-1}.
\end{aligned}
\tag{S2.33}
$$

Taking expectation on both sides of (S2.33), we obtain

$$P\big(\|n^{-1}\sum_{i=1}^{n}\tilde{G}_i\epsilon_i\|_\infty \le c_{10}\{\log(np_ns_n)/n\}^{1/2}\max_{l\le p_ns_n}\max_{i\le n}|\tilde{G}_{il}|\{1+\max_{i\le n}\text{var}(\epsilon_i|X_i)\}\big)$$
$$\ge 1-4n^{-1}.$$

Together with (S2.25) and Lemma 2 yields that with probability tending to 1,

$$\|n^{-1}\sum_{i=1}^{n}\tilde{G}_i\epsilon_i\|_\infty \le c_{11}\{\log^2(np_ns_n)/n\}^{1/2}\exp\{c_{11}q_n\log^{1/2}(nq_ns_n)\}, \tag{S2.34}$$

for some universal constant $c_{11} > 0$, which completes the proof of part 8). Part 9) is due to part 8) and (A3.1). Part 10) is based on part 8) and Lemma 2. Before showing part 11), first note that it follows from mean value theorem that for any $i \le n$,

$$b'(\delta_i) - b'(\delta_i^*) = b''(\delta_i^* + t_i\sum_{j=1}^{q_n}\sum_{k=s_n+1}^{\infty}\theta_{ijk}\eta_{jk})\cdot\sum_{j=1}^{q_n}\sum_{k=s_n+1}^{\infty}\theta_{ijk}\eta_{jk}, \tag{S2.35}$$

for some $t_i \in [0,1]$. In addition, we have

$$\max_{i\le n}|b''(\delta_i^* + t_i\sum_{j=1}^{q_n}\sum_{k=s_n+1}^{\infty}\theta_{ijk}\eta_{jk})| \le \exp(\max_{i\le n}|\delta_i^* + t_i\sum_{j=1}^{q_n}\sum_{k=s_n+1}^{\infty}\theta_{ijk}\eta_{jk}|)$$
$$\le \exp(\max_{i\le n}|\delta_i^*|)\cdot\exp(\max_{i\le n}|\sum_{j=1}^{q_n}\sum_{k=s_n+1}^{\infty}\theta_{ijk}\eta_{jk}|)$$
$$\le c_{12}\exp\{c_{12}q_n\log^{1/2}(nq_ns_n)\}, \tag{S2.36}$$

for some universal constant $c_{12} > 0$, with probability tending to 1, where the first inequality holds from (A1), and the last inequality is based on (S2.21), (S2.23), and (A4.3). To this end, note that

$$
\begin{aligned}
&|n^{-1}\sum_{i=1}^{n}\{Y_i - b'(\delta_i^*)\}| = |n^{-1}\sum_{i=1}^{n}\{\epsilon_i + b'(\delta_i) - b'(\delta_i^*)\}| \\
\le&|n^{-1}\sum_{i=1}^{n}\epsilon_i| + |n^{-1}\sum_{i=1}^{n}\{b'(\delta_i) - b'(\delta_i^*)\}| \\
=&|n^{-1}\sum_{i=1}^{n}\epsilon_i| + |n^{-1}\sum_{i=1}^{n} b''(\delta_i^* + t_i\sum_{j=1}^{q_n}\sum_{k=s_n+1}^{\infty}\theta_{ijk}\eta_{jk})\cdot\sum_{j=1}^{q_n}\sum_{k=s_n+1}^{\infty}\theta_{ijk}\eta_{jk}| \\
\le&|n^{-1}\sum_{i=1}^{n}\epsilon_i| + \{\max_{i\le n}|b''(\delta_i^* + t_i\sum_{j=1}^{q_n}\sum_{k=s_n+1}^{\infty}\theta_{ijk}\eta_{jk})|\}\cdot\{n^{-1}\sum_{i=1}^{n}(\sum_{j=1}^{q_n}\sum_{k=s_n+1}^{\infty}\theta_{ijk}\eta_{jk})^2\}^{1/2} \\
\le&c_{13}[q_n s_n^{-\delta+1/2} + \{\log(n)/n\}^{1/2}]\exp\{c_{13}q_n\log^{1/2}(nq_ns_n)\},
\end{aligned}
\tag{S2.37}
$$

for some universal constant $c_{13} > 0$, with probability tending to 1, where the second equality holds from (S2.35), and the last inequality is based on (S2.31), (S2.36), and (S2.22). This completes the proof of part 11). To show part 12), note that

$$
\begin{aligned}
&\|n^{-1}\sum_{i=1}^{n}\tilde{G}_i\{Y_i - b'(\delta_i^*)\}\|_\infty \le \|n^{-1}\sum_{i=1}^{n}\tilde{G}_i\epsilon_i\|_\infty + \|n^{-1}\sum_{i=1}^{n}\tilde{G}_i\{b'(\delta_i) - b'(\delta_i^*)\}\|_\infty \\
=&\|n^{-1}\sum_{i=1}^{n}\tilde{G}_i\epsilon_i\|_\infty + \max_{l\le p_ns_n}|n^{-1}\sum_{i=1}^{n}\tilde{G}_{il}b''(\delta_i^* + t_i\sum_{j=1}^{q_n}\sum_{k=s_n+1}^{\infty}\theta_{ijk}\eta_{jk})\cdot\sum_{j=1}^{q_n}\sum_{k=s_n+1}^{\infty}\theta_{ijk}\eta_{jk}| \\
\le&\|n^{-1}\sum_{i=1}^{n}\tilde{G}_i\epsilon_i\|_\infty + \{\max_{l\le p_ns_n}n^{-1}\sum_{i=1}^{n}\tilde{G}_{il}^2\}^{1/2}\cdot\{n^{-1}\sum_{i=1}^{n}(\sum_{j=1}^{q_n}\sum_{k=s_n+1}^{\infty}\theta_{ijk}\eta_{jk})^2\}^{1/2} \\
&\cdot\max_{i\le n}|b''(\delta_i^* + t_i\sum_{j=1}^{q_n}\sum_{k=s_n+1}^{\infty}\theta_{ijk}\eta_{jk})|.
\end{aligned}
\tag{S2.38}
$$

For the term $\max_{l\le p_ns_n} n^{-1}\sum_{i=1}^{n}\tilde{G}_{il}^2$, it follows from Lemma 2, (A2.1), and

(A4.1) that with probability tending to 1,

$$\max_{l \le p_n s_n} n^{-1} \sum_{i=1}^{n} \tilde{G}_{il}^2 \le c_{14}, \tag{S2.39}$$

for some universal constant $c_{14} > 0$. By combining (S2.39), (S2.36), (S2.22), (S2.34) with (S2.38), it can be deduced that with probability tending to 1,

$$\|n^{-1} \sum_{i=1}^{n} \tilde{G}_i \{Y_i - b'(\delta_i^*)\}\|_\infty \le c_{15} q_n s_n^{-\delta+1/2} \exp\{c_{15} q_n \log^{1/2}(n q_n s_n)\} +$$

$$c_{15}\{\log^2(n p_n s_n)/n\}^{1/2} \exp\{c_{15} q_n \log^{1/2}(n q_n s_n)\}, \tag{S2.40}$$

for some universal constant $c_{15} > 0$, which completes the proof of part 12). Piecing (S2.40) and (A3.1) together yields part 13). Finally, part 14) holds from (S2.40) and Lemma 2. □

Lemma 4. *Under conditions (A2.1), (A3.1), (A3.3), (A4.1)-(A4.2) and (B1)-(B5), we have that with probability tending to 1:*

1) $\sum_{j=1}^{p_n} \|\Theta_j(\eta_j - \eta_j^*)\|_2^2 \le c_1 n \|\tilde{\nu}\|_2^2$, *for some universal constant* $c_1 > 0$.

2) $\|\tilde{\nu}\|_1 \le c_2 n^{-1/2} s_n^{1/2} \sum_{j=1}^{p_n} \|\Theta_j(\eta_j - \eta_j^*)\|_2$, *for some universal constant* $c_2 > 0$.

3) $\lambda_n \|\tilde{\nu}\|_1 \le c_3 s_n^{1/2} n^{1/18} \{P_{\lambda_n}(\eta^*) + P_{\lambda_n}(\eta) + n^{-1/9} \|\tilde{\nu}\|_2^2\}$, *for some universal constant* $c_3 > 0$.

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

Proof. First of all, note that with probability tending to 1,

$$
\begin{aligned}
&\sum_{j=1}^{p_n} \|\Theta_j(\eta_j-\eta_j^*)\|_2^2 \\
=&\sum_{j=1}^{p_n}\{\Lambda_j^{1/2}(\eta_j-\eta_j^*)\}'\{\Lambda_j^{-1/2}\Theta_j{}'\Theta_j\Lambda_j^{-1/2}-E(\Lambda_j^{-1/2}\Theta_j{}'\Theta_j\Lambda_j^{-1/2})\}\{\Lambda_j^{1/2}(\eta_j-\eta_j^*)\}+ \\
&\sum_{j=1}^{p_n}\{\Lambda_j^{1/2}(\eta_j-\eta_j^*)\}'E(\Lambda_j^{-1/2}\Theta_j{}'\Theta_j\Lambda_j^{-1/2})\{\Lambda_j^{1/2}(\eta_j-\eta_j^*)\} \\
\leq& n\{\lambda_{\max}(E(\tilde{G}_i\tilde{G}_i')) + s_n\|n^{-1}\sum_{i=1}^{n}\tilde{G}_i\tilde{G}_i' - E(\tilde{G}_i\tilde{G}_i')\|_\infty\}\cdot\sum_{j=1}^{p_n}\|\Lambda_j^{1/2}(\eta_j-\eta_j^*)\|_2^2 \\
\leq& c_1 n[1+\{s_n^2\log(np_ns_n)/n\}^{1/2}]\cdot\|\tilde{\nu}\|_2^2 \leq c_2 n\|\tilde{\nu}\|_2^2, \qquad \text{(S2.41)}
\end{aligned}
$$

for some universal constants $c_1, c_2 > 0$, where the second last inequality follows from (A3.3) and Lemma 2, and the last inequality is by (A4.2). This completes the proof of part 1). To show part 2), note that with probability tending to 1,

$$
\begin{aligned}
&\|\tilde{\nu}\|_1 = \sum_{j=1}^{p_n}\|\Lambda_j^{1/2}(\eta_j-\eta_j^*)\|_1 \leq s_n^{1/2}\sum_{j=1}^{p_n}\left[\{\Lambda_j^{1/2}(\eta_j-\eta_j^*)\}'\{\Lambda_j^{1/2}(\eta_j-\eta_j^*)\}\right]^{1/2} \\
\leq& s_n^{1/2}\{\lambda_{\min}(E(\tilde{G}_i\tilde{G}_i'))\}^{-1/2}\sum_{j=1}^{p_n}\left[\{\Lambda_j^{1/2}(\eta_j-\eta_j^*)\}'\{n^{-1}E(\Lambda_j^{-1/2}\Theta_j\Theta_j\Lambda_j^{-1/2})\}\{\Lambda_j^{1/2}(\eta_j-\eta_j^*)\}\right]^{1/2} \\
\leq& c_3 s_n^{1/2}\sum_{j=1}^{p_n}\left[\{\Lambda_j^{1/2}(\eta_j-\eta_j^*)\}'\{n^{-1}E(\Lambda_j^{-1/2}\Theta_j\Theta_j\Lambda_j^{-1/2})\}\{\Lambda_j^{1/2}(\eta_j-\eta_j^*)\}\right]^{1/2} \\
\leq& c_3 n^{-1/2}s_n^{1/2}\sum_{j=1}^{p_n}\|\Theta_j(\eta_j-\eta_j^*)\|_2 + c_3 s_n^{1/2}\|n^{-1}\sum_{i=1}^{n}\tilde{G}_i\tilde{G}_i' - E(\tilde{G}_i\tilde{G}_i')\|_\infty^{1/2}\|\tilde{\nu}\|_1 \\
\leq& c_3 n^{-1/2}s_n^{1/2}\sum_{j=1}^{p_n}\|\Theta_j(\eta_j-\eta_j^*)\|_2 + c_4\{s_n^2\log(np_ns_n)/n\}^{1/4}\|\tilde{\nu}\|_1,
\end{aligned}
$$

for some universal constants $c_3, c_4 > 0$, where the third inequality follows from (A3.3), and the last inequality is based on Lemma 2. Together with the fact that

$s_n^2 \log(np_n s_n)/n \to 0$ under (A4.2), it is seen that with probability tending to 1,

$$\|\tilde{\nu}\|_1 \lesssim n^{-1/2} s_n^{1/2} \sum_{j=1}^{p_n} \|\Theta_j(\eta_j - \eta_j^*)\|_2, \tag{S2.42}$$

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

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

where the first inequality holds from (S2.42), the second inequality is based on part 3) of Lemma 1, and the last inequality follows from the subadditivity of $\rho_\lambda(\cdot)$ and (S2.41). This completes the proof of part 3). □

Lemma 5. *Let $X_1, X_2, \ldots, X_n$ be centered independent random vectors, with each $X_i = (X_{i1}, \ldots, X_{ip})' \in \mathbb{R}^p$. Then, we have parts 1)–2) below,*

1) 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) There exists a sequence of constants $u_n \ge 1$ such that all X_{ij} are sub-exponential with variance proxy u_n.

(c) $u_n^6 \log^9(np)/n \to 0$.

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* $A = \{\omega \in \mathbb{R}^p : a_j \le \omega_j \le b_j, j \le p\}$ *for some* $-\infty \le a_j \le b_j \le \infty$ *for all* $j \le p$*. 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*

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

(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:

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

2) 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') There exists a sequence of constants $v_n \ge 1$ *such that all* X_{ij} *are sub-gaussian with variance proxy* v_n*.*

(c') $v_n^3 \log^9(np)/n \to 0$.

Then, under (a')–(c'), we have:

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

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 (A2.1), (A3.1), (A3.3), (A4.1), (A5.3) and (A5.4), we have*

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

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

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

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

$$P\big[\bigcap_{j=1}^{h_ns_n}\{\|\hat{\Lambda}^{1/2}_{\mathcal{H}_n^c}(\hat{w}_j - w_j)\|_2 \le c_2\{\rho_{nj}\log(np_ns_n)/n\}^{1/2}\}\big] \to 1,$$

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

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

$$P\big[\bigcap_{j=1}^{h_ns_n}\{\|\hat{\Lambda}^{1/2}_{\mathcal{H}_n^c}(\hat{w}_j - w_j)\|_1 \le c_3\{\rho_{nj}^2\log(np_ns_n)/n\}^{1/2}\}\big] \to 1,$$

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

Proof. By the definition of $\hat{w}$ in (3.8) of the main paper, it holds true for all $j \le h_ns_n$ that

$$(2n)^{-1}\sum_{i=1}^n(\breve{E}_{ij} - F_i'\hat{w}_j)^2 + \lambda_n^*\|\hat{\Lambda}^{1/2}_{\mathcal{H}_n^c}\hat{w}_j\|_1 \le (2n)^{-1}\sum_{i=1}^n(\breve{E}_{ij} - F_i'w_j)^2 + \lambda_n^*\|\hat{\Lambda}^{1/2}_{\mathcal{H}_n^c}w_j\|_1,$$

which implies that for all $j \le h_ns_n$,

$$0 \le (\hat{w}_j - w_j)'\{(2n)^{-1}\sum_{i=1}^n F_iF_i'\}(\hat{w}_j - w_j) \le \|n^{-1}\sum_{i=1}^n \breve{F}_i(\breve{E}_{ij} - F_i'w_j)\|_\infty \cdot \|\hat{\Lambda}^{1/2}_{\mathcal{H}_n^c}(\hat{w}_j - w_j)\|_1$$
$$+ \lambda_n^*\|\hat{\Lambda}^{1/2}_{\mathcal{H}_n^c}w_j\|_1 - \lambda_n^*\|\hat{\Lambda}^{1/2}_{\mathcal{H}_n^c}\hat{w}_j\|_1. \quad \text{(S2.43)}$$

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

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

Together with (S2.43) yields that for all $j \leq h_n s_n$,

$$\begin{aligned}0 \leq &(\hat{w}_j - w_j)'\{(2n)^{-1}\sum_{i=1}^n F_iF_i'\}(\hat{w}_j - w_j) \qquad \text{(S2.44)}\\ \leq &\{\lambda_n^* + \|n^{-1}\sum_{i=1}^n \breve{F}_i(\breve{E}_{ij} - F_i'w_j)\|_\infty\} \cdot \|\big(\hat{\Lambda}_{\mathcal{H}_n^c}^{1/2}(\hat{w}_j - w_j)\big)_{\mathcal{S}_j}\|_1 -\\ &\{\lambda_n^* - \|n^{-1}\sum_{i=1}^n \breve{F}_i(\breve{E}_{ij} - F_i'w_j)\|_\infty\} \cdot \|\big(\hat{\Lambda}_{\mathcal{H}_n^c}^{1/2}(\hat{w}_j - w_j)\big)_{\mathcal{S}_j^c}\|_1\\ \leq &\{\lambda_n^* + \max_{l \leq h_n s_n}\|n^{-1}\sum_{i=1}^n \breve{F}_i(\breve{E}_{il} - F_i'w_l)\|_\infty\} \cdot \|\big(\hat{\Lambda}_{\mathcal{H}_n^c}^{1/2}(\hat{w}_j - w_j)\big)_{\mathcal{S}_j}\|_1 -\\ &\{\lambda_n^* - \max_{l \leq h_n s_n}\|n^{-1}\sum_{i=1}^n \breve{F}_i(\breve{E}_{il} - F_i'w_l)\|_\infty\} \cdot \|\big(\hat{\Lambda}_{\mathcal{H}_n^c}^{1/2}(\hat{w}_j - w_j)\big)_{\mathcal{S}_j^c}\|_1.\end{aligned}$$

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

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

By choosing $K_1 \geq 2c_1$ in (A5.3), we have that with probability tending to 1:

$$\max_{j \leq h_n s_n}\|n^{-1}\sum_{i=1}^n \breve{F}_i(\breve{E}_{ij} - F_i'w_j)\|_\infty \leq 2^{-1}\lambda_n^*. \qquad \text{(S2.45)}$$

It follows from (S2.44) and (S2.45) that

$$P\big[\bigcap_{j=1}^{h_n s_n}\big\{\|\big(\hat{\Lambda}_{\mathcal{H}_n^c}^{1/2}(\hat{w}_j - w_j)\big)_{\mathcal{S}_j^c}\|_1 \leq 3\|\big(\hat{\Lambda}_{\mathcal{H}_n^c}^{1/2}(\hat{w}_j - w_j)\big)_{\mathcal{S}_j}\|_1\big\}\big] \to 1,$$

which further implies that

$$P\big[\bigcap_{j=1}^{h_n s_n}\big\{\|\hat{\Lambda}_{\mathcal{H}_n^c}^{1/2}(\hat{w}_j - w_j)\|_1 \leq 4\rho_{nj}^{1/2}\|\hat{\Lambda}_{\mathcal{H}_n^c}^{1/2}(\hat{w}_j - w_j)\|_2\big\}\big] \to 1. \tag{S2.46}$$

Based on (3.10) of the main paper and the Karush-Kuhn-Tucker condition, it is seen that

$$\max_{j\leq h_n s_n}\|n^{-1}\sum_{i=1}^{n}\breve{F}_i(\breve{E}_{ij} - F_i'\hat{w}_j)\|_\infty \leq \lambda_n^*. \tag{S2.47}$$

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

$$\begin{aligned}
&\max_{j\leq h_n s_n + k_n}\|n^{-1}\sum_{i=1}^{n}\breve{F}_i F_i'(\hat{w}_j - w_j)\|_\infty \\
\leq &\max_{j\leq h_n s_n}\|n^{-1}\sum_{i=1}^{n}\breve{F}_i(\breve{E}_{ij} - F_i'\hat{w}_j)\|_\infty + \max_{j\leq h_n s_n}\|n^{-1}\sum_{i=1}^{n}\breve{F}_i(\breve{E}_{ij} - F_i' w_j)\|_\infty \\
\leq &3\lambda_n^*/2 \leq (3K_2/2)\cdot\{\log(np_n s_n)/n\}^{1/2},
\end{aligned} \tag{S2.48}$$

with probability tending to 1, where the second inequality follows from (S2.45) and (S2.47), and the last inequality holds from (A5.3). Together with Lemma 2, it can be deduced that with probability tending to 1:

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

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

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

Together with (S2.46) and (S2.48) yields

$$
\begin{aligned}
&P\Big[\bigcap_{j=1}^{h_n s_n} \big\{(\hat{w}_j - w_j)'(n^{-1}\sum_{i=1}^{n} F_i F_i')(\hat{w}_j - w_j) \le 6K_2\{\rho_{nj}\log(np_n s_n)/n\}^{1/2} \\
&\quad \|\hat{\Lambda}^{1/2}_{\mathcal{H}_n^c}(\hat{w}_j - w_j)\|_2\big\}\Big] \to 1. \qquad \text{(S2.49)}
\end{aligned}
$$

Also note that for all $j \le h_n s_n$,

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

for some universal constant $c_2 > 0$, where the last inequality is by (A3.3). Together with Lemma 2, (S2.46), and (A5.4), it can be deduced that

$$
P\Big[\bigcap_{j=1}^{h_n s_n} \big\{(\hat{w}_j - w_j)'(n^{-1}\sum_{i=1}^{n} F_i F_i')(\hat{w}_j - w_j) \ge 2^{-1}c_2\|\hat{\Lambda}^{1/2}_{\mathcal{H}_n^c}(\hat{w}_j - w_j)\|_2^2\big\}\Big] \to 1.
$$

Together with (S2.49) yields

$$P\Big[\bigcap_{j=1}^{h_ns_n}\big\{\|\hat{\Lambda}_{\mathcal{H}_n^c}^{1/2}(\hat{w}_j-w_j)\|_2\leq 12c_2^{-1}K_2\{\rho_{nj}\log(np_ns_n)/n\}^{1/2}\big\}\Big]\to 1, \tag{S2.50}$$

which further implies that

$$P\Big[\max_{j\leq h_ns_n}\|\hat{\Lambda}_{\mathcal{H}_n^c}^{1/2}(\hat{w}_j-w_j)\|_2\leq 12c_2^{-1}K_2\{\rho_n\log(np_ns_n)/n\}^{1/2}\Big]\to 1.$$

This completes the proof of part 2). By combining (S2.50) with (S2.46), it can be deduced that

$$P\Big[\bigcap_{j=1}^{h_ns_n}\big\{\|\hat{\Lambda}_{\mathcal{H}_n^c}^{1/2}(\hat{w}_j-w_j)\|_1\leq 48c_2^{-1}K_2\{\rho_{nj}^2\log(np_ns_n)/n\}^{1/2}\big\}\Big]\to 1,$$

which further implies that

$$P\Big[\max_{j\leq h_ns_n}\|\hat{\Lambda}_{\mathcal{H}_n^c}^{1/2}(\hat{w}_j-w_j)\|_1\leq 48c_2^{-1}K_2\{\rho_n^2\log(np_ns_n)/n\}^{1/2}\Big]\to 1.$$

This completes the proof of part 3). □

Next, we state Lemma 7 as follows.

Lemma 7. *Under conditions (A1)–(A5) and (B1)–(B5), we have*

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

$$\max_{l \le h_n s_n} |n^{-1/2} \sum_{i=1}^{n} (S_{il} - S_{il}^*)|$$

$$\le c_1 n^{-1/2} \rho_n \{\log(np_n s_n)\}^{3/2} \exp\{c_1 q_n \log^{1/2}(nq_n s_n)\} +$$

$$c_1 n^{1/2} q_n s_n^{-\delta+1/2} \exp\{c_1 q_n \log^{1/2}(nq_n s_n)\} +$$

$$c_1 \lambda_n n^{-1/18} s_n^{1/2} q_n \{\log(np_n s_n)\}^{1/2} \exp\{c_1 q_n \log^{1/2}(nq_n s_n)\} +$$

$$c_1 \lambda_n^2 n^{7/18} q_n \{\log(np_n s_n)\}^{1/2} \exp\{c_1 q_n \log^{1/2}(nq_n s_n)\} +$$

$$c_1 \lambda_n n^{-1/18} \rho_n^{1/2} q_n^{1/2} \{\log(np_n s_n)\}^{1/2} \exp\{c_1 q_n \log^{1/2}(nq_n s_n)\}.$$

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

$$\max_{l \le h_n s_n} n^{-1} \sum_{i=1}^{n} (\hat{S}_{il} - S_{il}^*)^2$$

$$\le c_2 q_n^2 s_n^{-2\delta+1} \{\log(np_n s_n)\} \exp\{c_2 q_n \log^{1/2}(nq_n s_n)\} +$$

$$c_2 \lambda_n^2 n^{-1/9} q_n \{\log(np_n s_n)\} \exp\{c_2 q_n \log^{1/2}(nq_n s_n)\} +$$

$$c_2 n^{-1} \rho_n \log^2(n) \{\log(np_n s_n)\} \exp\{c_2 q_n \log^{1/2}(nq_n s_n)\}.$$

3) There is a universal constant $c_3 > 0$ that with probability tending to 1:

$$\max_{l \le h_n s_n} n^{-1} \sum_{i=1}^{n} \hat{S}_{il}^2 \le c_3 \log^2(n) \exp\{c_3 q_n \log^{1/2}(nq_n s_n)\}.$$

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

$$\max_{l\le h_ns_n}|n^{-1/2}\sum_{i=1}^{n}(S_{il}-S^*_{il})|$$
$$\le\Delta_1+\Delta_2+\Delta_3+\Delta_4+\Delta_5+\Delta_6+\Delta_7+\Delta_8, \qquad \text{(S2.51)}$$

where

$$\Delta_1=n^{-1/2}\max_{l\le h_ns_n}|(\hat{w}_l-w_l)'\sum_{i=1}^{n}F_i\epsilon_i|,$$
$$\Delta_2=n^{-1/2}\max_{l\le h_ns_n}|\sum_{i=1}^{n}(\tilde{E}_{il}-\breve{E}_{il})\epsilon_i|,$$
$$\Delta_3=n^{-1/2}\max_{l\le h_ns_n}|\sum_{i=1}^{n}(w_l'F_i-\tilde{E}_{il})\{b'(\alpha_0+\sum_{j=1}^{p_n}\sum_{k=1}^{\infty}\theta_{ijk}\eta_{jk})-b'(\alpha_0+\sum_{j=1}^{p_n}\sum_{k=1}^{s_n}\theta_{ijk}\eta_{jk})\}|,$$
$$\Delta_4=n^{-1/2}\max_{l\le h_ns_n}|(\hat{w}_l-w_l)'\sum_{i=1}^{n}F_i\{b'(\alpha_0+\sum_{j=1}^{p_n}\sum_{k=1}^{\infty}\theta_{ijk}\eta_{jk})-b'(\alpha_0+\sum_{j=1}^{p_n}\sum_{k=1}^{s_n}\theta_{ijk}\eta_{jk})\}|,$$
$$\Delta_5=n^{-1/2}\max_{l\le h_ns_n}|\sum_{i=1}^{n}(\tilde{E}_{il}-\breve{E}_{il})\{b'(\alpha_0+\sum_{j=1}^{p_n}\sum_{k=1}^{\infty}\theta_{ijk}\eta_{jk})-b'(\alpha_0+\sum_{j=1}^{p_n}\sum_{k=1}^{s_n}\theta_{ijk}\eta_{jk})\}|,$$
$$\Delta_6=n^{-1/2}\max_{l\le h_ns_n}|\sum_{i=1}^{n}(w_l'F_i-\tilde{E}_{il})\{b'(\hat{\alpha}_0+E_i'\eta_{\mathcal{H}_n}+F_i'\hat{\eta}_{\mathcal{H}_n^c})-b'(\alpha_0+E_i'\eta_{\mathcal{H}_n}+F_i'\eta_{\mathcal{H}_n^c})\}|,$$
$$\Delta_7=n^{-1/2}\max_{l\le h_ns_n}|(\hat{w}_l-w_l)'\sum_{i=1}^{n}F_i\{b'(\hat{\alpha}_0+E_i'\eta_{\mathcal{H}_n}+F_i'\hat{\eta}_{\mathcal{H}_n^c})-b'(\alpha_0+E_i'\eta_{\mathcal{H}_n}+F_i'\eta_{\mathcal{H}_n^c})\}|,$$
$$\Delta_8=n^{-1/2}\max_{l\le h_ns_n}|\sum_{i=1}^{n}(\tilde{E}_{il}-\breve{E}_{il})\{b'(\hat{\alpha}_0+E_i'\eta_{\mathcal{H}_n}+F_i'\hat{\eta}_{\mathcal{H}_n^c})-b'(\alpha_0+E_i'\eta_{\mathcal{H}_n}+F_i'\eta_{\mathcal{H}_n^c})\}|.$$

To bound Δ_1, note that

$$\begin{aligned}\Delta_1 =& n^{1/2} \max_{l \leq h_n s_n} |\{\hat{\Lambda}^{1/2}_{\mathcal{H}_n^c}(\hat{w}_l - w_l)\}' n^{-1} \sum_{i=1}^{n} \breve{F}_i \epsilon_i| \\ \leq& n^{1/2} \|n^{-1} \sum_{i=1}^{n} \breve{F}_i \epsilon_i\|_\infty \cdot \max_{l \leq h_n s_n} \|\hat{\Lambda}^{1/2}_{\mathcal{H}_n^c}(\hat{w}_l - w_l)\|_1 \\ \leq& c_1 n^{-1/2} \rho_n \{\log(n p_n s_n)\}^{3/2} \exp\{c_1 q_n \log^{1/2}(n q_n s_n)\}, \end{aligned} \tag{S2.52}$$

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 3 and Lemma 6. To bound Δ_2, note that

$$\begin{aligned}\Delta_2 \leq& n^{1/2} \|\Lambda^{1/2} \hat{\Lambda}^{-1/2} - I\|_\infty \cdot \|n^{-1} \sum_{i=1}^{n} \tilde{G}_i \epsilon_i\|_\infty \\ \leq& c_2 n^{-1/2} \{\log(n p_n s_n)\}^{3/2} \exp\{c_2 q_n \log^{1/2}(n q_n s_n)\}, \end{aligned} \tag{S2.53}$$

for some universal constant $c_2 > 0$, with probability tending to 1, where the last inequality is based on Lemma 2 and Lemma 3. To bound Δ_3, recall that

$\delta_i^* = \alpha_0 + \sum_{j=1}^{p_n} \sum_{k=1}^{s_n} \theta_{ijk}\eta_{jk}$ and $\delta_i = \alpha_0 + \sum_{j=1}^{p_n} \sum_{k=1}^{\infty} \theta_{ijk}\eta_{jk}$, we have

$$\begin{aligned}
\Delta_3 =& n^{-1/2} \max_{l \le h_n s_n} |\sum_{i=1}^n (w_l' F_i - \tilde{E}_{il})\{b'(\delta_i) - b'(\delta_i^*)\}| \\
=& n^{-1/2} \max_{l \le h_n s_n} |\sum_{i=1}^n (w_l' F_i - \tilde{E}_{il})(\sum_{j=1}^{q_n} \sum_{k=s_n+1}^{\infty} \theta_{ijk}\eta_{jk}) b''(\delta_i^* + t_i \sum_{j=1}^{q_n} \sum_{k=s_n+1}^{\infty} \theta_{ijk}\eta_{jk})| \\
\le& \{\max_{l \le h_n s_n} n^{-1} \sum_{i=1}^n (w_l' F_i - \tilde{E}_{il})^2\}^{1/2} \cdot \{\sum_{i=1}^n (\sum_{j=1}^{q_n} \sum_{k=s_n+1}^{\infty} \theta_{ijk}\eta_{jk})^2\}^{1/2} \\
& \cdot \max_{i \le n} |b''(\delta_i^* + t_i \sum_{j=1}^{q_n} \sum_{k=s_n+1}^{\infty} \theta_{ijk}\eta_{jk})| \\
\le& c_3 n^{1/2} q_n s_n^{-\delta+1/2} \exp\{c_3 q_n \log^{1/2}(n q_n s_n)\}, \qquad \text{(S2.54)}
\end{aligned}$$

for some universal constant $c_3 > 0$, with probability tending to 1, where the second equality holds from (S2.35), and the last inequality is based on (S2.36), Lemma 2, Lemma 3, and (A2.1). To bound Δ_4, note that

$$\begin{aligned}
\Delta_4 =& n^{-1/2} \max_{l \le h_n s_n} |(\hat{w}_l - w_l)' \sum_{i=1}^n F_i \{b'(\delta_i) - b'(\delta_i^*)\}| \\
=& n^{-1/2} \max_{l \le h_n s_n} |(\hat{w}_l - w_l)' \sum_{i=1}^n F_i (\sum_{j=1}^{q_n} \sum_{k=s_n+1}^{\infty} \theta_{ijk}\eta_{jk}) b''(\delta_i^* + t_i \sum_{j=1}^{q_n} \sum_{k=s_n+1}^{\infty} \theta_{ijk}\eta_{jk})| \\
\le& \{\max_{l \le h_n s_n} (\hat{w}_l - w_l)' (n^{-1} \sum_{i=1}^n F_i F_i')(\hat{w}_l - w_l)\}^{1/2} \cdot \{\sum_{i=1}^n (\sum_{j=1}^{q_n} \sum_{k=s_n+1}^{\infty} \theta_{ijk}\eta_{jk})^2\}^{1/2} \\
& \cdot \max_{i \le n} |b''(\delta_i^* + t_i \sum_{j=1}^{q_n} \sum_{k=s_n+1}^{\infty} \theta_{ijk}\eta_{jk})|, \qquad \text{(S2.55)}
\end{aligned}$$

where the second equality is based on (S2.35). For the term $\max_{l \le h_n s_n}(\hat{w}_l -$

$w_l)'(n^{-1}\sum_{i=1}^n F_iF_i')(\hat{w}_l - w_l)$, note that

$$\begin{aligned}
&\max_{l\le h_ns_n}(\hat{w}_l - w_l)'(n^{-1}\sum_{i=1}^n F_iF_i')(\hat{w}_l - w_l)\\
\le&\max_{l\le h_ns_n}\|n^{-1}\sum_{i=1}^n \breve{F}_iF_i'(\hat{w}_l - w_l)\|_\infty\cdot\max_{l\le h_ns_n}\|\hat{\Lambda}^{1/2}_{\mathcal{H}_n^c}(\hat{w}_l - w_l)\|_1\\
\lesssim& n^{-1}\rho_n\log(np_ns_n), \qquad\qquad \text{(S2.56)}
\end{aligned}$$

with probability tending to 1, where the last inequality holds from Lemma 6. By combining (S2.56), (S2.36), Lemma 3 with (S2.55), we have

$$\Delta_4 \le c_4\rho_n^{1/2}q_ns_n^{-\delta+1/2}\{\log(np_ns_n)\}^{1/2}\exp\{c_4q_n\log^{1/2}(nq_ns_n)\}, \qquad \text{(S2.57)}$$

for some universal constant $c_4 > 0$, with probability tending to 1. To bound Δ_5, note that

$$\begin{aligned}
\Delta_5 =& n^{-1/2}\max_{l\le h_ns_n}|\sum_{i=1}^n(\tilde{E}_{il} - \breve{E}_{il})\{b'(\delta_i) - b'(\delta_i^*)\}|\\
=& n^{-1/2}\max_{l\le h_ns_n}|\sum_{i=1}^n(\tilde{E}_{il} - \breve{E}_{il})(\sum_{j=1}^{q_n}\sum_{k=s_n+1}^{\infty}\theta_{ijk}\eta_{jk})b''(\delta_i^* + t_i\sum_{j=1}^{q_n}\sum_{k=s_n+1}^{\infty}\theta_{ijk}\eta_{jk})|\\
\le&\{\max_{l\le h_ns_n}n^{-1}\sum_{i=1}^n(\tilde{E}_{il} - \breve{E}_{il})^2\}^{1/2}\cdot\{\sum_{i=1}^n(\sum_{j=1}^{q_n}\sum_{k=s_n+1}^{\infty}\theta_{ijk}\eta_{jk})^2\}^{1/2}\cdot\max_{i\le n}|b''(\delta_i^* + t_i\sum_{j=1}^{q_n}\sum_{k=s_n+1}^{\infty}\theta_{ijk}\eta_{jk})|\\
\le&\|\Lambda^{1/2}\hat{\Lambda}^{-1/2} - I\|_\infty\cdot\{\max_{l\le p_ns_n}n^{-1}\sum_{i=1}^n\tilde{G}_{il}^2\}^{1/2}\cdot\{\sum_{i=1}^n(\sum_{j=1}^{q_n}\sum_{k=s_n+1}^{\infty}\theta_{ijk}\eta_{jk})^2\}^{1/2}\\
&\cdot\max_{i\le n}|b''(\delta_i^* + t_i\sum_{j=1}^{q_n}\sum_{k=s_n+1}^{\infty}\theta_{ijk}\eta_{jk})|\\
\le& c_5q_ns_n^{-\delta+1/2}\{\log(np_ns_n)\}^{1/2}\exp\{c_5q_n\log^{1/2}(nq_ns_n)\}, \qquad \text{(S2.58)}
\end{aligned}$$

for some universal constant $c_5 > 0$, with probability tending to 1, where the second equality is by (S2.35), and the last inequality is based on (S2.36), Lemma 2, Lemma 3, and (A2.1). It follows from mean value theorem that

$$
\begin{aligned}
&b'(\hat{\alpha}_0 + E_i'\eta_{\mathcal{H}_n} + F_i'\hat{\eta}_{\mathcal{H}_n^c}) - b'(\alpha_0 + E_i'\eta_{\mathcal{H}_n} + F_i'\eta_{\mathcal{H}_n^c}) \\
=&b''(\delta_i^* + \bar{t}_i(\hat{\alpha}_0 - \alpha_0) + \bar{t}_i F_i'(\hat{\eta}_{\mathcal{H}_n^c} - \eta_{\mathcal{H}_n^c})) \cdot \{(\hat{\alpha}_0 - \alpha_0) + F_i'(\hat{\eta}_{\mathcal{H}_n^c} - \eta_{\mathcal{H}_n^c})\},
\end{aligned}
\tag{S2.59}
$$

for some $\bar{t}_i \in [0,1]$. In addition, we have

$$
\begin{aligned}
&\max_{i\le n}|b''(\delta_i^* + \bar{t}_i(\hat{\alpha}_0 - \alpha_0) + \bar{t}_i F_i'(\hat{\eta}_{\mathcal{H}_n^c} - \eta_{\mathcal{H}_n^c}))| \\
\le&\exp\{\max_{i\le n}|\delta_i^* + \bar{t}_i(\hat{\alpha}_0 - \alpha_0) + \bar{t}_i F_i'(\hat{\eta}_{\mathcal{H}_n^c} - \eta_{\mathcal{H}_n^c})|\} \\
\le&\exp(\max_{i\le n}|\delta_i^*|)\cdot\exp\{|\hat{\alpha}_0 - \alpha_0| + \max_{i\le n}\max_{l\le p_n s_n}|\tilde{G}_{il}|\cdot\|\Lambda^{1/2}(\hat{\eta} - \eta)\|_1\} \\
\le&c_6\exp\{c_6 q_n \log^{1/2}(n q_n s_n)\},
\end{aligned}
\tag{S2.60}
$$

for some universal constant $c_6 > 0$, with probability tending to 1, where the first inequality is by (A1), and the last inequality is based on Lemma 2, Lemma 3, and Theorem 1. Moreover, we have that for every $i \le n$,

$$
\begin{aligned}
&|b''(\delta_i^* + \bar{t}_i(\hat{\alpha}_0 - \alpha_0) + \bar{t}_i F_i'(\hat{\eta}_{\mathcal{H}_n^c} - \eta_{\mathcal{H}_n^c})) - b''(\delta_i^*)| \\
\le&\exp(3\max_{i\le n}|\delta_i^*|)\cdot\exp(3|\hat{\alpha}_0 - \alpha_0|)\cdot\exp\{3\max_{i\le n}|F_i'(\hat{\eta}_{\mathcal{H}_n^c} - \eta_{\mathcal{H}_n^c})|\}\cdot|(\hat{\alpha}_0 - \alpha_0) + F_i'(\hat{\eta}_{\mathcal{H}_n^c} - \eta_{\mathcal{H}_n^c})| \\
\le&\exp(3\max_{i\le n}|\delta_i^*|)\cdot\exp(3|\hat{\alpha}_0 - \alpha_0|)\cdot\exp\{3\max_{i\le n}\max_{l\le p_n s_n}|\tilde{G}_{il}|\cdot\|\Lambda^{1/2}(\hat{\eta} - \eta)\|_1\}\cdot \\
&|(\hat{\alpha}_0 - \alpha_0) + F_i'(\hat{\eta}_{\mathcal{H}_n^c} - \eta_{\mathcal{H}_n^c})|,
\end{aligned}
$$

where the first inequality is by (A1). Together with Lemma 2, Lemma 3, and

Theorem 1, it can be deduced that there is a universal constant $c_7 > 0$ such that

$$P\Big[\cap_{i=1}^{n}\big\{|b''(\delta_i^* + \bar{t}_i(\hat{\alpha}_0 - \alpha_0) + \bar{t}_i F_i'(\hat{\eta}_{\mathcal{H}_n^c} - \eta_{\mathcal{H}_n^c})) - b''(\delta_i^*)| \le c_7 \exp\{c_7 q_n \log^{1/2}(n q_n s_n)\}|(\hat{\alpha}_0 - \alpha_0) + F_i'(\hat{\eta}_{\mathcal{H}_n^c} - \eta_{\mathcal{H}_n^c})|\big\}\Big] \to 1. \quad \text{(S2.61)}$$

To bound Δ_6, note that

$$\begin{aligned}\Delta_6 &= n^{-1/2} \max_{l \le h_n s_n} |\sum_{i=1}^{n} (w_l' F_i - \tilde{E}_{il}) b''(\delta_i^* + \bar{t}_i(\hat{\alpha}_0 - \alpha_0) + \bar{t}_i F_i'(\hat{\eta}_{\mathcal{H}_n^c} - \eta_{\mathcal{H}_n^c}))\{(\hat{\alpha}_0 - \alpha_0) + F_i'(\hat{\eta}_{\mathcal{H}_n^c} - \eta_{\mathcal{H}_n^c})\}| \\ &\le \Omega_1 + \Omega_2, \quad \text{(S2.62)}\end{aligned}$$

where the first equality is by (S2.59), and

$$\begin{aligned}\Omega_1 =& n^{-1/2} \max_{l \le h_n s_n} |\sum_{i=1}^{n} (w_l' F_i - \tilde{E}_{il}) b''(\delta_i^*)\{(\hat{\alpha}_0 - \alpha_0) + F_i'(\hat{\eta}_{\mathcal{H}_n^c} - \eta_{\mathcal{H}_n^c})\}|, \\ \Omega_2 =& n^{-1/2} \max_{l \le h_n s_n} |\sum_{i=1}^{n} (w_l' F_i - \tilde{E}_{il})\{b''(\delta_i^* + \bar{t}_i(\hat{\alpha}_0 - \alpha_0) + \bar{t}_i F_i'(\hat{\eta}_{\mathcal{H}_n^c} - \eta_{\mathcal{H}_n^c})) - b''(\delta_i^*)\} \cdot \\ & \{(\hat{\alpha}_0 - \alpha_0) + F_i'(\hat{\eta}_{\mathcal{H}_n^c} - \eta_{\mathcal{H}_n^c})\}|.\end{aligned}$$

To bound Ω_1, note that

$$\begin{aligned}\Omega_1 \le& n^{1/2} \max_{l \le h_n s_n} \|n^{-1} \sum_{i=1}^{n} \binom{\tilde{F}_i}{1} (w_l{}' F_i - \tilde{E}_{il}) b''(\delta_i^*)\|_\infty \{|\hat{\alpha}_0 - \alpha_0| + \|\Lambda^{1/2}(\hat{\eta} - \eta)\|_1\} \\ \le& c_8 n^{1/2} \max_{i \le n} b''(\delta_i^*) \cdot \max_{l \le h_n s_n} \|n^{-1} \sum_{i=1}^{n} \binom{\tilde{F}_i}{1} (w_l{}' F_i - \tilde{E}_{il})\|_\infty \cdot \{|\hat{\alpha}_0 - \alpha_0| + \|\Lambda^{1/2}(\hat{\eta} - \eta)\|_1\} \\ \le& c_8 n^{1/2} \exp(\max_{i \le n} |\delta_i^*|) \cdot \max_{l \le h_n s_n} \|n^{-1} \sum_{i=1}^{n} \binom{\tilde{F}_i}{1} (w_l{}' F_i - \tilde{E}_{il})\|_\infty \cdot \{|\hat{\alpha}_0 - \alpha_0| + \|\Lambda^{1/2}(\hat{\eta} - \eta)\|_1\} \\ \le& c_9 \lambda_n n^{-1/18} s_n^{1/2} q_n \{\log(n p_n s_n)\}^{1/2} \exp\{c_9 q_n \log^{1/2}(n q_n s_n)\}, \quad \text{(S2.63)}\end{aligned}$$

for some universal constants $c_8, c_9 > 0$, with probability tending to 1, where the second inequality is by (A2), the third inequality is by (A1), and the last inequality is due to Lemma 2, Lemma 3, and Theorem 1. To bound Ω_2, note that

$$\begin{aligned}\Omega_2 \le & c_7 n^{1/2}\exp\{c_7 q_n \log^{1/2}(nq_n s_n)\}\max_{l\le h_n s_n}\sum_{i=1}^{n} n^{-1}|w_l' F_i - \tilde{E}_{il}|\{(\hat{\alpha}_0-\alpha_0)+F_i'(\hat{\eta}_{\mathcal{H}_n^c}-\eta_{\mathcal{H}_n^c})\}^2\\ \le & 2c_7 n^{1/2}\exp\{c_7 q_n \log^{1/2}(nq_n s_n)\}\cdot(\max_{l\le h_n s_n}\max_{i\le n}|w_l' F_i - \tilde{E}_{il}|)\cdot\\ & \{(\hat{\alpha}_0-\alpha_0)^2+(\hat{\eta}_{\mathcal{H}_n^c}-\eta_{\mathcal{H}_n^c})'(n^{-1}\sum_{i=1}^{n}F_iF_i')(\hat{\eta}_{\mathcal{H}_n^c}-\eta_{\mathcal{H}_n^c})\},\end{aligned} \tag{S2.64}$$

with probability tending to 1, where the first inequality is by (S2.61). To bound the term $(\hat{\eta}_{\mathcal{H}_n^c}-\eta_{\mathcal{H}_n^c})'(n^{-1}\sum_{i=1}^{n}F_iF_i')(\hat{\eta}_{\mathcal{H}_n^c}-\eta_{\mathcal{H}_n^c})$, note that

$$\begin{aligned}&(\hat{\eta}_{\mathcal{H}_n^c}-\eta_{\mathcal{H}_n^c})'(n^{-1}\sum_{i=1}^{n}F_iF_i')(\hat{\eta}_{\mathcal{H}_n^c}-\eta_{\mathcal{H}_n^c})\\ =&\{\Lambda_{\mathcal{H}_n^c}^{1/2}(\hat{\eta}_{\mathcal{H}_n^c}-\eta_{\mathcal{H}_n^c})\}'\{n^{-1}\sum_{i=1}^{n}\tilde{F}_i\tilde{F}_i' - E(\tilde{F}_1\tilde{F}_1')\}\{\Lambda_{\mathcal{H}_n^c}^{1/2}(\hat{\eta}_{\mathcal{H}_n^c}-\eta_{\mathcal{H}_n^c})\}+\\ &\{\Lambda_{\mathcal{H}_n^c}^{1/2}(\hat{\eta}_{\mathcal{H}_n^c}-\eta_{\mathcal{H}_n^c})\}'E(\tilde{F}_1\tilde{F}_1')\{\Lambda_{\mathcal{H}_n^c}^{1/2}(\hat{\eta}_{\mathcal{H}_n^c}-\eta_{\mathcal{H}_n^c})\}\\ \le&\lambda_{\max}(E(\tilde{F}_1\tilde{F}_1'))\cdot\|\Lambda^{1/2}(\hat{\eta}-\eta)\|_2^2+\|n^{-1}\sum_{i=1}^{n}\tilde{F}_i\tilde{F}_i' - E(\tilde{F}_1\tilde{F}_1')\|_\infty\cdot\|\Lambda^{1/2}(\hat{\eta}-\eta)\|_1^2\\ \le& c_{10}\lambda_n^2 q_n n^{-1/9},\end{aligned} \tag{S2.65}$$

for some universal constants $c_{10} > 0$, with probability tending to 1, where the last inequality is based on (A3.3), Lemma 2, and Theorem 1. By combining (S2.65), Lemma 2, Theorem 1 with (S2.64), it can be deduced that

$$\Omega_2 \le c_{11}\lambda_n^2 n^{7/18} q_n\{\log(np_n s_n)\}^{1/2}\exp\{c_{11} q_n \log^{1/2}(nq_n s_n)\}, \tag{S2.66}$$

for some universal constants $c_{11} > 0$, with probability tending to 1. By combining (S2.66), (S2.63) with (S2.62), it can be deduced that

$$\Delta_6 \leq c_{12}\lambda_n n^{-1/18} s_n^{1/2} q_n \{\log(np_n s_n)\}^{1/2} \exp\{c_{12} q_n \log^{1/2}(nq_n s_n)\} + c_{12}\lambda_n^2 n^{7/18} q_n \{\log(np_n s_n)\}^{1/2} \exp\{c_{12} q_n \log^{1/2}(nq_n s_n)\}, \tag{S2.67}$$

for some universal constants $c_{12} > 0$, with probability tending to 1. To bound Δ_7, note that

$$\begin{aligned}
&\Delta_7 \\
=& n^{-1/2} \max_{l \leq h_n s_n} |(\hat{w}_l - w_l)' \sum_{i=1}^n F_i b''(\delta_i^* + \bar{t}_i(\hat{\alpha}_0 - \alpha_0) + \bar{t}_i F_i'(\hat{\eta}_{\mathcal{H}_n^c} - \eta_{\mathcal{H}_n^c}))\{(\hat{\alpha}_0 - \alpha_0) + F_i'(\hat{\eta}_{\mathcal{H}_n^c} - \eta_{\mathcal{H}_n^c})\}| \\
\leq& 2n^{1/2} \max_{i \leq n} |b''(\delta_i^* + \bar{t}_i(\hat{\alpha}_0 - \alpha_0) + \bar{t}_i F_i'(\hat{\eta}_{\mathcal{H}_n^c} - \eta_{\mathcal{H}_n^c}))| \cdot \{\max_{l \leq h_n s_n} (\hat{w}_l - w_l)'(n^{-1} \sum_{i=1}^n F_i F_i')(\hat{w}_l - w_l)\}^{1/2} \\
&\cdot \{(\hat{\alpha}_0 - \alpha_0)^2 + (\hat{\eta}_{\mathcal{H}_n^c} - \eta_{\mathcal{H}_n^c})'(n^{-1} \sum_{i=1}^n F_i F_i')(\hat{\eta}_{\mathcal{H}_n^c} - \eta_{\mathcal{H}_n^c})\}^{1/2} \\
\leq& c_{13}\lambda_n n^{-1/18} \rho_n^{1/2} q_n^{1/2} \{\log(np_n s_n)\}^{1/2} \exp\{c_{13} q_n \log^{1/2}(nq_n s_n)\},
\end{aligned} \tag{S2.68}$$

for some universal constants $c_{13} > 0$, with probability tending to 1, where the first equality is by (S2.59), and the last inequality is based on (S2.56), (S2.60), (S2.65), and Theorem 1. To bound Δ_8, note that

$$\begin{aligned}\Delta_8 =& n^{-1/2}\max_{l\le h_n s_n}|\sum_{i=1}^{n}(\tilde{E}_{il}-\breve{E}_{il})b''(\delta_i^*+\bar{t}_i(\hat{\alpha}_0-\alpha_0)+\bar{t}_iF_i'(\hat{\eta}_{\mathcal{H}_n^c}-\eta_{\mathcal{H}_n^c}))\{(\hat{\alpha}_0-\alpha_0)+F_i'(\hat{\eta}_{\mathcal{H}_n^c}-\eta_{\mathcal{H}_n^c})\}| \\ \le& 2n^{1/2}\|\Lambda^{1/2}\hat{\Lambda}^{-1/2}-I\|_\infty\cdot\max_{i\le n}|b''(\delta_i^*+\bar{t}_i(\hat{\alpha}_0-\alpha_0)+\bar{t}_iF_i'(\hat{\eta}_{\mathcal{H}_n^c}-\eta_{\mathcal{H}_n^c}))|\cdot(\max_{l\le p_n s_n}n^{-1}\sum_{i=1}^{n}\tilde{G}_{il}^2)^{1/2} \\ &\cdot\{(\hat{\alpha}_0-\alpha_0)^2+(\hat{\eta}_{\mathcal{H}_n^c}-\eta_{\mathcal{H}_n^c})'(n^{-1}\sum_{i=1}^{n}F_iF_i')(\hat{\eta}_{\mathcal{H}_n^c}-\eta_{\mathcal{H}_n^c})\}^{1/2} \\ \le& c_{14}\lambda_n n^{-1/18}q_n^{1/2}\{\log(np_ns_n)\}^{1/2}\exp\{c_{14}q_n\log^{1/2}(nq_ns_n)\}, \qquad \text{(S2.69)}\end{aligned}$$

for some universal constants $c_{14}>0$, with probability tending to 1, where the first equality is by (S2.59), and the last inequality is based on (S2.39), (S2.60), (S2.65), Lemma 2 and Theorem 1. By combining (S2.69), (S2.68), (S2.67), (S2.58), (S2.57), (S2.54), (S2.53), (S2.52) with (S2.51), it can be concluded that there exists a universal constant $c_{15}>0$ such that

$$\begin{aligned}&\max_{l\le h_n s_n}|n^{-1/2}\sum_{i=1}^{n}(S_{il}-S_{il}^*)| \\ \le& c_{15}n^{-1/2}\rho_n\{\log(np_ns_n)\}^{3/2}\exp\{c_{15}q_n\log^{1/2}(nq_ns_n)\}+ \\ & c_{15}n^{1/2}q_ns_n^{-\delta+1/2}\exp\{c_{15}q_n\log^{1/2}(nq_ns_n)\}+ \\ & c_{15}\lambda_n n^{-1/18}s_n^{1/2}q_n\{\log(np_ns_n)\}^{1/2}\exp\{c_{15}q_n\log^{1/2}(nq_ns_n)\}+ \\ & c_{15}\lambda_n^2n^{7/18}q_n\{\log(np_ns_n)\}^{1/2}\exp\{c_{15}q_n\log^{1/2}(nq_ns_n)\}+ \\ & c_{15}\lambda_n n^{-1/18}\rho_n^{1/2}q_n^{1/2}\{\log(np_ns_n)\}^{1/2}\exp\{c_{15}q_n\log^{1/2}(nq_ns_n)\},\end{aligned}$$

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

2), it follows from triangle inequality that

$$\max_{l\le h_n s_n} n^{-1}\sum_{i=1}^{n}(\hat{S}_{il}-S^*_{il})^2$$

$$\lesssim \Delta_1^*+\Delta_2^*+\Delta_3^*+\Delta_4^*+\Delta_5^*+\Delta_6^*+\Delta_7^*+\Delta_8^*, \tag{S2.70}$$

where

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

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

$$\Delta_3^* = \max_{l\le h_n s_n} n^{-1}\sum_{i=1}^{n}[(w_l'F_i-\tilde{E}_{il})\{b'(\alpha_0+\sum_{j=1}^{p_n}\sum_{k=1}^{\infty}\theta_{ijk}\eta_{jk})-b'(\alpha_0+\sum_{j=1}^{p_n}\sum_{k=1}^{s_n}\theta_{ijk}\eta_{jk})\}]^2,$$

$$\Delta_4^* = \max_{l\le h_n s_n} n^{-1}\sum_{i=1}^{n}[(\hat{w}_l-w_l)'F_i\{b'(\alpha_0+\sum_{j=1}^{p_n}\sum_{k=1}^{\infty}\theta_{ijk}\eta_{jk})-b'(\alpha_0+\sum_{j=1}^{p_n}\sum_{k=1}^{s_n}\theta_{ijk}\eta_{jk})\}]^2,$$

$$\Delta_5^* = \max_{l\le h_n s_n} n^{-1}\sum_{i=1}^{n}[(\tilde{E}_{il}-\breve{E}_{il})\{b'(\alpha_0+\sum_{j=1}^{p_n}\sum_{k=1}^{\infty}\theta_{ijk}\eta_{jk})-b'(\alpha_0+\sum_{j=1}^{p_n}\sum_{k=1}^{s_n}\theta_{ijk}\eta_{jk})\}]^2,$$

$$\Delta_6^* = \max_{l\le h_n s_n} n^{-1}\sum_{i=1}^{n}[(w_l'F_i-\tilde{E}_{il})\{b'(\hat{\alpha}_0+G_i'\hat{\eta})-b'(\alpha_0+G_i'\eta)\}]^2,$$

$$\Delta_7^* = \max_{l\le h_n s_n} n^{-1}\sum_{i=1}^{n}[(\hat{w}_l-w_l)'F_i\{b'(\hat{\alpha}_0+G_i'\hat{\eta})-b'(\alpha_0+G_i'\eta)\}]^2,$$

$$\Delta_8^* = \max_{l\le h_n s_n} n^{-1}\sum_{i=1}^{n}[(\tilde{E}_{il}-\breve{E}_{il})\{b'(\hat{\alpha}_0+G_i'\hat{\eta})-b'(\alpha_0+G_i'\eta)\}]^2.$$

To bound Δ_1^*, note that

$$\Delta_1^* \le (\max_{i\le n}|\epsilon_i|)^2\cdot \max_{l\le h_n s_n}(\hat{w}_l-w_l)'(n^{-1}\sum_{i=1}^{n}F_iF_i')(\hat{w}_l-w_l)$$

$$\le c_{16}n^{-1}\rho_n\log^2(n)\{\log(np_ns_n)\}\exp\{c_{16}q_n\log^{1/2}(nq_ns_n)\}, \tag{S2.71}$$

for some universal constant $c_{16} > 0$, with probability tending to 1, where the last inequality is based on Lemma 3 and (S2.56). To bound Δ_2^*, note that

$$\begin{aligned}\Delta_2^* \leq &\|\Lambda^{1/2}\hat{\Lambda}^{-1/2} - I\|_\infty^2 \cdot (\max_{i\leq n}|\epsilon_i|)^2 \cdot (\max_{l\leq h_n s_n + k_n} n^{-1}\sum_{i=1}^n \tilde{E}_{il}^2) \\ \leq & c_{17} n^{-1}\log^2(n)\{\log(np_n s_n)\}\exp\{c_{17} q_n \log^{1/2}(nq_n s_n)\},\end{aligned} \tag{S2.72}$$

for some universal constant $c_{17} > 0$, with probability tending to 1, where the last inequality is based on Lemma 2, Lemma 3, and (S2.39). To bound Δ_3^*, note that

$$\begin{aligned}\Delta_3^* =& \max_{l\leq h_n s_n} n^{-1}\sum_{i=1}^n [(w_l' F_i - \tilde{E}_{il})\{b'(\delta_i) - b'(\delta_i^*)\}]^2 \\ =& \max_{l\leq h_n s_n} n^{-1}\sum_{i=1}^n \{(w_l' F_i - \tilde{E}_{il}) b''(\delta_i^* + t_i \sum_{j=1}^{q_n}\sum_{k=s_n+1}^\infty \theta_{ijk}\eta_{jk}) \cdot (\sum_{j=1}^{q_n}\sum_{k=s_n+1}^\infty \theta_{ijk}\eta_{jk})\}^2 \\ \leq & \{\max_{i\leq n}|b''(\delta_i^* + t_i\sum_{j=1}^{q_n}\sum_{k=s_n+1}^\infty \theta_{ijk}\eta_{jk})|\}^2 \cdot (\max_{l\leq h_n s_n}\max_{i\leq n}|w_l' F_i - \tilde{E}_{il}|)^2 \cdot \\ & n^{-1}\sum_{i=1}^n(\sum_{j=1}^{q_n}\sum_{k=s_n+1}^\infty \theta_{ijk}\eta_{jk})^2 \\ \leq & c_{18} q_n^2 s_n^{-2\delta+1}\{\log(np_n s_n)\}\exp\{c_{18} q_n \log^{1/2}(nq_n s_n)\},\end{aligned} \tag{S2.73}$$

for some universal constant $c_{18} > 0$, with probability tending to 1, where the second equality is by (S2.35), and the last inequality is based on Lemma 2, Lemma 3, and (S2.36). To bound Δ_4^*, note that

$$\begin{aligned}
\Delta_4^* =& \max_{l \le h_n s_n} n^{-1} \sum_{i=1}^{n} [(\hat{w}_l - w_l)' F_i \{b'(\delta_i) - b'(\delta_i^*)\}]^2 \\
=& \max_{l \le h_n s_n} n^{-1} \sum_{i=1}^{n} \{(\hat{w}_l - w_l)' F_i b''(\delta_i^* + t_i \sum_{j=1}^{q_n} \sum_{k=s_n+1}^{\infty} \theta_{ijk}\eta_{jk}) \cdot (\sum_{j=1}^{q_n} \sum_{k=s_n+1}^{\infty} \theta_{ijk}\eta_{jk})\}^2 \\
\le& \{\max_{i \le n} |b''(\delta_i^* + t_i \sum_{j=1}^{q_n} \sum_{k=s_n+1}^{\infty} \theta_{ijk}\eta_{jk})|\}^2 \cdot \{n^{-1} \sum_{i=1}^{n} (\sum_{j=1}^{q_n} \sum_{k=s_n+1}^{\infty} \theta_{ijk}\eta_{jk})^2\} \cdot \\
& (\max_{i \le n} \|\breve{G}_i\|_\infty)^2 \cdot \max_{l \le h_n s_n} \|\hat{\Lambda}_{\mathcal{H}_n^c}^{1/2} (\hat{w}_l - w_l)\|_1^2 \\
\le& c_{19} n^{-1} \rho_n^2 q_n^2 s_n^{-2\delta+1} \{\log(n p_n s_n)\}^2 \exp\{c_{19} q_n \log^{1/2}(n q_n s_n)\}, \qquad \text{(S2.74)}
\end{aligned}$$

for some universal constant $c_{19} > 0$, with probability tending to 1, where the second equality is by (S2.35), and the last inequality is based on Lemma 2, Lemma 3, Lemma 6, and (S2.36). To bound Δ_5^*, note that

$$\begin{aligned}
\Delta_5^* =& \max_{l \le h_n s_n} n^{-1} \sum_{i=1}^{n} [(\tilde{E}_{il} - \breve{E}_{il}) \{b'(\delta_i) - b'(\delta_i^*)\}]^2 \\
=& \max_{l \le h_n s_n} n^{-1} \sum_{i=1}^{n} \{(\tilde{E}_{il} - \breve{E}_{il}) b''(\delta_i^* + t_i \sum_{j=1}^{q_n} \sum_{k=s_n+1}^{\infty} \theta_{ijk}\eta_{jk}) \cdot (\sum_{j=1}^{q_n} \sum_{k=s_n+1}^{\infty} \theta_{ijk}\eta_{jk})\}^2 \\
\le& \{\max_{i \le n} |b''(\delta_i^* + t_i \sum_{j=1}^{q_n} \sum_{k=s_n+1}^{\infty} \theta_{ijk}\eta_{jk})|\}^2 \cdot \{n^{-1} \sum_{i=1}^{n} (\sum_{j=1}^{q_n} \sum_{k=s_n+1}^{\infty} \theta_{ijk}\eta_{jk})^2\} \cdot \\
& (\max_{i \le n} \|\tilde{G}_i\|_\infty)^2 \cdot \|\Lambda^{1/2} \hat{\Lambda}^{-1/2} - I\|_\infty^2 \\
\le& c_{20} n^{-1} q_n^2 s_n^{-2\delta+1} \{\log(n p_n s_n)\}^2 \exp\{c_{20} q_n \log^{1/2}(n q_n s_n)\}, \qquad \text{(S2.75)}
\end{aligned}$$

for some universal constant $c_{20} > 0$, with probability tending to 1, where the second equality is by (S2.35), and the last inequality is based on Lemma 2, Lemma 3, and (S2.36). To bound Δ_6^*, first note that it follows from mean value theorem that for any $i \le n$

$$
\begin{aligned}
& b'(\hat{\alpha}_0 + G_i'\hat{\eta}) - b'(\alpha_0 + G_i'\eta) \\
=& b''(\delta_i^* + \bar{t}_i^*(\hat{\alpha}_0 - \alpha_0) + \bar{t}_i^* G_i'(\hat{\eta} - \eta)) \cdot \{(\hat{\alpha}_0 - \alpha_0) + G_i'(\hat{\eta} - \eta)\}, \qquad \text{(S2.76)}
\end{aligned}
$$

for some $\bar{t}_i^* \in [0, 1]$. In addition, we have

$$
\begin{aligned}
& \max_{i \le n} |b''(\delta_i^* + \bar{t}_i^*(\hat{\alpha}_0 - \alpha_0) + \bar{t}_i^* G_i'(\hat{\eta} - \eta))| \\
\le & \exp\{\max_{i \le n} |\delta_i^* + \bar{t}_i^*(\hat{\alpha}_0 - \alpha_0) + \bar{t}_i^* G_i'(\hat{\eta} - \eta)|\} \\
\le & \exp(\max_{i \le n} |\delta_i^*|) \cdot \exp\{|\hat{\alpha}_0 - \alpha_0| + \max_{i \le n} \max_{l \le p_n s_n} |\tilde{G}_{il}| \cdot \|\Lambda^{1/2}(\hat{\eta} - \eta)\|_1\} \\
\le & c_{21} \exp\{c_{21} q_n \log^{1/2}(n q_n s_n)\}, \qquad \text{(S2.77)}
\end{aligned}
$$

for some universal constant $c_{21} > 0$, with probability tending to 1, where the first inequality is by (A1), and the last inequality is based on Lemma 2, Lemma 3, and Theorem 1. To bound Δ_6^*, note that

$$
\begin{aligned}
\Delta_6^* =& \max_{l \le h_n s_n} n^{-1} \sum_{i=1}^n [(w_l' F_i - \tilde{E}_{il}) b''(\delta_i^* + \bar{t}_i^*(\hat{\alpha}_0 - \alpha_0) + \bar{t}_i^* G_i'(\hat{\eta} - \eta)) \cdot \{(\hat{\alpha}_0 - \alpha_0) + G_i'(\hat{\eta} - \eta)\}]^2 \\
\le & 2\{\max_{i \le n} |b''(\delta_i^* + \bar{t}_i^*(\hat{\alpha}_0 - \alpha_0) + \bar{t}_i^* G_i'(\hat{\eta} - \eta))|\}^2 \cdot (\max_{l \le h_n s_n} \max_{i \le n} |w_l' F_i - \tilde{E}_{il}|)^2 \cdot \\
& \{(\hat{\alpha}_0 - \alpha_0)^2 + (\hat{\eta} - \eta)'(n^{-1} \sum_{i=1}^n G_i G_i')(\hat{\eta} - \eta)\}, \qquad \text{(S2.78)}
\end{aligned}
$$

where the first equality is by (S2.76). Similar reasoning as (S2.65) leads to

$$
(\hat{\eta} - \eta)'(n^{-1} \sum_{i=1}^n G_i G_i')(\hat{\eta} - \eta) \le c_{22} \lambda_n^2 q_n n^{-1/9}, \qquad \text{(S2.79)}
$$

for some universal constant $c_{22} > 0$, with probability tending to 1. By combining (S2.77), (S2.79), Lemma 2, Theorem 1 with (S2.78), we have

$$\Delta_6^* \le c_{23} n^{-1/9} \lambda_n^2 q_n \{\log(np_n s_n)\} \exp\{c_{23} q_n \log^{1/2}(nq_n s_n)\}, \tag{S2.80}$$

for some universal constant $c_{23} > 0$, with probability tending to 1. To bound Δ_7^*, note that

$$\begin{aligned}
\Delta_7^* =& \max_{l \le h_n s_n} n^{-1} \sum_{i=1}^n [(\hat{w}_l - w_l)' F_i b''(\delta_i^* + \bar{t}_i^*(\hat{\alpha}_0 - \alpha_0) + \bar{t}_i^* G_i'(\hat{\eta} - \eta)) \cdot \{(\hat{\alpha}_0 - \alpha_0) + G_i'(\hat{\eta} - \eta)\}]^2 \\
\le& 2\{\max_{i \le n} |b''(\delta_i^* + \bar{t}_i^*(\hat{\alpha}_0 - \alpha_0) + \bar{t}_i^* G_i'(\hat{\eta} - \eta))|\}^2 \cdot \max_{l \le h_n s_n} \|\hat{\Lambda}_{\mathcal{H}_n^c}^{1/2}(\hat{w}_l - w_l)\|_1^2 \cdot (\max_{i \le n} \|\breve{G}_i\|_\infty)^2 \\
& \cdot \{(\hat{\alpha}_0 - \alpha_0)^2 + (\hat{\eta} - \eta)'(n^{-1} \sum_{i=1}^n G_i G_i')(\hat{\eta} - \eta)\} \\
\le& c_{24} n^{-10/9} \lambda_n^2 q_n \rho_n^2 \{\log(np_n s_n)\}^2 \exp\{c_{24} q_n \log^{1/2}(nq_n s_n)\},
\end{aligned} \tag{S2.81}$$

for some universal constant $c_{24} > 0$, with probability tending to 1, where the first equality is by (S2.76), and the last inequality is based on Lemma 2, Lemma 6, Theorem 1, (S2.79), and (S2.77). To bound Δ_8^*, note that

$$\begin{aligned}
\Delta_8^* =& \max_{l \le h_n s_n} n^{-1} \sum_{i=1}^n [(\tilde{E}_{il} - \breve{E}_{il}) b''(\delta_i^* + \bar{t}_i^*(\hat{\alpha}_0 - \alpha_0) + \bar{t}_i^* G_i'(\hat{\eta} - \eta)) \cdot \{(\hat{\alpha}_0 - \alpha_0) + G_i'(\hat{\eta} - \eta)\}]^2 \\
\le& 2\{\max_{i \le n} |b''(\delta_i^* + \bar{t}_i^*(\hat{\alpha}_0 - \alpha_0) + \bar{t}_i^* G_i'(\hat{\eta} - \eta))|\}^2 \cdot \|\Lambda^{1/2} \hat{\Lambda}^{-1/2} - I\|_\infty^2 \cdot (\max_{i \le n} \|\tilde{G}_i\|_\infty)^2 \cdot \\
& \{(\hat{\alpha}_0 - \alpha_0)^2 + (\hat{\eta} - \eta)'(n^{-1} \sum_{i=1}^n G_i G_i')(\hat{\eta} - \eta)\} \\
\le& c_{25} n^{-10/9} \lambda_n^2 q_n \{\log(np_n s_n)\}^2 \exp\{c_{25} q_n \log^{1/2}(nq_n s_n)\},
\end{aligned} \tag{S2.82}$$

for some universal constant $c_{25} > 0$, with probability tending to 1, where the first equality is by (S2.76), and the last inequality is based on Lemma 2, Theorem 1, (S2.79), and (S2.77). By combining (S2.82), (S2.81), (S2.80), (S2.75), (S2.74), (S2.73), (S2.72), (S2.71) with (S2.70), we have

$$\begin{aligned}
&\max_{l\leq h_n s_n} n^{-1}\sum_{i=1}^{n}(\hat{S}_{il}-S_{il}^{*})^2\\
\leq& c_{26}q_n^2 s_n^{-2\delta+1}\{\log(np_n s_n)\}\exp\{c_{26}q_n\log^{1/2}(nq_n s_n)\}+\\
&c_{26}\lambda_n^2 n^{-1/9}q_n\{\log(np_n s_n)\}\exp\{c_{26}q_n\log^{1/2}(nq_n s_n)\}+\\
&c_{26}n^{-1}\rho_n\log^2(n)\{\log(np_n s_n)\}\exp\{c_{26}q_n\log^{1/2}(nq_n s_n)\},
\end{aligned} \tag{S2.83}$$

for some universal constant $c_{26}>0$, with probability tending to 1. This completes the proof of part 2). To show part 3), first note that

$$\max_{l\leq h_n s_n} n^{-1}\sum_{i=1}^{n}\hat{S}_{il}^2 \lesssim \max_{l\leq h_n s_n} n^{-1}\sum_{i=1}^{n}(\hat{S}_{il}-S_{il}^{*})^2+\max_{l\leq h_n s_n} n^{-1}\sum_{i=1}^{n}S_{il}^{*2}. \tag{S2.84}$$

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

$$\begin{aligned}
&\max_{l\leq h_n s_n} n^{-1}\sum_{i=1}^{n}S_{il}^{*2}=\max_{l\leq h_n s_n} n^{-1}\sum_{i=1}^{n}(w_l'F_i-\tilde{E}_{il})^2\epsilon_i^2\\
\leq&(\max_{i\leq n}|\epsilon_i|)^2\cdot\max_{l\leq h_n s_n} n^{-1}\sum_{i=1}^{n}(w_l'F_i-\tilde{E}_{il})^2\\
\leq& c_{27}\log^2(n)\exp\{c_{27}q_n\log^{1/2}(nq_n s_n)\},
\end{aligned} \tag{S2.85}$$

for some universal constant $c_{27}>0$, with probability tending to 1, where the last inequality is based on Lemma 2, Lemma 3, (A2.1), and (A4.1). By combining (S2.85) and (S2.83) with (S2.84), the assertion in part 3) holds apparently. □

S3 Proofs of Main Theorems

Proof of Theorem 1. First note that with some abuse of notation, we write (η^*, α_0^*) as the true version of some estimator (η, α_0), and denotes the differences $\nu = \eta - \eta^*$ and $\tilde{\nu} = \Lambda^{1/2}\nu$. Based on the first order necessary condition of the optimization theory, any local minima $(\hat{\eta}, \hat{\alpha}_0)$ of $Q_n(\eta, \alpha_0)$ from (2.5) in the main article must satisfy $(\hat{\eta}, \hat{\alpha}_0) \in \{(\eta, \alpha_0) : \langle \nabla_\eta L_n(\eta, \alpha_0) + \nabla_\eta P_{\lambda_n}(\eta), \nu\rangle \leq 0, \;\; \nabla_{\alpha_0} L_n(\eta, \alpha_0) = 0, \;\; \|\eta\|_1 + |\alpha_0| \leq B_n\}$. Hence, to show Theorem 1, it suffices to justify that any estimator $(\eta, \alpha_0) \in \{(\eta, \alpha_0) : \langle \nabla_\eta L_n(\eta, \alpha_0) + \nabla_\eta P_{\lambda_n}(\eta), \nu\rangle \leq 0, \;\; \nabla_{\alpha_0} L_n(\eta, \alpha_0) = 0, \;\; \|\eta\|_1 + |\alpha_0| \leq B_n\}$ satisfies parts 1)–2) of Theorem 1. Therefore, we start the proof with an arbitrary estimator (η, α_0) satisfying

$$
\begin{aligned}
(\eta, \alpha_0) \in \{(\eta, \alpha_0) :& \langle \nabla_\eta L_n(\eta, \alpha_0) + \nabla_\eta P_{\lambda_n}(\eta), \nu\rangle \leq 0, \\
& \nabla_{\alpha_0} L_n(\eta, \alpha_0) = 0, \;\; \|\eta\|_1 + |\alpha_0| \leq B_n\}. \qquad \text{(S3.86)}
\end{aligned}
$$

In addition, it can be verified that

$$
\begin{aligned}
&\langle \nabla L_n(\eta, \alpha_0) - \nabla L_n(\eta^*, \alpha_0^*), (\nu', \alpha_0 - \alpha_0^*)'\rangle \\
=& n^{-1} \sum_{i=1}^{n} \{b'(\alpha_0 + \sum_{j=1}^{p_n}\sum_{k=1}^{s_n} \theta_{ijk}\eta_{jk}) - b'(\alpha_0^* + \sum_{j=1}^{p_n}\sum_{k=1}^{s_n} \theta_{ijk}\eta_{jk}^*)\} \cdot \{\tilde{G}_i'\tilde{\nu} + (\alpha_0 - \alpha_0^*)\} \\
=& n^{-1} \sum_{i=1}^{n} b''(\alpha_0^* + \sum_{j=1}^{p_n}\sum_{k=1}^{s_n} \theta_{ijk}\eta_{jk}^* + t_i^*\tilde{G}_i'\tilde{\nu} + t_i^*(\alpha_0 - \alpha_0^*)) \\
&\cdot \{\tilde{G}_i'\tilde{\nu} + (\alpha_0 - \alpha_0^*)\}^2, \qquad \text{(S3.87)}
\end{aligned}
$$

for some $t_i^* \in [0,1]$, where the last equality holds from mean value theorem. Similar reasoning as (the proof of) Corollary 2 in Loh and Wainwright (2015) yields that there exist constants $c_1, c_2 > 0$ such that with probability tending to 1:

$$\begin{aligned}
&\langle \nabla L_n(\eta, \alpha_0) - \nabla L_n(\eta^*, \alpha_0^*), (\nu', \alpha_0 - \alpha_0^*)' \rangle \\
\geq & c_1\{\|\tilde{\nu}\|_2^2 + (\alpha_0 - \alpha_0^*)^2\} - c_2\{\log(np_n s_n)/n\}^{1/2} \cdot (\|\tilde{\nu}\|_1 + |\alpha_0 - \alpha_0^*|) \\
& \cdot \{\|\tilde{\nu}\|_2^2 + (\alpha_0 - \alpha_0^*)^2\}^{1/2}, \qquad \forall \ \|\tilde{\nu}\|_2^2 + (\alpha_0 - \alpha_0^*)^2 \leq 1. \qquad \text{(S3.88)}
\end{aligned}$$

It follows from the arithmetic mean-geometric mean inequality that

$$\begin{aligned}
& c_2\{\log(np_n s_n)/n\}^{1/2} \cdot (\|\tilde{\nu}\|_1 + |\alpha_0 - \alpha_0^*|) \cdot \{\|\tilde{\nu}\|_2^2 + (\alpha_0 - \alpha_0^*)^2\}^{1/2} \\
\leq & 2^{-1} c_2^2 c_1^{-1} \{\log(np_n s_n)/n\}(\|\tilde{\nu}\|_1 + |\alpha_0 - \alpha_0^*|)^2 + 2^{-1} c_1 \{\|\tilde{\nu}\|_2^2 + (\alpha_0 - \alpha_0^*)^2\}.
\end{aligned}$$

Together with (S3.88) yields that with probability tending to 1:

$$\begin{aligned}
& \langle \nabla L_n(\eta, \alpha_0) - \nabla L_n(\eta^*, \alpha_0^*), (\nu', \alpha_0 - \alpha_0^*)' \rangle \\
\geq & 2^{-1} c_1 \{\|\tilde{\nu}\|_2^2 + (\alpha_0 - \alpha_0^*)^2\} - 2^{-1} c_2^2 c_1^{-1} \{\log(np_n s_n)/n\}(\|\tilde{\nu}\|_1 + |\alpha_0 - \alpha_0^*|)^2, \\
& \forall \ \|\tilde{\nu}\|_2^2 + (\alpha_0 - \alpha_0^*)^2 \leq 1. \qquad \text{(S3.89)}
\end{aligned}$$

Together with Lemma 8 in Loh and Wainwright (2015) yields that with probability tending to 1:

$$\begin{aligned}&\langle \nabla L_n(\eta,\alpha_0)-\nabla L_n(\eta^*,\alpha_0^*),(\nu',\alpha_0-\alpha_0^*)'\rangle\\ \geq& 2^{-1}c_1\{\|\tilde{\nu}\|_2^2+(\alpha_0-\alpha_0^*)^2\}^{1/2}-\{\log(np_ns_n)/n\}^{1/2}(\|\tilde{\nu}\|_1+|\alpha_0-\alpha_0^*|),\\ &\forall\ \|\tilde{\nu}\|_2^2+(\alpha_0-\alpha_0^*)^2\geq 1. \end{aligned} \tag{S3.90}$$

Denoting the function $P_{\lambda_n,\mu}(\eta)=P_{\lambda_n}(\eta)+2^{-1}\mu n^{-10/9}\sum_{j=1}^{p_n}\|\Theta_j\eta_j\|_2^2$, it then follows from condition (B5) that $P_{\lambda_n,\mu}(\eta)$ is convex in η, which entails that $P_{\lambda_n,\mu}(\eta^*)-P_{\lambda_n,\mu}(\eta)\geq-\langle\nabla P_{\lambda_n,\mu}(\eta),(\nu',\alpha_0-\alpha_0^*)'\rangle$. This further implies

$$\begin{aligned}&-\langle\nabla P_{\lambda_n}(\eta),(\nu',\alpha_0-\alpha_0^*)'\rangle\\ \leq& 2^{-1}\mu n^{-10/9}\sum_{j=1}^{p_n}\|\Theta_j(\eta_j-\eta_j^*)\|_2^2+P_{\lambda_n}(\eta^*)-P_{\lambda_n}(\eta). \end{aligned} \tag{S3.91}$$

Next, we start to show that $P\{\|\tilde{\nu}\|_2^2+(\alpha_0-\alpha_0^*)^2\leq 1\}\to 1$. It follows from (S3.90) and (S3.86) that conditional on the event $\{\|\tilde{\nu}\|_2^2+(\alpha_0-\alpha_0^*)^2>1\}$, we have

$$\begin{aligned}&\langle-\nabla P_{\lambda_n}(\eta)-\nabla L_n(\eta^*,\alpha_0^*),(\nu',\alpha_0-\alpha_0^*)'\rangle\geq 2^{-1}c_1\{\|\tilde{\nu}\|_2^2+(\alpha_0-\alpha_0^*)^2\}^{1/2}\\ &-\{\log(np_ns_n)/n\}^{1/2}(\|\tilde{\nu}\|_1+|\alpha_0-\alpha_0^*|), \end{aligned} \tag{S3.92}$$

with probability tending to 1. To bound the term $|\langle \nabla L_n(\eta^*, \alpha_0^*), (\nu', \alpha_0 - \alpha_0^*)' \rangle|$, note that

$$
\begin{aligned}
& |\langle \nabla L_n(\eta^*, \alpha_0^*), (\nu', \alpha_0 - \alpha_0^*)' \rangle| \\
=& |n^{-1} \sum_{i=1}^{n} \{Y_i - b'(\alpha_0^* + \sum_{j=1}^{p_n} \sum_{k=1}^{s_n} \theta_{ijk}\eta_{jk}^*)\}\{\tilde{G}_i'\tilde{\nu} + (\alpha_0 - \alpha_0^*)\}| \\
\leq& \|n^{-1} \sum_{i=1}^{n} \tilde{G}_i\{Y_i - b'(\alpha_0^* + \sum_{j=1}^{p_n} \sum_{k=1}^{s_n} \theta_{ijk}\eta_{jk}^*)\}\|_\infty \cdot \|\tilde{\nu}\|_1 + \\
& |n^{-1} \sum_{i=1}^{n} \{Y_i - b'(\alpha_0^* + \sum_{j=1}^{p_n} \sum_{k=1}^{s_n} \theta_{ijk}\eta_{jk}^*)\}| \cdot |\alpha_0 - \alpha_0^*| \\
\leq& c_3[q_n s_n^{-\delta+1/2} + \{\log^2(np_n s_n)/n\}^{1/2}] \exp\{c_3 q_n \log^{1/2}(nq_n s_n)\} \\
& \cdot (\|\tilde{\nu}\|_1 + |\alpha_0 - \alpha_0^*|), \qquad \text{(S3.93)}
\end{aligned}
$$

for some universal constants $c_3 > 0$, with probability tending to 1, where the last inequality holds from Lemma 3. To bound the term $|\langle \nabla P_{\lambda_n}(\eta), (\nu', \alpha_0 - \alpha_0^*)' \rangle|$, note that

$$
\begin{aligned}
&|\langle \nabla P_{\lambda_n}(\eta), (\nu', \alpha_0 - \alpha_0^*)' \rangle| \\
=&|\sum_{j=1}^{p_n} \rho'_{\lambda_n}\big(n^{-5/9}\|\Theta_j\eta_j\|_2\big) n^{-5/9}\|\Theta_j\eta_j\|_2^{-1}(\Theta_j\eta_j)'\Theta_j(\eta_j - \eta_j^*)| \\
\leq& n^{-5/9}\sum_{j=1}^{p_n} |\rho'_{\lambda_n}\big(n^{-5/9}\|\Theta_j\eta_j\|_2\big)| \cdot \|\tilde{\Theta}_j(\tilde{\eta}_j - \tilde{\eta}_j^*)\|_2 \\
\leq& c_4 n^{-5/9}\lambda_n \sum_{j=1}^{p_n} \|\tilde{\Theta}_j(\tilde{\eta}_j - \tilde{\eta}_j^*)\|_2 \\
=& c_4\lambda_n n^{-1/18}\sum_{j=1}^{p_n} \big[(\tilde{\eta}_j - \tilde{\eta}_j^*)' E(n^{-1}\tilde{\Theta}_j'\tilde{\Theta}_j)(\tilde{\eta}_j - \tilde{\eta}_j^*) + \\
&(\tilde{\eta}_j - \tilde{\eta}_j^*)'\{n^{-1}\tilde{\Theta}_j'\tilde{\Theta}_j - E(n^{-1}\tilde{\Theta}_j'\tilde{\Theta}_j)\}(\tilde{\eta}_j - \tilde{\eta}_j^*)\big]^{1/2} \\
\leq& c_5\lambda_n n^{-1/18}\{\|n^{-1}\sum_{i=1}^{n}\tilde{G}_i\tilde{G}_i' - E(\tilde{G}_i\tilde{G}_i')\|_\infty^{1/2} + \|E(\tilde{G}_i\tilde{G}_i')\|_\infty^{1/2}\}\|\tilde{\nu}\|_1 \\
\leq& c_6\lambda_n n^{-1/18}\|\tilde{\nu}\|_1, \qquad \text{(S3.94)}
\end{aligned}
$$

for some universal constants $c_4, c_5, c_6 > 0$, with probability tending to 1, where the second inequality follows from Lemma 1, and the last inequality is based on (A2.1) and Lemma 2. By combining (S3.93) and (S3.94) with (S3.92), it can be deduced that conditional on the event $\{\|\tilde{\nu}\|_2^2 + (\alpha_0 - \alpha_0^*)^2 > 1\}$, we have

$$
\begin{aligned}
&\{\|\tilde{\nu}\|_2^2 + (\alpha_0 - \alpha_0^*)^2\}^{1/2} \\
\leq &[c_7\lambda_n n^{-1/18} + c_7 q_n s_n^{-\delta+1/2}\exp\{c_7 q_n \log^{1/2}(nq_n s_n)\} + \\
&c_7\{\log^2(np_n s_n)/n\}^{1/2}\exp\{c_7 q_n \log^{1/2}(nq_n s_n)\}] \cdot (\|\tilde{\nu}\|_1 + |\alpha_0 - \alpha_0^*|) \\
\leq &[c_8\lambda_n n^{-1/18} + c_8 q_n s_n^{-\delta+1/2}\exp\{c_8 q_n \log^{1/2}(nq_n s_n)\} + \\
&c_8\{\log^2(np_n s_n)/n\}^{1/2}\exp\{c_8 q_n \log^{1/2}(nq_n s_n)\}] \cdot B_n \\
\leq &1,
\end{aligned}
$$

for some universal constants $c_7, c_8 > 0$, with probability tending to 1, where the second inequality is based on the fact that $\|\tilde{\nu}\|_1 + |\alpha_0 - \alpha_0^*| \lesssim \|\nu\|_1 + |\alpha_0 - \alpha_0^*| \lesssim B_n$, and the last inequality holds from (A4) and (A5). This further entails that

$$
P\{\|\tilde{\nu}\|_2^2 + (\alpha_0 - \alpha_0^*)^2 \leq 1\} \to 1. \tag{S3.95}
$$

Together with (S3.89) yields that with probability tending to 1:

$$
\begin{aligned}
&\langle \nabla L_n(\eta, \alpha_0) - \nabla L_n(\eta^*, \alpha_0^*), (\nu', \alpha_0 - \alpha_0^*)' \rangle \\
\geq &c_9\{\|\tilde{\nu}\|_2^2 + (\alpha_0 - \alpha_0^*)^2\} - c_{10}\{\log(np_n s_n)/n\}(\|\tilde{\nu}\|_1 + |\alpha_0 - \alpha_0^*|)^2,
\end{aligned}
$$

for some universal constants $c_9, c_{10} > 0$. Together with (S3.86) yields that with probability tending to 1:

$$
\begin{aligned}
&\langle -\nabla P_{\lambda_n}(\eta) - \nabla L_n(\eta^*, \alpha_0^*), (\nu', \alpha_0 - \alpha_0^*)' \rangle \\
\geq &c_9\{\|\tilde{\nu}\|_2^2 + (\alpha_0 - \alpha_0^*)^2\} - c_{10}\{\log(np_n s_n)/n\}(\|\tilde{\nu}\|_1 + |\alpha_0 - \alpha_0^*|)^2.
\end{aligned}
$$

Based on the above discussion, we have that with probability tending to 1:

$$
\begin{aligned}
& c_9\{\|\tilde{\nu}\|_2^2 + (\alpha_0 - \alpha_0^*)^2\} - c_{10}\{\log(np_n s_n)/n\}(\|\tilde{\nu}\|_1 + |\alpha_0 - \alpha_0^*|)^2 \\
\le & \langle -\nabla P_{\lambda_n}(\eta) - \nabla L_n(\eta^*, \alpha_0^*), (\nu', \alpha_0 - \alpha_0^*)' \rangle \\
\le & 2^{-1}\mu n^{-10/9} \sum_{j=1}^{p_n} \|\Theta_j(\eta_j - \eta_j^*)\|_2^2 + P_{\lambda_n}(\eta^*) - P_{\lambda_n}(\eta) + \langle -\nabla L_n(\eta^*, \alpha_0^*), (\nu', \alpha_0 - \alpha_0^*)' \rangle \\
\le & c_{11} n^{-1/9}\{\|\tilde{\nu}\|_2^2 + (\alpha_0 - \alpha_0^*)^2\} + P_{\lambda_n}(\eta^*) - P_{\lambda_n}(\eta) + \\
& c_{11}[q_n s_n^{-\delta+1/2} + \{\log^2(np_n s_n)/n\}^{1/2}] \exp\{c_{11} q_n \log^{1/2}(nq_n s_n)\} \\
& \cdot (\|\tilde{\nu}\|_1 + |\alpha_0 - \alpha_0^*|), \qquad \text{(S3.96)}
\end{aligned}
$$

for some universal constant $c_{11} > 0$, where the second inequality follows from (S3.91), and the last inequality holds from Lemma 4 and (S3.93). Some rearrangement of (S3.96) leads to

$$
\begin{aligned}
0 \le & (c_9 - c_{11} n^{-1/9})\{\|\tilde{\nu}\|_2^2 + (\alpha_0 - \alpha_0^*)^2\} \\
\le & c_{11}[q_n s_n^{-\delta+1/2} + \{\log^2(np_n s_n)/n\}^{1/2}] \exp\{c_{11} q_n \log^{1/2}(nq_n s_n)\}(\|\tilde{\nu}\|_1 + |\alpha_0 - \alpha_0^*|) \\
& + c_{10}\{\log(np_n s_n)/n\}(\|\tilde{\nu}\|_1 + |\alpha_0 - \alpha_0^*|)^2 + P_{\lambda_n}(\eta^*) - P_{\lambda_n}(\eta) \\
\le & c_{12}[q_n s_n^{-\delta+1/2} + \{\log^2(np_n s_n)/n\}^{1/2}] \exp\{c_{12} q_n \log^{1/2}(nq_n s_n)\}(\|\tilde{\nu}\|_1 + |\alpha_0 - \alpha_0^*|) \\
& + P_{\lambda_n}(\eta^*) - P_{\lambda_n}(\eta), \qquad \text{(S3.97)}
\end{aligned}
$$

for some universal constant $c_{12} > 0$, with probability tending to 1, where the last inequality is based on the fact that $\|\tilde{\nu}\|_1 + |\alpha_0 - \alpha_0^*| \lesssim B_n$. It then follows from (S3.97) that conditional on the event $\big\{|\alpha_0 - \alpha_0^*| > 2c_9^{-1}c_{12}[q_n s_n^{-\delta+1/2} +$

$\{\log^2(np_ns_n)/n\}^{1/2}]\exp\{c_{12}q_n\log^{1/2}(nq_ns_n)\}\big\}$, we have

$$
\begin{aligned}
0 \leq &(2^{-1}c_9 - c_{11}n^{-1/9})\{\|\tilde{\nu}\|_2^2 + (\alpha_0 - \alpha_0^*)^2\} \\
\leq & c_{12}[q_ns_n^{-\delta+1/2} + \{\log^2(np_ns_n)/n\}^{1/2}]\exp\{c_{12}q_n\log^{1/2}(nq_ns_n)\}\|\tilde{\nu}\|_1 \\
&+ P_{\lambda_n}(\eta^*) - P_{\lambda_n}(\eta),
\end{aligned}
$$

with probability tending to 1. Together with Lemma 4 and (A5.2), it can be deduced that conditional on the event $\big\{|\alpha_0-\alpha_0^*| > 2c_9^{-1}c_{12}[q_ns_n^{-\delta+1/2}+\{\log^2(np_ns_n)/n\}^{1/2}]$ $\exp\{c_{12}q_n\log^{1/2}(nq_ns_n)\}\big\}$, we have $0 \leq \|\tilde{\nu}\|_2^2+(\alpha_0-\alpha_0^*)^2 \lesssim P_{\lambda_n}(\eta^*)-P_{\lambda_n}(\eta)$, with probability tending to 1. Together with part 4) of Lemma 1 yields that conditional on the event $\big\{|\alpha_0 - \alpha_0^*| > 2c_9^{-1}c_{12}[q_ns_n^{-\delta+1/2} + \{\log^2(np_ns_n)/n\}^{1/2}]$ $\exp\{c_{12}q_n\log^{1/2}(nq_ns_n)\}\big\}$, we have

$$
\begin{aligned}
&\|\tilde{\nu}\|_2^2 + (\alpha_0 - \alpha_0^*)^2 \\
\lesssim & \lambda_n\Big\{\sum_{j\in\mathcal{A}_n} n^{-5/9}\|\Theta_j(\eta_j - \eta_j^*)\|_2 - \sum_{j\in\mathcal{A}_n^c} n^{-5/9}\|\Theta_j(\eta_j - \eta_j^*)\|_2\Big\},
\end{aligned}
\qquad \text{(S3.98)}
$$

with probability tending to 1. On one hand, (S3.98) implies that conditional on the event $\big\{|\alpha_0-\alpha_0^*| > 2c_9^{-1}c_{12}[q_ns_n^{-\delta+1/2}+\{\log^2(np_ns_n)/n\}^{1/2}]\exp\{c_{12}q_n\log^{1/2}(nq_ns_n)\}\big\}$, we have

$$\begin{aligned}
\|\tilde{\nu}\|_2^2 + (\alpha_0 - \alpha_0^*)^2 &\lesssim \lambda_n \sum_{j\in\mathcal{A}_n} n^{-5/9}\|\Theta_j(\eta_j - \eta_j^*)\|_2 \\
&\lesssim \lambda_n n^{-5/9} q_n^{1/2}\Big\{\sum_{j\in\mathcal{A}_n}\|\Theta_j(\eta_j - \eta_j^*)\|_2^2\Big\}^{1/2} \\
&\lesssim \lambda_n n^{-1/18} q_n^{1/2}\|\tilde{\nu}\|_2 \\
&\lesssim \lambda_n n^{-1/18} q_n^{1/2}\{\|\tilde{\nu}\|_2^2 + (\alpha_0 - \alpha_0^*)^2\}^{1/2},
\end{aligned}$$

with probability tending to 1, where the second last inequality holds from part 1) of Lemma 4. This further entails that conditional on the event $\big\{|\alpha_0 - \alpha_0^*| > 2c_9^{-1}c_{12}[q_n s_n^{-\delta+1/2} + \{\log^2(np_n s_n)/n\}^{1/2}]\exp\{c_{12}q_n\log^{1/2}(nq_n s_n)\}\big\}$, we have

$$\{\|\tilde{\nu}\|_2^2 + (\alpha_0 - \alpha_0^*)^2\}^{1/2} \lesssim \lambda_n n^{-1/18} q_n^{1/2}, \tag{S3.99}$$

with probability tending to 1. On the other hand, (S3.98) also implies that conditional on the event $\big\{|\alpha_0 - \alpha_0^*| > 2c_9^{-1}c_{12}[q_n s_n^{-\delta+1/2} + \{\log^2(np_n s_n)/n\}^{1/2}]$ $\exp\{c_{12}q_n\log^{1/2}(nq_n s_n)\}\big\}$, we have

$$\sum_{j\in\mathcal{A}_n^c}\|\Theta_j(\eta_j - \eta_j^*)\|_2 \le \sum_{j\in\mathcal{A}_n}\|\Theta_j(\eta_j - \eta_j^*)\|_2, \tag{S3.100}$$

with probability tending to 1. Therefore, conditional on the event $\big\{|\alpha_0 - \alpha_0^*| > 2c_9^{-1}c_{12}[q_n s_n^{-\delta+1/2} + \{\log^2(np_ns_n)/n\}^{1/2}]\exp\{c_{12}q_n\log^{1/2}(nq_ns_n)\}\big\}$, we have

$$\begin{aligned}\|\tilde{\nu}\|_1 &\lesssim n^{-1/2}s_n^{1/2}\sum_{j=1}^{p_n}\|\Theta_j(\eta_j-\eta_j^*)\|_2 \lesssim n^{-1/2}s_n^{1/2}\sum_{j\in\mathcal{A}_n}\|\Theta_j(\eta_j-\eta_j^*)\|_2\\ &\lesssim n^{-1/2}s_n^{1/2}q_n^{1/2}\{\sum_{j\in\mathcal{A}_n}\|\Theta_j(\eta_j-\eta_j^*)\|_2^2\}^{1/2}\\ &\lesssim s_n^{1/2}q_n^{1/2}\|\tilde{\nu}\|_2 \lesssim \lambda_n s_n^{1/2}q_n n^{-1/18}, \qquad \text{(S3.101)}\end{aligned}$$

with probability tending to 1, where the first inequality holds from part 2) of Lemma 4, the second inequality is based on (S3.100), the fourth inequality follows from part 1) of Lemma 4, and the last inequality is due to (S3.99). Moreover, it can be verified that

$$\begin{aligned}&\langle\nabla_\eta L_n(\eta,\alpha_0)-\nabla_\eta L_n(\eta^*,\alpha_0),\nu\rangle\\ =&n^{-1}\sum_{i=1}^{n}\{b'(\alpha_0+\sum_{j=1}^{p_n}\sum_{k=1}^{s_n}\theta_{ijk}\eta_{jk})-b'(\alpha_0+\sum_{j=1}^{p_n}\sum_{k=1}^{s_n}\theta_{ijk}\eta_{jk}^*)\}\cdot(\tilde{G}_i'\tilde{\nu})\\ =&n^{-1}\sum_{i=1}^{n}b''(\alpha_0+\sum_{j=1}^{p_n}\sum_{k=1}^{s_n}\theta_{ijk}\eta_{jk}^*+\tilde{t}_i\tilde{G}_i'\tilde{\nu})\cdot(\tilde{G}_i'\tilde{\nu})^2,\end{aligned}$$

for some $\tilde{t}_i \in [0,1]$, where the last equality is by mean value theorem. Similar reasoning as (the proof of) Corollary 2 in Loh and Wainwright (2015) yields that there exist constants $c_{13}, c_{14} > 0$ such that with probability tending to 1:

$$\begin{aligned}&\langle\nabla_\eta L_n(\eta,\alpha_0)-\nabla_\eta L_n(\eta^*,\alpha_0),\nu\rangle\\ \geq&c_{13}\|\tilde{\nu}\|_2^2 - c_{14}\{\log(np_ns_n)/n\}^{1/2}\|\tilde{\nu}\|_1\|\tilde{\nu}\|_2, \qquad \forall\ \|\tilde{\nu}\|_2\leq 1. \qquad \text{(S3.102)}\end{aligned}$$

It follows from the arithmetic mean-geometric mean inequality that

$$
\begin{aligned}
&c_{14}\{\log(np_n s_n)/n\}^{1/2}\|\tilde{\nu}\|_1\|\tilde{\nu}\|_2 \\
\le & 2^{-1}c_{14}^2 c_{13}^{-1}\{\log(np_n s_n)/n\}\|\tilde{\nu}\|_1^2 + 2^{-1}c_{13}\|\tilde{\nu}\|_2^2.
\end{aligned}
$$

Together with (S3.102), (S3.95), and (S3.86) yields that with probability tending to 1:

$$
\begin{aligned}
&\langle -\nabla_\eta P_{\lambda_n}(\eta) - \nabla_\eta L_n(\eta^*, \alpha_0), \nu\rangle \\
\ge & 2^{-1}c_{13}\|\tilde{\nu}\|_2^2 - 2^{-1}c_{14}^2 c_{13}^{-1}\{\log(np_n s_n)/n\}\|\tilde{\nu}\|_1^2. \qquad \text{(S3.103)}
\end{aligned}
$$

To bound the term $|\langle \nabla_\eta L_n(\eta^*, \alpha_0), \nu\rangle|$, note that

$$
\begin{aligned}
&|\langle \nabla_\eta L_n(\eta^*, \alpha_0), \nu\rangle| \\
\le & |n^{-1}\sum_{i=1}^n \{Y_i - b'(\alpha_0^* + \sum_{j=1}^{p_n}\sum_{k=1}^{s_n}\theta_{ijk}\eta_{jk}^*)\} \cdot (\tilde{G}_i'\tilde{\nu})| + \\
& |n^{-1}\sum_{i=1}^n \{b'(\alpha_0 + \sum_{j=1}^{p_n}\sum_{k=1}^{s_n}\theta_{ijk}\eta_{jk}^*) - b'(\alpha_0^* + \sum_{j=1}^{p_n}\sum_{k=1}^{s_n}\theta_{ijk}\eta_{jk}^*)\} \cdot (\tilde{G}_i'\tilde{\nu})| \\
= & |n^{-1}\sum_{i=1}^n \{Y_i - b'(\alpha_0^* + \sum_{j=1}^{p_n}\sum_{k=1}^{s_n}\theta_{ijk}\eta_{jk}^*)\} \cdot (\tilde{G}_i'\tilde{\nu})| + \\
& |n^{-1}\sum_{i=1}^n (\alpha_0 - \alpha_0^*) b''(\alpha_0^* + \sum_{j=1}^{p_n}\sum_{k=1}^{s_n}\theta_{ijk}\eta_{jk}^* + \tilde{t}_i^*(\alpha_0 - \alpha_0^*)) \cdot (\tilde{G}_i'\tilde{\nu})| \\
\le & \|n^{-1}\sum_{i=1}^n \tilde{G}_i\{Y_i - b'(\alpha_0^* + \sum_{j=1}^{p_n}\sum_{k=1}^{s_n}\theta_{ijk}\eta_{jk}^*)\}\|_\infty \cdot \|\tilde{\nu}\|_1 + \\
& \|n^{-1}\sum_{i=1}^n (\alpha_0 - \alpha_0^*)\tilde{G}_i b''(\alpha_0^* + \sum_{j=1}^{p_n}\sum_{k=1}^{s_n}\theta_{ijk}\eta_{jk}^* + \tilde{t}_i^*(\alpha_0 - \alpha_0^*))\|_\infty \cdot \|\tilde{\nu}\|_1,
\end{aligned}
$$

(S3.104)

for some $\tilde{t}_i^* \in [0,1]$, where the first equality holds from mean value theorem. To bound the term $\|n^{-1}\sum_{i=1}^n(\alpha_0-\alpha_0^*)\tilde{G}_i b''(\alpha_0^*+\sum_{j=1}^{p_n}\sum_{k=1}^{s_n}\theta_{ijk}\eta_{jk}^*+\tilde{t}_i^*(\alpha_0-\alpha_0^*))\|_\infty$, note that

$$\begin{aligned}
&\|n^{-1}\sum_{i=1}^n(\alpha_0-\alpha_0^*)\tilde{G}_i b''(\alpha_0^*+\sum_{j=1}^{p_n}\sum_{k=1}^{s_n}\theta_{ijk}\eta_{jk}^*+\tilde{t}_i^*(\alpha_0-\alpha_0^*))\|_\infty \\
\le&|\alpha_0-\alpha_0^*|\cdot\{\max_{l\le p_n s_n} n^{-1}\sum_{i=1}^n \tilde{G}_{il}^2\}^{1/2}\cdot\max_{i\le n}|b''(\alpha_0^*+\sum_{j=1}^{p_n}\sum_{k=1}^{s_n}\theta_{ijk}\eta_{jk}^*+\tilde{t}_i^*(\alpha_0-\alpha_0^*))| \\
\le&|\alpha_0-\alpha_0^*|\cdot\{\max_{l\le p_n s_n} n^{-1}\sum_{i=1}^n \tilde{G}_{il}^2\}^{1/2}\cdot\exp(\max_{i\le n}|\alpha_0^*+\sum_{j=1}^{p_n}\sum_{k=1}^{s_n}\theta_{ijk}\eta_{jk}^*+\tilde{t}_i^*(\alpha_0-\alpha_0^*)|) \\
\le&|\alpha_0-\alpha_0^*|\exp(|\alpha_0-\alpha_0^*|)\cdot\{\max_{l\le p_n s_n} n^{-1}\sum_{i=1}^n \tilde{G}_{il}^2\}^{1/2}\cdot\exp(\max_{i\le n}|\alpha_0^*+\sum_{j=1}^{p_n}\sum_{k=1}^{s_n}\theta_{ijk}\eta_{jk}^*|) \\
\le&c_{15}|\alpha_0-\alpha_0^*|\exp(|\alpha_0-\alpha_0^*|)\exp\{c_{15}q_n\log^{1/2}(nq_ns_n)\}, \qquad \text{(S3.105)}
\end{aligned}$$

for some universal constants $c_{15}>0$, with probability tending to 1, where the second inequality is based on (A1), and the last inequality follows from (S2.39) and Lemma 3. Based on (S3.105), it can be deduced that conditional on the event $\{|\alpha_0-\alpha_0^*|\le 2c_9^{-1}c_{12}[q_ns_n^{-\delta+1/2}+\{\log^2(np_ns_n)/n\}^{1/2}]\exp\{c_{12}q_n\log^{1/2}(nq_ns_n)\}\}$, we have

$$\begin{aligned}
&\|n^{-1}\sum_{i=1}^n(\alpha_0-\alpha_0^*)\tilde{G}_i b''(\alpha_0^*+\sum_{j=1}^{p_n}\sum_{k=1}^{s_n}\theta_{ijk}\eta_{jk}^*+\tilde{t}_i^*(\alpha_0-\alpha_0^*))\|_\infty \\
\le&c_{16}[q_ns_n^{-\delta+1/2}+\{\log^2(np_ns_n)/n\}^{1/2}]\exp\{c_{16}q_n\log^{1/2}(nq_ns_n)\}, \qquad \text{(S3.106)}
\end{aligned}$$

for some universal constants $c_{16}>0$, with probability tending to 1. By combining (S3.106) and Lemma 3 with (S3.104), it can be deduced that conditional on the event $\{|\alpha_0-\alpha_0^*| \leq 2c_9^{-1}c_{12}[q_n s_n^{-\delta+1/2}+\{\log^2(np_n s_n)/n\}^{1/2}]\exp\{c_{12}q_n\log^{1/2}(nq_n s_n)\}\}$, we have

$$\begin{aligned}&|\langle\nabla_\eta L_n(\eta^*,\alpha_0),\nu\rangle|\\ \leq& c_{17}[q_n s_n^{-\delta+1/2}+\{\log^2(np_n s_n)/n\}^{1/2}]\exp\{c_{17}q_n\log^{1/2}(nq_n s_n)\}\cdot\|\tilde{\nu}\|_1,\end{aligned} \tag{S3.107}$$

for some universal constants $c_{17}>0$, with probability tending to 1. By combining (S3.107) and (S3.91) with (S3.103), it can be deduced that conditional on the event $\{|\alpha_0-\alpha_0^*| \leq 2c_9^{-1}c_{12}[q_n s_n^{-\delta+1/2}+\{\log^2(np_n s_n)/n\}^{1/2}]\exp\{c_{12}q_n\log^{1/2}(nq_n s_n)\}\}$, we have

$$\begin{aligned}&2^{-1}c_{13}\|\tilde{\nu}\|_2^2-2^{-1}c_{14}^2c_{13}^{-1}\{\log(np_n s_n)/n\}\|\tilde{\nu}\|_1^2\\ \leq& c_{17}[q_n s_n^{-\delta+1/2}+\{\log^2(np_n s_n)/n\}^{1/2}]\exp\{c_{17}q_n\log^{1/2}(nq_n s_n)\}\cdot\|\tilde{\nu}\|_1+\\ &2^{-1}\mu n^{-10/9}\sum_{j=1}^{p_n}\|\Theta_j(\eta_j-\eta_j^*)\|_2^2+P_{\lambda_n}(\eta^*)-P_{\lambda_n}(\eta),\end{aligned}$$

with probability tending to 1. Together with Lemma 4, (A5.2) and the fact that $\|\tilde{\nu}\|_1 \lesssim B_n$, it can be deduced that conditional on the event $\{|\alpha_0-\alpha_0^*| \leq 2c_9^{-1}c_{12}[q_n s_n^{-\delta+1/2}+\{\log^2(np_n s_n)/n\}^{1/2}]\exp\{c_{12}q_n\log^{1/2}(nq_n s_n)\}\}$, we have $0\leq\|\tilde{\nu}\|_2^2\lesssim P_{\lambda_n}(\eta^*)-P_{\lambda_n}(\eta)$, with probability tending to 1. Together with part 4) of Lemma 1 yields that conditional on the event $\{|\alpha_0-\alpha_0^*| \leq$

$2c_9^{-1}c_{12}[q_n s_n^{-\delta+1/2} + \{\log^2(np_n s_n)/n\}^{1/2}]\exp\{c_{12}q_n\log^{1/2}(nq_n s_n)\}\big\}$, we have

$$\|\tilde{\nu}\|_2^2 \lesssim \lambda_n\big\{\sum_{j\in\mathcal{A}_n} n^{-5/9}\|\Theta_j(\eta_j-\eta_j^*)\|_2 - \sum_{j\in\mathcal{A}_n^c} n^{-5/9}\|\Theta_j(\eta_j-\eta_j^*)\|_2\big\},$$

(S3.108)

with probability tending to 1. On one hand, (S3.108) implies that conditional on the event $\big\{|\alpha_0-\alpha_0^*| \le 2c_9^{-1}c_{12}[q_n s_n^{-\delta+1/2}+\{\log^2(np_n s_n)/n\}^{1/2}]\exp\{c_{12}q_n\log^{1/2}(nq_n s_n)\}\big\}$, we have

$$\|\tilde{\nu}\|_2^2 \lesssim \lambda_n\sum_{j\in\mathcal{A}_n} n^{-5/9}\|\Theta_j(\eta_j-\eta_j^*)\|_2 \lesssim \lambda_n n^{-5/9}q_n^{1/2}\big\{\sum_{j\in\mathcal{A}_n}\|\Theta_j(\eta_j-\eta_j^*)\|_2^2\big\}^{1/2}$$
$$\lesssim \lambda_n n^{-1/18}q_n^{1/2}\|\tilde{\nu}\|_2,$$

with probability tending to 1, where the last inequality holds from part 1) of Lemma 4. This further entails that conditional on the event $\big\{|\alpha_0-\alpha_0^*| \le 2c_9^{-1}c_{12}[q_n s_n^{-\delta+1/2} + \{\log^2(np_n s_n)/n\}^{1/2}]\exp\{c_{12}q_n\log^{1/2}(nq_n s_n)\}\big\}$, we have

$$\|\tilde{\nu}\|_2 \lesssim \lambda_n n^{-1/18}q_n^{1/2}, \tag{S3.109}$$

with probability tending to 1. On the other hand, (S3.108) also implies that conditional on the event $\big\{|\alpha_0-\alpha_0^*| \le 2c_9^{-1}c_{12}[q_n s_n^{-\delta+1/2}+\{\log^2(np_n s_n)/n\}^{1/2}]$ $\exp\{c_{12}q_n\log^{1/2}(nq_n s_n)\}\big\}$, we have

$$\sum_{j\in\mathcal{A}_n^c}\|\Theta_j(\eta_j-\eta_j^*)\|_2 \le \sum_{j\in\mathcal{A}_n}\|\Theta_j(\eta_j-\eta_j^*)\|_2, \tag{S3.110}$$

with probability tending to 1. Therefore, conditional on the event $\big\{|\alpha_0 - \alpha_0^*| \leq 2c_9^{-1}c_{12}[q_n s_n^{-\delta+1/2} + \{\log^2(np_n s_n)/n\}^{1/2}]\exp\{c_{12}q_n\log^{1/2}(nq_n s_n)\}\big\}$, we have

$$\begin{aligned}\|\tilde{\nu}\|_1 &\lesssim n^{-1/2}s_n^{1/2}\sum_{j=1}^{p_n}\|\Theta_j(\eta_j - \eta_j^*)\|_2 \lesssim n^{-1/2}s_n^{1/2}\sum_{j\in\mathcal{A}_n}\|\Theta_j(\eta_j - \eta_j^*)\|_2 \\ &\lesssim n^{-1/2}s_n^{1/2}q_n^{1/2}\{\sum_{j\in\mathcal{A}_n}\|\Theta_j(\eta_j - \eta_j^*)\|_2^2\}^{1/2} \\ &\lesssim s_n^{1/2}q_n^{1/2}\|\tilde{\nu}\|_2 \lesssim \lambda_n s_n^{1/2}q_n n^{-1/18}, \end{aligned} \tag{S3.111}$$

with probability tending to 1, where the first inequality holds from part 2) of Lemma 4, the second inequality is based on (S3.110), the fourth inequality follows from part 1) of Lemma 4, and the last inequality is due to (S3.109). It follows from (A5.2) that conditional on the event $\big\{|\alpha_0 - \alpha_0^*| \leq 2c_9^{-1}c_{12}[q_n s_n^{-\delta+1/2} + \{\log^2(np_n s_n)/n\}^{1/2}]\exp\{c_{12}q_n\log^{1/2}(nq_n s_n)\}\big\}$, we have

$$|\alpha_0 - \alpha_0^*| = o(\lambda_n n^{-1/18}q_n^{1/2}). \tag{S3.112}$$

Based on (S3.112), (S3.111), (S3.109), (S3.101), and (S3.99), it can be deduced that

$$\begin{aligned}&P(\|\tilde{\nu}\|_2 \lesssim \lambda_n q_n^{1/2}n^{-1/18}) \to 1, \\ &P(|\alpha_0 - \alpha_0^*| \lesssim \lambda_n q_n^{1/2}n^{-1/18}) \to 1, \\ &P(\|\tilde{\nu}\|_1 \lesssim \lambda_n s_n^{1/2}q_n n^{-1/18}) \to 1,\end{aligned}$$

which completes the proof of Theorem 1. □

Proof of Theorem 2. Recall the four quantities

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

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

$$\min_{l\leq h_ns_n} n^{-1}\sum_{i=1}^{n} E(S_{il}^{*2}) = \min_{l\leq h_ns_n} n^{-1}\sum_{i=1}^{n} E\{(w_l'F_i - \tilde{E}_{il})^2\epsilon_i^2\} \geq c_1, \quad \text{(S3.113)}$$

for some universal constant $c_1 > 0$, where the last inequality is based on (A2.4). Next, we proceed the proof by discussing two cases as follows.

Case one: Assume we have

$$\epsilon_i|X_i \sim \text{sub-Gaussian}(\sigma^{*2}\{1 + \text{var}(\epsilon_i|X_i)\}), \quad \text{(S3.114)}$$

for all $i \leq n$ under (A2.3). It then follows that

$$S_{il}^*|X_i \sim \text{sub-Gaussian}(\sigma^{*2}(w_l'F_i - \tilde{E}_{il})^2\{1 + \text{var}(\epsilon_i|X_i)\}), \quad \text{(S3.115)}$$

for all $i \leq n$ and $l \leq h_ns_n$, where $S_{il}^* = (w_l'F_i - \tilde{E}_{il})\epsilon_i$. Based on Lemma 2 and Lemma 3, there exists a universal constant $c_2 > 0$ such that

$$P(\mathcal{D}_n) \to 1, \quad \text{(S3.116)}$$

where the event $\mathcal{D}_n = \big\{\{X_i\}_{i=1}^n : \max_{i\leq n}\max_{l\leq h_ns_n}\sigma^{*2}(w_l'F_i - \tilde{E}_{il})^2\{1 +$

var$(\epsilon_i|X_i)\} \leq c_2\{\log(np_n s_n)\}\exp\{c_2 q_n \log^{1/2}(nq_n s_n)\}\big\}$. Thus, we have

$$
\begin{aligned}
&P\Big[\cap_{i=1}^{n}\cap_{l=1}^{h_n s_n}\big\{S_{il}^{*} \sim \text{sub-Gaussian}(c_2\{\log(np_n s_n)\}\exp\{c_2 q_n \log^{1/2}(nq_n s_n)\})\big\}\Big] \\
\geq &\sum_{\{X_i\}_{i=1}^{n}\in\mathcal{D}_n} P\Big[\cap_{i=1}^{n}\cap_{l=1}^{h_n s_n}\big\{S_{il}^{*} \sim \text{sub-Gaussian}(c_2\{\log(np_n s_n)\} \\
&\exp\{c_2 q_n \log^{1/2}(nq_n s_n)\})\big\}\Big|\{X_i\}_{i=1}^{n}\Big] \cdot P(\{X_i\}_{i=1}^{n}) \\
= &\sum_{\{X_i\}_{i=1}^{n}\in\mathcal{D}_n} P\Big[\cap_{i=1}^{n}\cap_{l=1}^{h_n s_n}\big\{S_{il}^{*}|X_i \sim \text{sub-Gaussian}(c_2\{\log(np_n s_n)\} \\
&\exp\{c_2 q_n \log^{1/2}(nq_n s_n)\})\big\}\Big|\{X_i\}_{i=1}^{n}\Big] \cdot P(\{X_i\}_{i=1}^{n}) \\
= &\sum_{\{X_i\}_{i=1}^{n}\in\mathcal{D}_n} P(\{X_i\}_{i=1}^{n}) = P(\mathcal{D}_n),
\end{aligned}
$$

where the second equality holds from (S3.115) and the definition of $\mathcal{D}_n$. Together with (S3.116) yields that

$$P\Big[\cap_{i=1}^{n}\cap_{l=1}^{h_n s_n}\big\{S_{il}^{*} \sim \text{sub-Gaussian}(v_n)\big\}\Big] \to 1, \tag{S3.117}$$

where $v_n = c_2\{\log(np_n s_n)\}\exp\{c_2 q_n \log^{1/2}(nq_n s_n)\}$. Based on (A4.1), it can be verified that

$$v_n^3\{\log(np_n s_n)\}^9/n \to 0. \tag{S3.118}$$

By combining (S3.113), (S3.117), (S3.118) with part 2) of Lemma 5, we have

$$\lim_{n\to\infty}\sup_{A\in\mathcal{A}^{Re}}\Big|P(T^{*}\in A) - P_e(T_e^{*}\in A)\Big| = 0, \tag{S3.119}$$

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

such that

$$P(\|\hat{T}(\beta_{\mathcal{H}_n}) - T^*\|_\infty \geq f_n) \to 0, \tag{S3.120}$$

where

$$\begin{aligned} f_n =& c_3 n^{-1/2} \rho_n \{\log(np_n s_n)\}^{3/2} \exp\{c_3 q_n \log^{1/2}(nq_n s_n)\} + \\ & c_3 n^{1/2} q_n s_n^{-\delta+1/2} \exp\{c_3 q_n \log^{1/2}(nq_n s_n)\} + \\ & c_3 \lambda_n n^{-1/18} s_n^{1/2} q_n \{\log(np_n s_n)\}^{1/2} \exp\{c_3 q_n \log^{1/2}(nq_n s_n)\} + \\ & c_3 \lambda_n^2 n^{7/18} q_n \{\log(np_n s_n)\}^{1/2} \exp\{c_3 q_n \log^{1/2}(nq_n s_n)\} + \\ & c_3 \lambda_n n^{-1/18} \rho_n^{1/2} q_n^{1/2} \{\log(np_n s_n)\}^{1/2} \exp\{c_3 q_n \log^{1/2}(nq_n s_n)\}. \end{aligned}$$

To bound $\|\hat{T}_e - T_e^*\|_\infty$, note that for any $t > 0$,

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

where the first inequality is by union bound inequality, and the second inequality is based on Hoeffding inequality. Plugging $t = \{2 \max_{l \leq h_n s_n} n^{-1} \sum_{i=1}^n (\hat{S}_{il} -$

$S_{il}^*)^2\}^{1/2}\log^{1/2}(np_ns_n)$ into the above inequality yields

$$P_e(\|\hat{T}_e - T_e^*\|_\infty \geq \{2\max_{l\leq h_ns_n} n^{-1}\sum_{i=1}^n(\hat{S}_{il} - S_{il}^*)^2\}^{1/2}\log^{1/2}(np_ns_n))$$
$$\leq 2n^{-1} \to 0.$$

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

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

where

$$\begin{aligned} g_n =& c_4q_ns_n^{-\delta+1/2}\{\log(np_ns_n)\}\exp\{c_4q_n\log^{1/2}(nq_ns_n)\}+ \\ & c_4\lambda_nn^{-1/18}q_n^{1/2}\{\log(np_ns_n)\}\exp\{c_4q_n\log^{1/2}(nq_ns_n)\}+ \\ & c_4n^{-1/2}\rho_n^{1/2}\log(n)\{\log(np_ns_n)\}\exp\{c_4q_n\log^{1/2}(nq_ns_n)\}. \end{aligned}$$

Together with (S3.120), there exists a universal constant $c_5 > 0$ such that

$$P(\|\hat{T}(\beta_{\mathcal{H}_n}) - T^*\|_\infty \geq a_n) \to 0,$$
$$P_e(\|\hat{T}_e - T_e^*\|_\infty \geq a_n) \xrightarrow{p} 0, \quad \text{(S3.121)}$$

where

$$
\begin{aligned}
a_n = & c_5 n^{-1/2} \rho_n \{\log(np_n s_n)\}^{3/2} \exp\{c_5 q_n \log^{1/2}(nq_n s_n)\} + \\
& c_5 n^{-1/2} \rho_n^{1/2} \log(n) \{\log(np_n s_n)\} \exp\{c_5 q_n \log^{1/2}(nq_n s_n)\} + \\
& c_5 n^{1/2} q_n s_n^{-\delta+1/2} \exp\{c_5 q_n \log^{1/2}(nq_n s_n)\} + \\
& c_5 \lambda_n n^{-1/18} s_n^{1/2} q_n \{\log(np_n s_n)\}^{1/2} \exp\{c_5 q_n \log^{1/2}(nq_n s_n)\} + \\
& c_5 \lambda_n n^{-1/18} \rho_n^{1/2} q_n^{1/2} \{\log(np_n s_n)\}^{1/2} \exp\{c_5 q_n \log^{1/2}(nq_n s_n)\} + \\
& c_5 \lambda_n n^{-1/18} q_n^{1/2} \{\log(np_n s_n)\} \exp\{c_5 q_n \log^{1/2}(nq_n s_n)\} + \\
& c_5 \lambda_n^2 n^{7/18} q_n \{\log(np_n s_n)\}^{1/2} \exp\{c_5 q_n \log^{1/2}(nq_n s_n)\}.
\end{aligned}
$$

Under (A4.3), (A5.1) and (A5.4), we have

$$
a_n^2 \{1 + \log(h_n s_n) - \log a_n\} \to 0. \tag{S3.122}
$$

By combining (S3.113), (S3.117), (S3.118), (S3.119), (S3.121), (S3.122) with part 2) of Lemma 5, it can be concluded that

$$
\begin{aligned}
& \lim_{n\to\infty} \sup_{A \in \mathcal{A}^{Re}} \left| P(\hat{T}(\beta_{\mathcal{H}_n}) \in A) - P_e(\hat{T}_e \in A) \right| \\
= & \lim_{n\to\infty} \sup_{t \geq 0} \left| P(\|\hat{T}(\beta_{\mathcal{H}_n})\|_\infty \leq t) - P_e(\|\hat{T}_e\|_\infty \leq t) \right| = 0,
\end{aligned} \tag{S3.123}
$$

under case one that is specified by (S3.114).

Case two: Assume we have

$$
\epsilon_i | X_i \sim \text{sub-Exponential}(\sigma^{*2}\{1 + \text{var}(\epsilon_i | X_i)\}), \tag{S3.124}
$$

for all $i \le n$ under (A2.3). It then follows that

$$S_{il}^*|X_i \sim \text{sub-Exponential}(\sigma^{*2}|w_l'F_i - \tilde{E}_{il}|\{1 + \text{var}(\epsilon_i|X_i)\}), \qquad \text{(S3.125)}$$

for all $i \le n$ and $l \le h_n s_n$, where $S_{il}^* = (w_l'F_i - \tilde{E}_{il})\epsilon_i$. Based on Lemma 2 and Lemma 3, there exists a universal constant $c_6 > 0$ such that

$$P(\mathcal{D}_n^*) \to 1, \qquad \text{(S3.126)}$$

where the event $\mathcal{D}_n^* = \big\{\{X_i\}_{i=1}^n : \max_{i\le n}\max_{l\le h_n s_n}\sigma^{*2}|w_l'F_i - \tilde{E}_{il}|\{1 + \text{var}(\epsilon_i|X_i)\} \le c_6\{\log(np_ns_n)\}^{1/2}\exp\{c_6 q_n \log^{1/2}(nq_ns_n)\}\big\}$. Thus, we have

$$\begin{aligned}
&P\big[\cap_{i=1}^n \cap_{l=1}^{h_ns_n}\big\{S_{il}^* \sim \text{sub-Exponential}(c_6\{\log(np_ns_n)\}^{1/2}\exp\{c_6q_n\log^{1/2}(nq_ns_n)\})\big\}\big]\\
&\ge \sum_{\{X_i\}_{i=1}^n\in\mathcal{D}_n^*} P\big[\cap_{i=1}^n \cap_{l=1}^{h_ns_n}\big\{S_{il}^* \sim \text{sub-Exponential}(c_6\{\log(np_ns_n)\}^{1/2}\\
&\exp\{c_6q_n\log^{1/2}(nq_ns_n)\})\big\}\big|\{X_i\}_{i=1}^n\big]\cdot P(\{X_i\}_{i=1}^n)\\
&= \sum_{\{X_i\}_{i=1}^n\in\mathcal{D}_n^*} P\big[\cap_{i=1}^n \cap_{l=1}^{h_ns_n}\big\{S_{il}^*|X_i \sim \text{sub-Exponential}(c_6\{\log(np_ns_n)\}^{1/2}\\
&\exp\{c_6q_n\log^{1/2}(nq_ns_n)\})\big\}\big|\{X_i\}_{i=1}^n\big]\cdot P(\{X_i\}_{i=1}^n)\\
&= \sum_{\{X_i\}_{i=1}^n\in\mathcal{D}_n^*} P(\{X_i\}_{i=1}^n) = P(\mathcal{D}_n^*),
\end{aligned}$$

where the second equality holds from (S3.125) and the definition of $\mathcal{D}_n^*$. Together with (S3.126) yields that

$$P\big[\cap_{i=1}^n \cap_{l=1}^{h_ns_n}\big\{S_{il}^* \sim \text{sub-Exponential}(u_n)\big\}\big] \to 1, \qquad \text{(S3.127)}$$

where $u_n = c_6\{\log(np_ns_n)\}^{1/2}\exp\{c_6q_n\log^{1/2}(nq_ns_n)\}$. Based on (A4.1), it can be verified that

$$u_n^6\{\log(np_ns_n)\}^9/n \to 0. \tag{S3.128}$$

By combining (S3.113), (S3.127), (S3.128), (S3.121), (S3.122) with part 1) of Lemma 5, it can be concluded that

$$\begin{aligned}&\lim_{n\to\infty}\sup_{A\in\mathcal{A}^{Re}}\left|P(\hat{T}(\beta_{\mathcal{H}_n})\in A)-P_e(\hat{T}_e\in A)\right|\\ =&\lim_{n\to\infty}\sup_{t\geq 0}\left|P(\|\hat{T}(\beta_{\mathcal{H}_n})\|_\infty\leq t)-P_e(\|\hat{T}_e\|_\infty\leq t)\right|=0,\end{aligned} \tag{S3.129}$$

under case two that is specified by (S3.124).

Finally, the assertion in Theorem 2 holds from (S3.123) and (S3.129). □

Proof of Theorem 3. Given the true $\beta_{\mathcal{H}_n}$, we have

$$\begin{aligned}&\lim_{n\to\infty}\left|Power(\beta_{\mathcal{H}_n})-Power^*(\beta_{\mathcal{H}_n})\right|\\ =&\lim_{n\to\infty}\left|P[\|\hat{T}(\beta_{\mathcal{H}_n})+n^{-1/2}\sum_{i=1}^n(\hat{w}'F_i-\breve{E}_i)\{b'(\hat{\alpha}_0+E_i'\eta_{\mathcal{H}_n}+F_i'\hat{\eta}_{\mathcal{H}_n^c})-\right.\\ &b'(\hat{\alpha}_0+F_i'\hat{\eta}_{\mathcal{H}_n^c})\}\|_\infty\leq c_B(\alpha)]-P_{e^*}[\|\hat{T}_{e^*}+n^{-1/2}\sum_{i=1}^n(\hat{w}'F_i-\breve{E}_i)\cdot\\ &\left.\{b'(\hat{\alpha}_0+E_i'\eta_{\mathcal{H}_n}+F_i'\hat{\eta}_{\mathcal{H}_n^c})-b'(\hat{\alpha}_0+F_i'\hat{\eta}_{\mathcal{H}_n^c})\}\|_\infty\leq c_B(\alpha)]\right|\\ \leq&\lim_{n\to\infty}\sup_{A\in\mathcal{A}^{Re}}\left|P(\hat{T}(\beta_{\mathcal{H}_n})\in A)-P_{e^*}(\hat{T}_{e^*}\in A)\right|=0,\end{aligned}$$

where the last equality is by (S3.123) and (S3.129). This completes the proof.

□

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

$$\begin{aligned}
&Power^*(\beta_{\mathcal{H}_n})\\
=&P_{e^*}[\|\hat{T}_{e^*} + n^{-1/2}\sum_{i=1}^{n}(\hat{w}'F_i - \breve{E}_i)\{b'(\hat{\alpha}_0 + E_i'\eta_{\mathcal{H}_n} + F_i'\hat{\eta}_{\mathcal{H}_n^c})\\
&- b'(\hat{\alpha}_0 + F_i'\hat{\eta}_{\mathcal{H}_n^c})\}\|_\infty > c_B(\alpha)]\\
\geq&1 - P_{e^*}[\|\hat{T}_{e^*}\|_\infty \geq \|n^{-1/2}\sum_{i=1}^{n}(\hat{w}'F_i - \breve{E}_i)\{b'(\hat{\alpha}_0 + E_i'\eta_{\mathcal{H}_n} + F_i'\hat{\eta}_{\mathcal{H}_n^c})\\
&- b'(\hat{\alpha}_0 + F_i'\hat{\eta}_{\mathcal{H}_n^c})\}\|_\infty - c_B(\alpha)]. \qquad \text{(S3.130)}
\end{aligned}$$

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

$$\begin{aligned}
&P_{e^*}(\|\hat{T}_{e^*}\|_\infty \geq t) = P_{e^*}(\max_{l\leq h_n s_n}|n^{-1/2}\sum_{i=1}^{n}e_i^*\hat{S}_{il}| \geq t)\\
\leq&\sum_{l=1}^{h_n s_n}P_{e^*}(|n^{-1/2}\sum_{i=1}^{n}e_i^*\hat{S}_{il}| \geq t) \leq 2\sum_{l=1}^{h_n s_n}\exp[-\{2n^{-1}\sum_{i=1}^{n}\hat{S}_{il}^2\}^{-1}t^2]\\
\leq&2p_n s_n\exp[-\{2\max_{l\leq h_n s_n}n^{-1}\sum_{i=1}^{n}\hat{S}_{il}^2\}^{-1}t^2], \qquad \text{(S3.131)}
\end{aligned}$$

where the second inequality is by Hoeffding inequality. Plugging $t = c_B(\alpha)$ into (S3.131) yields

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

Together with Lemma 7, there exists a universal constant $c_1 > 0$ such that with probability tending to 1:

$$c_B(\alpha) \leq c_1\log(n)\{\log(p_n s_n)\}^{1/2}\exp\{c_1 q_n\log^{1/2}(nq_n s_n)\}. \qquad \text{(S3.132)}$$

To bound the term $\|n^{-1/2}\sum_{i=1}^{n}(\hat{w}'F_i - \breve{E}_i)\{b'(\hat{\alpha}_0 + E_i'\eta_{\mathcal{H}_n} + F_i'\hat{\eta}_{\mathcal{H}_n^c}) - b'(\hat{\alpha}_0 + F_i'\hat{\eta}_{\mathcal{H}_n^c})\}\|_\infty$, note that

$$\|n^{-1/2}\sum_{i=1}^{n}(\hat{w}'F_i - \breve{E}_i)\{b'(\hat{\alpha}_0 + E_i'\eta_{\mathcal{H}_n} + F_i'\hat{\eta}_{\mathcal{H}_n^c}) - b'(\hat{\alpha}_0 + F_i'\hat{\eta}_{\mathcal{H}_n^c})\}\|_\infty$$
$$\geq \Pi_1 - \Pi_2 - \Pi_3, \tag{S3.133}$$

where

$$\Pi_1 = \|n^{-1/2}\sum_{i=1}^{n}(w'F_i - \tilde{E}_i)\{b'(\hat{\alpha}_0 + E_i'\eta_{\mathcal{H}_n} + F_i'\hat{\eta}_{\mathcal{H}_n^c}) - b'(\hat{\alpha}_0 + F_i'\hat{\eta}_{\mathcal{H}_n^c})\}\|_\infty,$$
$$\Pi_2 = \|n^{-1/2}\sum_{i=1}^{n}(\hat{w} - w)'F_i\{b'(\hat{\alpha}_0 + E_i'\eta_{\mathcal{H}_n} + F_i'\hat{\eta}_{\mathcal{H}_n^c}) - b'(\hat{\alpha}_0 + F_i'\hat{\eta}_{\mathcal{H}_n^c})\}\|_\infty,$$
$$\Pi_3 = \|n^{-1/2}\sum_{i=1}^{n}(\tilde{E}_i - \breve{E}_i)\{b'(\hat{\alpha}_0 + E_i'\eta_{\mathcal{H}_n} + F_i'\hat{\eta}_{\mathcal{H}_n^c}) - b'(\hat{\alpha}_0 + F_i'\hat{\eta}_{\mathcal{H}_n^c})\}\|_\infty.$$

For Π_1, it follows from the definition of $\mathcal{F}_n$ that with probability tending to 1,

$$\Pi_1 \geq K(\rho_n^{1/2} + \log n)\{\log(np_ns_n)\}^{1/2}\exp\{Kq_n\log^{1/2}(nq_ns_n)\}. \tag{S3.134}$$

Before bounding Π_2, first note that it follows from mean value theorem that for any $i \leq n$

$$b'(\hat{\alpha}_0 + E_i'\eta_{\mathcal{H}_n} + F_i'\hat{\eta}_{\mathcal{H}_n^c}) - b'(\hat{\alpha}_0 + F_i'\hat{\eta}_{\mathcal{H}_n^c})$$
$$= b''(\hat{\alpha}_0 + F_i'\hat{\eta}_{\mathcal{H}_n^c} + \check{t}_iE_i'\eta_{\mathcal{H}_n}) \cdot (E_i'\eta_{\mathcal{H}_n}) \tag{S3.135}$$

for some $\check{t}_i \in [0,1]$. Similar reasoning as (S2.21) leads to

$$\max_{i\leq n}|\hat{\alpha}_0 + F_i'\hat{\eta}_{\mathcal{H}_n^c} + \check{t}_iE_i'\eta_{\mathcal{H}_n}| + \max_{i\leq n}|E_i'\eta_{\mathcal{H}_n}| \leq c_2q_n\log^{1/2}(nq_ns_n), \tag{S3.136}$$

for some universal constant $c_2 > 0$, with probability tending to 1. Together with (A1) yields that with probability tending to 1,

$$\max_{i \le n} |b''(\hat{\alpha}_0 + F_i'\hat{\eta}_{\mathcal{H}_n^c} + \check{t}_i E_i'\eta_{\mathcal{H}_n})| \le \exp\{c_2 q_n \log^{1/2}(nq_n s_n)\}. \qquad \text{(S3.137)}$$

For Π_2, we have

$$\begin{aligned}
\Pi_2 = & \max_{l \le h_n s_n} |n^{-1/2} \sum_{i=1}^n (\hat{w}_l - w_l)' F_i b''(\hat{\alpha}_0 + F_i'\hat{\eta}_{\mathcal{H}_n^c} + \check{t}_i E_i'\eta_{\mathcal{H}_n}) \cdot (E_i'\eta_{\mathcal{H}_n})| \\
\le & n^{1/2} \{\max_{i \le n} |b''(\hat{\alpha}_0 + F_i'\hat{\eta}_{\mathcal{H}_n^c} + \check{t}_i E_i'\eta_{\mathcal{H}_n})|\} \cdot \{\max_{i \le n} |E_i'\eta_{\mathcal{H}_n}|\} \cdot \\
& \{\max_{l \le h_n s_n} (\hat{w}_l - w_l)'(n^{-1}\sum_{i=1}^n F_i F_i')(\hat{w}_l - w_l)\}^{1/2} \\
\le & c_3 \rho_n^{1/2} \{\log(np_n s_n)\}^{1/2} \exp\{c_3 q_n \log^{1/2}(nq_n s_n)\}, \qquad \text{(S3.138)}
\end{aligned}$$

for some universal constant $c_3 > 0$, with probability tending to 1, where the first equality is by (S3.135), and the last inequality is based on (S3.136), (S3.137), and (S2.56). For Π_3, we have

$$\begin{aligned}
\Pi_3 = & \max_{l \le h_n s_n} |n^{-1/2} \sum_{i=1}^n (\tilde{E}_{il} - \breve{E}_{il}) b''(\hat{\alpha}_0 + F_i'\hat{\eta}_{\mathcal{H}_n^c} + \check{t}_i E_i'\eta_{\mathcal{H}_n}) \cdot (E_i'\eta_{\mathcal{H}_n})| \\
\le & n^{1/2} \|\Lambda^{1/2}\hat{\Lambda}^{-1/2} - I\|_\infty \cdot \{\max_{i \le n} |b''(\hat{\alpha}_0 + F_i'\hat{\eta}_{\mathcal{H}_n^c} + \check{t}_i E_i'\eta_{\mathcal{H}_n})|\} \cdot \\
& \{\max_{i \le n} |E_i'\eta_{\mathcal{H}_n}|\} \cdot \{\max_{l \le p_n s_n} n^{-1} \sum_{i=1}^n \tilde{G}_{il}^2\}^{1/2} \\
\le & c_4 \{\log(np_n s_n)\}^{1/2} \exp\{c_4 q_n \log^{1/2}(nq_n s_n)\}, \qquad \text{(S3.139)}
\end{aligned}$$

for some universal constant $c_4 > 0$, with probability tending to 1, where the first equality is by (S3.135), and the last inequality is based on (S3.136), (S3.137), (S2.39), and Lemma 2. By choosing $K \geq 4(c_1 + c_3 + c_4)$ in $\mathcal{F}_n$, it follows from (S3.139), (S3.138), (S3.134), (S3.133), and (S3.132) that with probability tending to 1:

$$\|n^{-1/2}\sum_{i=1}^{n}(\hat{w}'F_i - \breve{E}_i)\{b'(\hat{\alpha}_0 + E_i'\eta_{\mathcal{H}_n} + F_i'\hat{\eta}_{\mathcal{H}_n^c}) - b'(\hat{\alpha}_0 + F_i'\hat{\eta}_{\mathcal{H}_n^c})\}\|_\infty - c_B(\alpha)$$

$$\geq 4^{-1}K(\rho_n^{1/2} + \log n)\{\log(np_ns_n)\}^{1/2}\exp\{Kq_n\log^{1/2}(nq_ns_n)\}. \quad \text{(S3.140)}$$

Plugging $t = \|n^{-1/2}\sum_{i=1}^{n}(\hat{w}'F_i - \breve{E}_i)\{b'(\hat{\alpha}_0 + E_i'\eta_{\mathcal{H}_n} + F_i'\hat{\eta}_{\mathcal{H}_n^c}) - b'(\hat{\alpha}_0 + F_i'\hat{\eta}_{\mathcal{H}_n^c})\}\|_\infty - c_B(\alpha)$ into (S3.131) yields that with probability tending to 1:

$$P_{e^*}\{\|\hat{T}_{e^*}\|_\infty \geq \|n^{-1/2}\sum_{i=1}^{n}(\hat{w}'F_i - \breve{E}_i)\{b'(\hat{\alpha}_0 + E_i'\eta_{\mathcal{H}_n} + F_i'\hat{\eta}_{\mathcal{H}_n^c}) - b'(\hat{\alpha}_0 + F_i'\hat{\eta}_{\mathcal{H}_n^c})\}\|_\infty - c_B(\alpha)\}$$

$$\leq 2p_ns_n\exp\Big(-\{2\max_{l\leq h_ns_n} n^{-1}\sum_{i=1}^{n}\hat{S}_{il}^2\}^{-1}\cdot$$

$$[\|n^{-1/2}\sum_{i=1}^{n}(\hat{w}'F_i - \breve{E}_i)\{b'(\hat{\alpha}_0 + E_i'\eta_{\mathcal{H}_n} + F_i'\hat{\eta}_{\mathcal{H}_n^c}) - b'(\hat{\alpha}_0 + F_i'\hat{\eta}_{\mathcal{H}_n^c})\}\|_\infty - c_B(\alpha)]^2\Big).$$

(S3.141)

Based on part 3) of Lemma 7, there exists a universal constant $c_5 > 0$ such that with probability tending to 1:

$$\max_{l\leq h_ns_n} n^{-1}\sum_{i=1}^{n}\hat{S}_{il}^2 \leq c_5\log^2(n)\exp\{c_5q_n\log^{1/2}(nq_ns_n)\}. \quad \text{(S3.142)}$$

By choosing $K \geq 4(c_1 + c_3 + c_4 + c_5)$ in $\mathcal{F}_n$, it follows from (S3.140), (S3.141), (S3.142), and (S3.130) that with probability tending to 1:

$$Power^*(\beta_{\mathcal{H}_n}) \geq 1 - 2n^{-1}.$$

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.