Supplementary of An Adaptive Test on High-dimensional Parameters in Generalized Linear Models

Chong Wu[1*], Gongjun Xu[2], Wei Pan[1*]
for the Alzheimer's Disease Neuroimaging Initiative[†]
[1]*Division of Biostatistics, University of Minnesota*
[2]*Department of Statistics, University of Michigan*

January 8, 2018

The supplementary material includes proofs of the theoretical results and some additional simulation results.

1 Proofs

Lemma 1. *For* $1 \leq i \leq p$, *we have*
i) if a *is even and* $a = 2d$ $(d > 0)$, $\mu^{(i)}(a) = \frac{a!}{d!2^d} n^{-d} \sigma_{ii}^d + o(n^{-d})$, *where* $\sigma_{ii} = E\left[(S_{1i})^2\right]$.
ii) if $a \geq 3$ *is odd and* $a = 2d + 1$ $(d > 0)$, $\mu^{(i)}(a) = o(n^{-(d+1)})$.

*Correspondence: wuxx0845@umn.edu (C.W.), weip@biostat.umn.edu (W.P.)

†Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

Proof of Lemma 1. Similar as Lemma 1 in Xu et al. (2016), for $a = 2d$,

$$
\begin{aligned}
& E\Big[\Big(\frac{1}{n}\sum_{j=1}^{n} S_{ji}\Big)^a\Big] \\
&= \frac{1}{n^a} \sum_{\substack{t\geq 1; l_1,\ldots,l_t>0 \\ l_1+\cdots+l_t=a}} \binom{n}{t} \frac{a!}{l_1!\ldots l_t!} \prod_{s=1}^{t} E\left[(S_{1i})^{l_s}\right] \\
&\sim \frac{1}{n^a} \sum_{\substack{t\geq 1; l_1,\ldots,l_t>0 \\ l_1+\cdots+l_t=a}} \frac{n^t a!}{t! l_1!\ldots l_t!} \prod_{s=1}^{t} E\left[(S_{1i})^{l_s}\right] \\
&= \frac{1}{n^a} \sum_{\substack{t=d \\ l_1=\cdots=l_t=2}} \frac{n^t a!}{t! l_1!\ldots l_t!} \prod_{s=1}^{t} E\left[(S_{1i})^{l_s}\right] \\
&= \frac{a!}{d! 2^d} n^{-d} \sigma_{ii}^d + o(n^{-d}).
\end{aligned}
$$

For $a = 2d + 1$, similarly we have

$$
\begin{aligned}
& E\Big[\Big(\frac{1}{n}\sum_{j=1}^{n} S_{ji}\Big)^a\Big] \\
&\sim \frac{1}{n^a} \sum_{\substack{t\geq 1; l_1,\ldots,l_t>0 \\ l_1+\cdots+l_t=a}} \frac{n^t a!}{t! l_1!\ldots l_t!} \prod_{s=1}^{t} E\left[(S_{1i})^{l_s}\right] \\
&= \sum_{\substack{t=d \\ \text{one } l_s \text{ is } 3 \\ \text{others are } 2}} \frac{a!}{n^{a-t} t! l_1!\ldots l_t!} m_i \sigma_{ii}^{(d-1)} + o(n^{-(d+1)}) \\
&= o(n^{-(d+1)}),
\end{aligned}
$$

where $m_i = E[(S_{1i})^3] = 0$. This completes the proof of Lemma 1. ∎

Lemma 2. *For $1 \leq i, j \leq p$, consider integers $h, l \geq 1$,*
(i) if $h + l$ is an even number with $h + l = 2c$, we have

$$
E[L^{(i)}(h,\mu_0)L^{(j)}(l,\mu_0)] = \frac{1}{n^c} \sum_{\substack{2c_1+c_3=h \\ 2c_2+c_3=l}} \frac{h! l!}{c_3! c_1! c_2! 2^{c_1+c_2}} \sigma_{ii}^{c_1} \sigma_{jj}^{c_2} \sigma_{ij}^{c_3} + o(n^{-c}),
$$

where $\sigma_{ii} = E[(S_{1i})^2]$, and $\sigma_{ij} = E[S_{1i}S_{1j}]$.

If $h+l$ is an odd number with $h+l=2c+1$, we have

$$E[L^{(i)}(h,\mu_0)L^{(j)}(l,\mu_0)] = \frac{1}{n^{c+1}} \sum_{\substack{a+b=3 \\ 2c_1+c_3=h-a \\ 2c_2+c_3=l-b}} \frac{h!l!}{a!b!c_3!c_1!c_2!2^{c_1+c_2}} \sigma_{ii}^{c_1}\sigma_{jj}^{c_2}\sigma_{ij}^{c_3} m_{i^a j^b} + o(n^{-(c+1)}),$$

where $m_{i^a j^b} = E[(S_{1i})^a (S_{1j})^b]$.

Proof of Lemma 2. If $h+l=2c$,

$$\begin{aligned}
&E[L^{(i)}(h,\mu_0)L^{(j)}(l,\mu_0)] \\
&= E\Big[\frac{1}{n^{h+l}}\big(\sum_{s=1}^{n} S_{si}\big)^h \big(\sum_{t=1}^{n} S_{tj}\big)^l\Big] \\
&= \frac{1}{n^{2c}} \sum_{\substack{2c_1+c_3=h \\ 2c_2+c_3=l}} \binom{h}{c_3}\binom{l}{c_3} \frac{n^{c_1+c_2+c_3}(2c_1)!(2c_2)!c_3!}{c_1!c_2!2^{c_1+c_2}} \sigma_{ii}^{c_1}\sigma_{jj}^{c_2}\sigma_{ij}^{c_3} + o(n^{-c}) \\
&= \frac{1}{n^{c}} \sum_{\substack{2c_1+c_3=h \\ 2c_2+c_3=l}} \frac{h!l!}{c_3!c_1!c_2!2^{c_1+c_2}} \sigma_{ii}^{c_1}\sigma_{jj}^{c_2}\sigma_{ij}^{c_3} + o(n^{-c}).
\end{aligned}$$

If $h+l$ is odd with $h+l=2c+1$, similarly, we have

$$\begin{aligned}
&E[L^{(i)}(h,\mu_0)L^{(j)}(l,\mu_0)] \\
&= \frac{1}{n^{h+l}} \sum_{\substack{a+b=3 \\ 2c_1+c_3=h-a \\ 2c_2+c_3=l-b}} \binom{h}{a}\binom{l}{b}\binom{h-a}{c_3}\binom{l-b}{c_3} \frac{n^{c_1+c_2}(2c_1)!(2c_2)!}{c_1!c_2!2^{c_1+c_2}} n^{c_3} c_3! n \\
&\quad \times \sigma_{ii}^{c_1}\sigma_{jj}^{c_2}\sigma_{ij}^{c_3} m_{i^a j^b} + o(n^{-(c+1)}) \\
&= \frac{1}{n^{c+1}} \sum_{\substack{a+b=3 \\ 2c_1+c_3=h-a \\ 2c_2+c_3=l-b}} \frac{h!l!}{a!b!c_3!c_1!c_2!2^{c_1+c_2}} \sigma_{ii}^{c_1}\sigma_{jj}^{c_2}\sigma_{ij}^{c_3} m_{i^a j^b} + o(n^{-(c+1)}).
\end{aligned}$$

This completes the proof. ■

Proof of Proposition 2. Note that

$$\begin{aligned}
\sigma^2(\gamma) =& E\Big[\Big\{\sum_{i=1}^{p}\big(\frac{1}{n}\sum_{s=1}^{n} S_{si}\big)^{\gamma}\Big\}^2\Big] - E\Big[\sum_{i=1}^{p}\Big(\frac{1}{n}\sum_{s=1}^{n} S_{si}\Big)^{\gamma}\Big]^2 \\
=& \mu(2\gamma) - \sum_{i=1}^{p}\{\mu^{(i)}(\gamma)\}^2 + E\Big[\sum_{i\neq j} L^{(i)}(\gamma,\mu_0)L^{(j)}(\gamma,\mu_0)\Big] - \sum_{i\neq j}\mu^{(i)}(\gamma)\mu^{(j)}(\gamma).
\end{aligned}$$

For $\gamma = 1$, note that $\mu^{(i)}(1) = 0$ for $1 \le i \le p$. By Lemma 1, 2 and C2 mixing assumption, we have

$$\begin{aligned}\sigma^2(1) =& \mu(2) + E\Big[\sum_{i\neq j} L^{(i)}(1,\mu_0)L^{(j)}(1,\mu_0)\Big] \\ =& \mu(2) + \frac{1}{n}\sum_{i\neq j}\sigma_{ij} + o(pn^{-1}) = \frac{1}{n}\sum_{1\le i,j\le p}\sigma_{ij} + o(pn^{-1}).\end{aligned}$$

We use a similar argument as in the proof of Proposition 2 in Xu et al. (2016). According to Lemmas 1 and 2 and the α-mixing assumption, for $\gamma = 2d$, we have the following expressions

$$\begin{aligned}\sigma^2(\gamma) =& \mu(2\gamma) - \sum_{i=1}^{p}\{\mu^{(i)}(\gamma)\}^2 + \sum_{i\neq j}\Big\{\frac{1}{n^\gamma}\sum_{\substack{2c_1+c_3=\gamma\\2c_2+c_3=\gamma}}\frac{(\gamma!)^2}{c_3!c_1!c_2!2^{c_1+c_2}}\sigma_{ii}^{c_1}\sigma_{jj}^{c_2}\sigma_{ij}^{c_3}\Big\} \\ & - \sum_{i\neq j}\Big(\frac{1}{n^{\gamma/2}}\frac{\gamma!}{(\gamma/2)!2^{\gamma/2}}\sigma_{ii}^{(\gamma/2)}\Big)\Big(\frac{1}{n^{\gamma/2}}\frac{\gamma!}{(\gamma/2)!2^{\gamma/2}}\sigma_{jj}^{(\gamma/2)}\Big) + o(pn^{-\gamma}) \\ =& \mu(2\gamma) - \sum_{i=1}^{p}\{\mu^{(i)}(\gamma)\}^2 + \frac{1}{n^\gamma}\sum_{i\neq j}\sum_{\substack{2c_1+c_3=\gamma\\2c_2+c_3=\gamma\\c_3>0}}\frac{(\gamma!)^2}{c_3!c_1!c_2!2^{c_1+c_2}}\sigma_{ii}^{c_1}\sigma_{jj}^{c_2}\sigma_{ij}^{c_3} + o(pn^{-\gamma}).\end{aligned}$$

Similarly, for $\gamma = 2d+1$, we have

$$\begin{aligned}\sigma^2(\gamma) =& \mu(2\gamma) - \sum_{i=1}^{p}\{\mu^{(i)}(\gamma)\}^2 + \sum_{i\neq j}\Big\{\frac{1}{n^\gamma}\sum_{\substack{2c_1+c_3=\gamma\\2c_2+c_3=\gamma}}\frac{(\gamma!)^2}{c_3!c_1!c_2!2^{c_1+c_2}}\sigma_{ii}^{c_1}\sigma_{jj}^{c_2}\sigma_{ij}^{c_3}\Big\} + o(pn^{-\gamma}) \\ =& \mu(2\gamma) - \sum_{i=1}^{p}\{\mu^{(i)}(\gamma)\}^2 + \frac{1}{n^\gamma}\sum_{i\neq j}\sum_{\substack{2c_1+c_3=\gamma\\2c_2+c_3=\gamma\\c_3>0}}\frac{(\gamma!)^2}{c_3!c_1!c_2!2^{c_1+c_2}}\sigma_{ii}^{c_1}\sigma_{jj}^{c_2}\sigma_{ij}^{c_3} + o(pn^{-\gamma}).\end{aligned}$$

This completes the proof. ■

Proof of Proposition 3. Similar to the proposition 2, under the null hypothesis,

$$\begin{aligned}& cov\{L(t,\mu_0), L(s,\mu_0)\} \\ &= E[L(t,\mu_0)L(s,\mu_0)] - E[L(t,\mu_0)]E[L(s,\mu_0)] \\ &= \mu(t+s) + E\Big[\sum_{i\neq j} L^{(i)}(t,\mu_0)L^{(j)}(s,\mu_0)\Big] - \sum_{i=1}^{p}\mu^{(i)}(t)\mu^{(i)}(s) - \sum_{i\neq j}\mu^{(i)}(t)\mu^{(j)}(s).\end{aligned}$$

Suppose $t+s=2c$ and t, s is even, we have

$$
\begin{aligned}
&cov\{L(t,\mu_0),L(s,\mu_0)\}\\
&=\mu(t+s)-\sum_{i=1}^{p}\mu^{(i)}(t)\mu^{(i)}(s)+\sum_{i\neq j}\frac{1}{n^c}\sum_{\substack{2c_1+c_3=t\\2c_2+c_3=s}}\frac{t!s!}{c_3!c_1!c_2!2^{c_1+c_2}}\sigma_{ii}^{c_1}\sigma_{jj}^{c_2}\sigma_{ij}^{c_3}\\
&\quad-\sum_{i\neq j}\frac{1}{n^c}\frac{t!s!}{(t/2)!(s/2)!2^{(t+s)/2}}\sigma_{ii}^{(t/2)}\sigma_{jj}^{(s/2)}+o(pn^{-c})\\
&=\mu(t+s)-\sum_{i=1}^{p}\mu^{(i)}(t)\mu^{(i)}(s)+\frac{1}{n^c}\sum_{i\neq j}\sum_{\substack{2c_1+c_3=t\\2c_2+c_3=s\\c_3>0}}\frac{t!s!}{c_3!c_1!c_2!2^{c_1+c_2}}\sigma_{ii}^{c_1}\sigma_{jj}^{c_2}\sigma_{ij}^{c_3}+o(pn^{-c}).
\end{aligned}
$$

If $t+s=2c$ and t and s are odd, similarly we have

$$
\begin{aligned}
&cov\{L(t,\mu_0),L(s,\mu_0)\}\\
&=\mu(t+s)-\sum_{i=1}^{p}\mu^{(i)}(t)\mu^{(i)}(s)+\sum_{i\neq j}\frac{1}{n^c}\sum_{\substack{2c_1+c_3=t\\2c_2+c_3=s}}\frac{t!s!}{c_3!c_1!c_2!2^{c_1+c_2}}\sigma_{ii}^{c_1}\sigma_{jj}^{c_2}\sigma_{ij}^{c_3}+o(pn^{-c}).
\end{aligned}
$$

If $t+s=2c+1$ and is odd, we have

$$
\begin{aligned}
cov\{L(t,\mu_0),L(s,\mu_0)\}&=\mu(t+s)-\sum_{i=1}^{p}\mu^{(i)}(t)\mu^{(i)}(s)-\sum_{i\neq j}\mu^{(i)}(t)\mu^{(j)}(s)\\
&\quad+\textstyle\sum_{i\neq j}\frac{1}{n^{c+1}}\sum_{\substack{a+b=3\\2c_1+c_3=t-a\\2c_2+c_3=s-b}}\frac{t!s!}{a!b!c_3!c_1!c_2!2^{c_1+c_3}}\sigma_{ii}^{c_1}\sigma_{jj}^{c_2}\sigma_{ij}^{c_3}m_{i^a j^b}\\
&=O(pn^{-(t+s+1)/2})=o(pn^{-(t+s)/2}).
\end{aligned}
$$

This completes the proof. ■

Proof of Theorem 1. (i) For finite $\gamma\in\Gamma$, we first show the limiting distribution for each $L(\gamma,\mu_0)$ and then the joint distribution can be easily obtained by Cramér-Wold Theorem. We use Bernstein's block idea (Ibragimov, 1971) to derive the limiting distribution for each $L(\gamma,\mu_0)$. Specifically, we partition the sequence into different blocks and then by α-mixing assumptions (C2), blocks are almost independent and some well-established results can be applied accordingly. First, we partition the sequence $\sigma^{-1}(\gamma)\left(L^{(i)}(\gamma,\mu_0)-\mu^{(i)}(\gamma)\right)$, $1\le i\le p$, into $r+1$ blocks, where each of the first r block contains b variables such that $rb\le p<(r+1)b$. Then for each $1\le j\le r$, we partition the jth block into two sub-blocks with a larger one S_{j1}, which contains b_1 variables, and a smaller one S_{j2}, which contains

$b_2 = b - b_1$ variables. Let

$$S_{j1}(\gamma) = \sum_{i=1}^{b_1} \left(L^{(j-1)b+i}(\gamma, \mu_0) - \mu^{(j-1)b+i}(\gamma)\right), \qquad 1 \le j \le r;$$
$$S_{j2}(\gamma) = \sum_{i=1}^{b_2} \left(L^{(j-1)b+b_1+i}(\gamma, \mu_0) - \mu^{(j-1)b+b_1+i}(\gamma)\right), \qquad 1 \le j \le r.$$

We further define $\mathcal{L}_1 = \sigma^{-1}(\gamma) \sum_{j=1}^{r} S_{j1}(\gamma)$; $\mathcal{L}_2 = \sigma^{-1}(\gamma) \sum_{j=1}^{r} S_{j2}(\gamma)$; $\mathcal{L}_3 = \sigma^{-1}(\gamma) \sum_{i=rb+1}^{p} \left(L^{(i)}(\gamma, \mu_0) - \mu^{(i)}(\gamma)\right)$. As a result, $\sigma^{-1}(\gamma)\left(L(\gamma, \mu_0) - \mu(\gamma)\right) = \mathcal{L}_1 + \mathcal{L}_2 + \mathcal{L}_3$.

Suppose $r \to \infty$, $b_1 \to \infty$, $b_2 \to \infty$, $rb_1/p \to 1$ and $rb_2/p \to 0$ as $p \to \infty$, the Bernstein's block method makes S_{j1}'s "almost" independent and some well-studied results of sum of independent random variables may be applied to the study of $\mathcal{L}_1$. Further, by noting that since b_2 is small compared to b_1, $\mathcal{L}_2$ and $\mathcal{L}_3$ will be small and ignorable terms compared with $\mathcal{L}_1$. We next prove

$$\sigma^{-1}(\gamma)\left(L(\gamma) - \mu(\gamma)\right) = \mathcal{L}_1 + o_p(1).$$

Note that $E(\mathcal{L}_2) = E(\mathcal{L}_3) = 0$, it is sufficient to show that $\text{var}(\mathcal{L}_2) = \text{var}(\mathcal{L}_3) = o(1)$. Consider $\text{var}(\mathcal{L}_3)$ first and we have

$$\text{var}(\mathcal{L}_3) = \sigma^{-2}(\gamma)\text{var}\Big(\sum_{j=rb+1}^{p} L^{(i)}(\gamma)\Big) \le \sigma^{-2}(\gamma) \sum_{j_1=rb+1}^{p} \sum_{j_2=rb+1}^{p} \left|\text{cov}\left(L^{(j_1)}(\gamma), L^{(j_2)}(\gamma)\right)\right|.$$

For any $\epsilon > 0$, we have the following α-mixing inequality (Guyon, 1995)

$$\begin{aligned}
&\text{cov}\Big(n^{\frac{\gamma}{2}} L^{(i)}(\gamma), n^{\frac{\gamma}{2}} L^{(j)}(\gamma)\Big) \\
&\le 8\alpha_W(|i-j|)^{\frac{\epsilon}{2+\epsilon}} \max\Big(E[(n^{\frac{\gamma}{2}} L^{(i)}(\gamma))^{2+\epsilon}]^{\frac{\epsilon}{2+\epsilon}}, E[(n^{\frac{\gamma}{2}} L^{(j)}(\gamma))^{2+\epsilon}]^{\frac{\epsilon}{2+\epsilon}}\Big).
\end{aligned}$$

Then take $\epsilon = 2$ and from Lemma 1 we have

$$\text{cov}\left(n^{\gamma/2} L^{(j_1)}(\gamma), n^{\gamma/2} L^{(j_2)}(\gamma)\right) \le C\alpha_W(|j_1 - j_2|)^{1/2},$$

where C is some constant. By Proposition 2, we have $\sigma^{-2}(\gamma) = O(1)n^{\gamma}/p$, then

$$\begin{aligned}
\text{var}(\mathcal{L}_3) &= \sigma^{-2}(\gamma)\text{var}\left(\sum_{j=rb+1}^{p} L^{(i)}(\gamma)\right) \\
&\le O(1)\frac{n^{\gamma}}{p} \sum_{j_1=rb+1}^{p} \sum_{j_2=rb+1}^{p} Cn^{-\gamma}\delta^{|j_1-j_2|/2} \le O(1)\frac{(p-rb)}{p}.
\end{aligned}$$

Since $rb/p \to 1$ as $p \to \infty$, $\mathrm{var}(\mathcal{L}_3) = o(1)$, implying $\mathcal{L}_3 = o_p(1)$. Similarly, we have $\mathcal{L}_2 = o_p(1)$. Next, we focus on $\mathcal{L}_1$. Based on the similar arguments on page 338 in Ibragimov (1971), we can properly choose r and b_2 such that

$$\left|E\{\exp(it\mathcal{L}_1\} - E^r\left[\exp(it\sigma^{-1}(\gamma)S_{11}(\gamma))\right]\right| \leq 16r\alpha_W(b_2) \to 0.$$

This implies that the limiting distribution of $\mathcal{L}_1$ is the same as that of $\sigma^{-1}(\gamma)\sum_{j=1}^r \xi_j$, where ξ_j and $S_{j1}(\gamma)$ are identically distributed. Following Xu et al. (2016), we check the Lyapunov condition via using the moment bounds (Theorem 5 in Kim (1994)). Thus, for any finite $\gamma \in \Gamma$, we have proved the asymptotic normal distribution of $L(\gamma)$. For any linear combination of $L(\gamma)$'s with respect to different γ, we can derive the asymptotic normal distribution with similar techniques. Then the Cramér-Wold Theorem implies the asymptotic joint distribution of $\{L(\gamma); \gamma \in \Gamma'\}$.

(ii) The conclusion follows directly from the proof of Theorem 6 in Cai et al. (2014). In particular, define

$$V_{ij} = \frac{(Y_i - \mu_{0i})X_{ij}}{\sqrt{\sigma_{jj}}}, \qquad 1 \leq i \leq n, 1 \leq j \leq p.$$

Let $\hat{V}_{ij} = V_{ij}I(|V_{ij}| \leq \tau_n)$ for $i = 1, \ldots, n$ and $j = 1, \ldots, p$, where $\tau_n = 2\eta^{-1/2}\sqrt{\log(p+n)}$. Define $W_j = \sum_{i=1}^n V_{ij}/\sqrt{n}$ and $\hat{W}_j = \sum_{i=1}^n \hat{V}_{ij}/\sqrt{n}$. Note that $\max_{1\leq j\leq p}|W_j - \hat{W}_j| \geq 1/\log p$ only holds when at least one $|V_{ij}| \geq \tau_n$. Then by Markov's inequality and C4,

$$Pr\left(\max_{1\leq j\leq p}|W_j - \hat{W}_j| \geq \frac{1}{\log p}\right) \leq np\max_{1\leq j\leq p}Pr(|V_{1j}| \geq \tau_n) = O(p^{-1} + n^{-1}).$$

Following the proof in Cai et al. (2014),

$$|\max_{1\leq j\leq p}W_j^2 - \max_{1\leq j\leq p}\hat{W}_j^2| \leq 2\max_{1\leq j\leq p}|W_j|\max_{1\leq j\leq p}|W_j - \hat{W}_j| + \max_{1\leq j\leq p}|W_j - \hat{W}_j|^2.$$

Then we have when $n, p \to \infty$, $|\max_{1\leq j\leq p}W_j^2 - \max_{1\leq j\leq p}\hat{W}_j^2| \to 0$. Further by Cai et al. (2014),

$$Pr\{\max_{1\leq j\leq p}\hat{W}_j^2 - 2\log p + \log\log p \leq x\} \to \exp\{-\pi^{-1/2}\exp(-x/2)\}.$$

This gives the limiting distribution of $L(\infty, \mu_0)$.

(iii) The proof of the asymptotic independence follows from a similar argument as that in Hsing (1995) and Xu et al. (2016). Here we only present the key steps. The idea is that if $L^{(i)}(\gamma, \mu_0)$ is weakly dependent and $L(\gamma, \mu_0)$ is asymptotically normal, then the individual summands must be asymptotically negligible and the maximum term should play no role in the limiting distribution, leading to the asymptotic independence results. Specially, consider the sequence of random variables $\tilde{L}^{(j)}(\gamma)$ defined on another probability space such that

$$\tilde{Pr}\left(\tilde{L}^{(j)}(\gamma) \leq x_j, 1 \leq j \leq p\right) = Pr\left(L^{(j)}(\gamma, \mu_0) \leq x_j, 1 \leq j \leq p | L(\infty, \mu_0) < a_p + x\right).$$

The expectation with respect to $\tilde{Pr}$ is denoted by $\tilde{E}$. To complete this proof, we are to show the the asymptotic normality of $\sigma^{-1}(\gamma)\big(\tilde{L}(\gamma) - \mu(\gamma)\big)$ as in proof (i). Similar as in proof (i), partition the sequence $\sigma^{-1}(\gamma)\big(\tilde{L}^{(i)}(\gamma) - \mu(\gamma)\big)$, $1 \leq i \leq p$, into r blocks, where each block contains b variables such that $rb \leq p < (r+1)b$. Then for each $1 \leq j \leq r$, we partition the jth block into two sub-blocks with a larger one $\tilde{S}_{j1}$, which contains b_1 variables, and a smaller one $\tilde{S}_{j2}$, which contains $b_2 = b - b_1$ variables. This step makes $\tilde{S}_{j1}$, $1 \leq j \leq r$ almost independent by α-mixing assumption. Let

$$\tilde{S}_{j1}(\gamma) = \sum_{i=1}^{b_1}\left(\tilde{L}^{(j-1)b+i}(\gamma) - \mu^{(j-1)b+i}(\gamma)\right), \qquad 1 \leq j \leq r;$$
$$\tilde{S}_{j2}(\gamma) = \sum_{i=1}^{b_2}\left(\tilde{L}^{(j-1)b+b_1+i}(\gamma) - \mu^{(j-1)b+b_1+i}(\gamma)\right), \qquad 1 \leq j \leq r.$$

Further define $\tilde{\mathcal{L}}_1 = \sigma^{-1}(\gamma)\sum_{j=1}^{r}\tilde{S}_{j1}(\gamma)$, $\tilde{\mathcal{L}}_2 = \sigma^{-1}(\gamma)\sum_{j=1}^{r}\tilde{S}_{j2}(\gamma)$, and $\tilde{\mathcal{L}}_3 = \sigma^{-1}(\gamma)$ $\sum_{j=rb+1}^{p}\big(\tilde{L}^{(i)}(\gamma) - \mu^{(i)}(\gamma)\big)$. Then we prove $\sigma^{-1}(\gamma)\big(\tilde{L}(\gamma) - \mu(\gamma)\big) = \tilde{\mathcal{L}}_1 + o_p(1)$. Note that $\tilde{E}[\tilde{\mathcal{L}}_2] = \tilde{E}[\tilde{\mathcal{L}}_3] = 0$. It is sufficient to show that $\tilde{E}(\tilde{\mathcal{L}}_2^2) = \tilde{E}(\tilde{\mathcal{L}}_3^2) = o(1)$ and by Cauchy-Schwarz inequality,

$$\tilde{E}(\mathcal{L}_2^2) = \sigma^{-2}(\gamma)\tilde{E}\Big[\Big(\sum_{j=1}^{r}\tilde{S}_{j2}(\gamma)\Big)^2\Big] \leq 2\sigma^{-2}(\gamma)\Big(\sum_{i\neq j}\tilde{E}[\tilde{S}_{i2}^2(\gamma)]^{\frac{1}{2}}\tilde{E}[\tilde{S}_{j2}^2(\gamma)]^{\frac{1}{2}} + \sum_{j=1}^{p}\tilde{E}[\tilde{S}_{j2}^2(\gamma)]\Big).$$

By the definition of $\tilde{E}$, $\tilde{E}\{\tilde{S}_{j2}^2(\gamma)\} = \frac{E\{S_{j2}^2(\gamma)|L(\infty)<a_p+x\}}{Pr(L(\infty)<a_p+x)} \leq \frac{E\{S_{j2}^2(\gamma)\}}{Pr(L(\infty)<a_p+x)}$. Then we have

$$\tilde{E}(\mathcal{L}_2^2) \leq \sigma^{-2}(\gamma)Pr(L(\infty) < a_p + x)^{-1} \times \Big(\sum_{i\neq j}E[S_{i2}^2(\gamma)]^{1/2}E[S_{j2}^2(\gamma)]^{1/2} + \sum_{j=1}^{p}E[S_{j2}^2(\gamma)]\Big).$$

The above bound goes to 0 under the strong mixing assumption by choosing proper b_2. Similarly, we can show that $\tilde{E}(\tilde{\mathcal{L}}_3^2) = o(1)$.

Next, we focus on $\tilde{\mathcal{L}}_1$. Following Xu et al. (2016), based on the similar arguments on page 338 in Ibragimov (1971) and a similar argument as that of Lemma 2.2 in Hsing (1995), we can properly choose r and b_2 such that $\Big|\tilde{E}[\exp\{it\tilde{\mathcal{L}}_1\}] - \tilde{E}^r\big[\exp\{it\sigma^{-1}(\gamma)\tilde{S}_{1,1}(\gamma)\}\big]\Big| \to 0$. This implies that the limiting distribution of $\tilde{\mathcal{L}}_1$ is the same as that of $\sigma^{-1}(\gamma)\sum_{j=1}^{r}\tilde{\xi}_j$, where

$\tilde{\xi}_j$ and $\tilde{S}_{j1}(\gamma)$ are identically distributed under measure $\tilde{Pr}$. Then by checking the Lyapunov condition, we show that the central limit theorem holds for $\sigma^{-1}(\gamma)\sum_{j=1}^{r}\tilde{\xi}_j$. In particular,

$$\sigma^{-4}(\gamma)\sum_{i=1}^{r}\tilde{E}(\tilde{\xi}_i^4) \le \sigma^{-4}(\gamma)\frac{\sum_{i=1}^{r}E(\xi_i^4)}{Pr(L(\infty) < a_p + x)} \to 0,$$

where ξ_i's are defined as those in proof (i) and the convergence also follows from (i).

This completes the proof. ■

Proof of Theorem 2. (i) Define $\hat{\mathbb{D}} = (Y - \hat{\mu}_0) = \{Y_1 - \hat{\mu}_{01}, \ldots, Y_n - \hat{\mu}_{0n}\}^\intercal$ and $\mathbb{D} = (Y - \mu_0) = \{Y_1 - \mu_{01}, \ldots, Y_n - \mu_{0n}\}^\intercal$. Using the approach in Le Cessie and Van Houwelingen (1991) and Guo and Chen (2016), we have

$$\hat{\mathbb{D}} = [\mathbb{I}_n - \mathbb{W}\mathbb{Z}\{\mathbb{I}(\alpha)\}^{-1}\mathbb{Z}^\intercal]\mathbb{D} + o_p(n^{-1/2}),$$

where $\mathbb{I}_n$ is the $n \times n$ identity matrix, $\mathbb{W}$ is a diagonal matrix, which is defined as $\mathbb{W} = \mathrm{diag}\{E(\epsilon_{01}^2|\mathbb{Z}), \ldots, E(\epsilon_{0n}^2|\mathbb{Z})\}$, and $\mathbb{I}(\alpha)$ is a $q \times q$ matrix given by $\mathbb{I}(\alpha) = \mathbb{Z}^\intercal\mathbb{W}\mathbb{Z}$. Since we only need prove the leading term is small, the smaller order term $o_p(n^{-1/2})$ can be ignored in the subsequent proof. For notation simplicity, define $\mathbb{B} = \mathbb{W}\mathbb{Z}\{\mathbb{I}(\alpha)\}^{-1}\mathbb{Z}^\intercal = (b_{ij})_{n\times n}$. By law of large numbers and assumption C7, $\mathbb{I}(\alpha)/n$ converges to a weighted covariance matrix almost surely and thus $\mathbb{I}(\alpha) = O(n)$ almost surely. Then by assumption C6 ($\mathbb{Z}$ is bounded almost surely), we have $b_{ij} = O(1/n)$ almost surely.

By some simple linear algebra, we have $\mu_{0i} - \hat{\mu}_{0i} = \sum_{l=1}^{n} b_{il}\epsilon_{0l}$ for $1 \le i \le n$, where $b_{il} = O(n^{-1})$. For simplicity, we denote all the constants by C, which may vary from place to place. Then we discuss different γ separately.

For $\gamma = 1$: we decompose the statistic $L(1, \hat{\mu}_0)$ as

$$L(1, \hat{\mu}_0) = \frac{1}{n}\sum_{j=1}^{p}\sum_{i=1}^{n}(Y_i - \hat{\mu}_{0i})X_{ij} = \sum_{j=1}^{p}\sum_{i=1}^{n}\frac{1}{n}S_{ij} + \sum_{j=1}^{p}\sum_{i=1}^{n}\frac{(\mu_{0i} - \hat{\mu}_{0i})X_{ij}}{n} = T_{10} + T_{11}.$$

Under the null hypothesis and proposed assumptions, the same techniques used in the proof of Theorem 1 lead to

$$T_{10}/\sigma(1) \xrightarrow{d} N(0, 1) \qquad \text{as } n \to \infty \text{ and } p \to \infty.$$

For T_{11}, by noting that $\mu_{0i} - \hat{\mu}_{0i} = \sum_{l=1}^{n} b_{il}\epsilon_{0l}$, we have

$$\begin{aligned}
E[(T_{11})^2] &= \frac{1}{n^2}\sum_{j_1=1}^{p}\sum_{j_2=1}^{p}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n} E\Big[(\mu_{0i_1} - \hat{\mu}_{0i_1})X_{i_1j_1}(\mu_{0i_2} - \hat{\mu}_{0i_2})X_{i_2j_2}\Big] \\
&= \frac{1}{n^2}\sum_{j_1=1}^{p}\sum_{j_2=1}^{p}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n} E\Big[X_{i_1j_1}X_{i_2j_2}\sum_{l=1}^{n}\epsilon_{0l}b_{i_1l}\sum_{l=1}^{n}\epsilon_{0l}b_{i_2l}\Big] \\
&= \frac{1}{n^2}\sum_{j_1=1}^{p}\sum_{j_2=1}^{p}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n} E\Big[X_{i_1j_1}X_{i_2j_2}\big(\epsilon_{0i_1}b_{i_1i_1} + \epsilon_{0i_2}b_{i_1i_2} + \sum_{l\neq i_1,i_2}\epsilon_{0l}b_{i_1l}\big) \\
&\qquad\qquad \times \big(\epsilon_{0i_1}b_{i_2i_1} + \epsilon_{0i_2}b_{i_2i_2} + \sum_{l\neq i_1,i_2}\epsilon_{0l}b_{i_2l}\big)\Big].
\end{aligned}$$

Since i_1 and i_2 are symmetrical, we have

$$\begin{aligned}
&E[(T_{11})^2] \\
\leq&\frac{1}{n^2}\sum_{j_1,j_2,i_1,i_2} CE\big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_1}^2 b_{i_1i_1}b_{i_2i_1}\big] + \frac{1}{n^2}\sum_{j_1,j_2,i_1,i_2} CE\big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_1}b_{i_1i_1}\epsilon_{0i_2}b_{i_2i_2}\big] \\
&+ \frac{1}{n^2}\sum_{j_1,j_2,i_1,i_2} CE\big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_1}b_{i_1i_1}\sum_{l\neq i_1,i_2}\epsilon_{0l}b_{i_2l}\big] + \frac{1}{n^2}\sum_{j_1,j_2,i_1,i_2} CE\big[X_{i_1j_1}X_{i_2j_2}\sum_{l\neq i_1,i_2}\epsilon_{0l}^2 b_{i_1l}b_{i_2l}\big] \\
=&E[T_{111}] + E[T_{112}] + E[T_{113}] + E[T_{114}].
\end{aligned}$$

We discuss the order of each term and show that $|T_{11}| = o_p(\sqrt{p}n^{-1/2})$ and thus can be ignored. By assumption C6, $E[X_{ij}|\mathbb{Z}] \neq 0$ only holds for $j \in P_0$, then

$$\begin{aligned}
E[T_{111}|\mathbb{Z}] &= \frac{C}{n^2}\sum_{j_1=1}^{p}\sum_{j_2=1}^{p}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n} E\big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_1}^2 b_{i_1i_1}b_{i_2i_1}|\mathbb{Z}\big] \\
&= \frac{C}{n^2}\sum_{j_1\notin P_0}\sum_{j_2\notin P_0}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n} E\big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_1}^2 b_{i_1i_1}b_{i_2i_1}|\mathbb{Z}\big] \\
&\quad + \frac{C}{n^2}\sum_{j_1\in P_0}\sum_{j_2\notin P_0}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n} E\big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_1}^2 b_{i_1i_1}b_{i_2i_1}|\mathbb{Z}\big] \\
&\quad + \frac{C}{n^2}\sum_{j_1\notin P_0}\sum_{j_2\in P_0}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n} E\big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_1}^2 b_{i_1i_1}b_{i_2i_1}|\mathbb{Z}\big] \\
&\quad + \frac{C}{n^2}\sum_{j_1\in P_0}\sum_{j_2\in P_0}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n} E\big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_1}^2 b_{i_1i_1}b_{i_2i_1}|\mathbb{Z}\big].
\end{aligned}$$

For the first term, by $E[X_{ij_1}|\mathbb{Z}]=0$ for $j\notin P_0$, we have

$$\begin{aligned}
&\frac{C}{n^2}\sum_{j_1\notin P_0}\sum_{j_2\notin P_0}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}E\big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_1}^2 b_{i_1i_1}b_{i_2i_1}|\mathbb{Z}\big]\\
\leq&O(n^{-2})\sum_{j_1\notin P_0}\sum_{j_2\notin P_0}\sum_{i_1=1}^{n}E\big[X_{i_1j_1}X_{i_1j_2}\epsilon_{0i_1}^2|\mathbb{Z}\big]\times O(n^{-2})\\
\leq&O(n^{-2})\times O(np^2)\times O(n^{-2})=O(n^{-3}p^2)=o(pn^{-1}).
\end{aligned}$$

For the second term, we have

$$\begin{aligned}
&\frac{C}{n^2}\sum_{j_1\in P_0}\sum_{j_2\notin P_0}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}E\big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_1}^2 b_{i_1i_1}b_{i_2i_1}|\mathbb{Z}\big]\\
\leq&O(n^{-2})\sum_{j_1\in P_0}\sum_{j_2\notin P_0}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}E\big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_1}^2|\mathbb{Z}\big]\times O(n^{-2})\\
\leq&O(n^{-2})\times O(pp_0n^2)\times O(n^{-2})=O(pp_0n^{-2})=o(pn^{-1}).
\end{aligned}$$

Noting that $p_0=O(p^{1/2-\delta})=o(n)$, we can derive the last equation. Similar to the derivation of the second term, for the third term, we have

$$\frac{C}{n^2}\sum_{j_1\notin P_0}\sum_{j_2\in P_0}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}E\big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_1}^2 b_{i_1i_1}b_{i_2i_1}|\mathbb{Z}\big]=o(pn^{-1}).$$

For the last term, we have

$$\begin{aligned}
&\frac{C}{n^2}\sum_{j_1\in P_0}\sum_{j_2\in P_0}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}E\big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_1}^2 b_{i_1i_1}b_{i_2i_1}|\mathbb{Z}\big]\\
\leq&O(n^{-2})\sum_{j_1\in P_0}\sum_{j_2\in P_0}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}E\big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_1}^2|\mathbb{Z}\big]\times O(n^{-2})\\
\leq&O(n^{-2})\times O(p_0^2n^2)\times O(n^{-2})=O(n^{-2}p_0^2)=o(pn^{-1}).
\end{aligned}$$

Combining the above derivations, we have $E[T_{111}|\mathbb{Z}]=o(pn^{-1})$. Next, we discuss the order of $E[T_{112}|\mathbb{Z}]$. By noting that $E[X_{ij}\epsilon_{0i}|\mathbb{Z}]=0$, we have

$$
\begin{aligned}
E[T_{112}|\mathbb{Z}] &= \frac{C}{n^2}\sum_{j_1=1}^{p}\sum_{j_2=1}^{p}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n} E\big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_1}b_{i_1i_1}\epsilon_{0i_2}b_{i_2i_2}|\mathbb{Z}\big] \\
&= O(n^{-2})\sum_{j_1=1}^{p}\sum_{j_2=1}^{p}\sum_{i_1=1}^{n} E\big[X_{i_1j_1}\epsilon_{0i_1}^2X_{i_1j_2}|\mathbb{Z}\big]\times O(n^{-2}) \\
&= O(n^{-2})\times O(p^2n)\times O(n^{-2}) = O(pn^{-1}pn^{-2}) = o(pn^{-1}).
\end{aligned}
$$

Then we discuss the order of $E[T_{113}|\mathbb{Z}]$:

$$
\begin{aligned}
E[T_{113}|\mathbb{Z}] =&\frac{C}{n^2}\sum_{j_1,j_2,i_1,i_2} E\Big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_1}b_{i_1i_1}\sum_{l\neq i_1,i_2}\epsilon_{0l}b_{i_2l}|\mathbb{Z}\Big] \\
=&O(n^{-2})\sum_{j_1,j_2,i_1,i_2} E\Big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_1}b_{i_1i_1}|\mathbb{Z}\Big]E\Big[\sum_{l\neq i_1,i_2}\epsilon_{0l}b_{i_2l}|\mathbb{Z}\Big] \\
=&O(n^{-2})\sum_{j_1,j_2,i_1,i_2} E\Big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_1}b_{i_1i_1}|\mathbb{Z}\Big]\times 0 = 0.
\end{aligned}
$$

Next, we discuss the order of $E[T_{114}|\mathbb{Z}]$. Similar to before, we have

$$
\begin{aligned}
E[T_{114}|\mathbb{Z}] &= \frac{C}{n^2}\sum_{j_1,j_2,i_1,i_2} E\big[X_{i_1j_1}X_{i_2j_2}\sum_{l\neq i_1,i_2}\epsilon_{0l}^2b_{i_1l}b_{i_2l}|\mathbb{Z}\big] \\
&\leq \frac{C}{n^2}\sum_{j_1,j_2}\sum_{i_1,i_2} E\big[X_{i_1j_1}X_{i_2j_2}|\mathbb{Z}\big]E\big[\sum_{l\neq i_1,i_2}\epsilon_{0l}^2b_{i_1l}b_{i_2l}|\mathbb{Z}\big] \\
&\leq O(n^{-2})\sum_{j_1,j_2}\sum_{i_1,i_2} E\big[X_{i_1j_1}X_{i_2j_2}|\mathbb{Z}\big]\times O(n^{-1}) \\
&\leq O(n^{-2})\sum_{j_1\notin P_0,j_2\notin P_0}\sum_{i_1,i_2} E\big[X_{i_1j_1}X_{i_2j_2}|\mathbb{Z}\big]\times O(n^{-1}) \\
&\quad + O(n^{-2})\sum_{j_1\notin P_0,j_2\in P_0}\sum_{i_1,i_2} E\big[X_{i_1j_1}X_{i_2j_2}|\mathbb{Z}\big]\times O(n^{-1}) \\
&\quad + O(n^{-2})\sum_{j_1\in P_0,j_2\in P_0}\sum_{i_1,i_2} E\big[X_{i_1j_1}X_{i_2j_2}|\mathbb{Z}\big]\times O(n^{-1}).
\end{aligned}
$$

Similarly, we discuss each term of $E[T_{114}|\mathbb{Z}]$ separately. By conditionally α-mixing condition

(C9), we have

$$\begin{aligned}
&O(n^{-2}) \sum_{j_1 \notin P_0, j_2 \notin P_0} \sum_{i_1, i_2} E\big[X_{i_1 j_1} X_{i_2 j_2} | \mathbb{Z}\big] \times O(n^{-1}) \\
=&O(n^{-2}) \sum_{j_1 \notin P_0, j_2 \notin P_0} \sum_{i_1=1}^{n} E\big[X_{i_1 j_1} X_{i_1 j_2} | \mathbb{Z}\big] \times O(n^{-1}) \\
=&O(n^{-2}) \times O(pn) \times O(n^{-1}) = O(pn^{-2}) = o(pn^{-1}).
\end{aligned}$$

By $E[X_{ij}|\mathbb{Z}] = 0$ for $j \notin P_0$, we have

$$\begin{aligned}
&O(n^{-2}) \sum_{j_1 \notin P_0, j_2 \in P_0} \sum_{i_1, i_2} E\big[X_{i_1 j_1} X_{i_2 j_2} | \mathbb{Z}\big] \times O(n^{-1}) \\
=&O(n^{-2}) \sum_{j_1 \notin P_0, j_2 \in P_0} \sum_{i_1=1}^{n} E\big[X_{i_1 j_1} X_{i_1 j_2} | \mathbb{Z}\big] \times O(n^{-1}) \\
=&O(n^{-2}) \times O(p_0 n) \times O(n^{-1}) = O(n^{-1} p_0 n^{-1}) = o(pn^{-1}).
\end{aligned}$$

Similarly, we have

$$\begin{aligned}
&O(n^{-2}) \sum_{j_1 \in P_0, j_2 \in P_0} \sum_{i_1, i_2} E\big[X_{i_1 j_1} X_{i_2 j_2} | \mathbb{Z}\big] \times O(n^{-1}) \\
=&O(n^{-2}) \times O(p_0^2 n^2) \times O(n^{-1}) = O(p_0^2 n^{-1}) = o(pn^{-1}).
\end{aligned}$$

The last equation discussed above comes from the assumptions that $p_0 = O(p^{1/2-\delta})$ for a small positive δ and $p = o(n^2)$. Combining the above equations, we have $E[T_{114}|\mathbb{Z}] = o(pn^{-1})$. In summary, we have $E[(T_{11})^2] = E\big[E[T_{111}|\mathbb{Z}] + E[T_{112}|\mathbb{Z}] + E[T_{113}|\mathbb{Z}] + E[T_{113}|\mathbb{Z}]\big] = o(pn^{-1})$, leading to $|T_{11}| = o_p(n^{-1/2}\sqrt{p})$.

For $1 < \gamma < \infty$: we decompose the statistic $L(\gamma, \hat{\mu}_0)$ as

$$\begin{aligned}
L(\gamma, \hat{\mu}_0) &= \sum_{j=1}^{p} \Big(\frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{\mu}_{0i}) X_{ij} \Big)^{\gamma} \\
&= \sum_{j=1}^{p} \Big(\frac{1}{n} \sum_{i=1}^{n} \big((Y_i - \mu_{0i}) + (\mu_{0i} - \hat{\mu}_{0i}) \big) X_{ij} \Big)^{\gamma} \\
&= \sum_{j=1}^{p} \Big(\frac{1}{n} \sum_{i=1}^{n} S_{ij} \Big)^{\gamma} + \sum_{1 \le v \le \gamma} \binom{\gamma}{v} \sum_{j=1}^{p} \Big(\frac{1}{n} \sum_{i=1}^{n} S_{ij} \Big)^{\gamma - v} \Big(\frac{1}{n} \sum_{i=1}^{n} (\mu_{0i} - \hat{\mu}_{0i}) X_{ij} \Big)^{v} \\
&= T_{\gamma 0} + \sum_{v=1}^{\gamma} T_{\gamma v}, \qquad \text{say.}
\end{aligned}$$

Under the null hypothesis and proposed assumptions, the same techniques used in the proof of Theorem 1 lead to $T_{\gamma 0}/\sigma(\gamma) \xrightarrow{d} N(0,1)$ as $n \to \infty$ and $p \to \infty$. Then we discuss the orders of $T_{\gamma v}$, $1 \le v \le \gamma$ separately to complete the proof. Specially, we divide the discussion into two cases: $v = 1$ and $v > 1$.

When $v = 1$, we have

$$
\begin{aligned}
&E[(T_{\gamma 1})^2] \\
&= E\Big[\frac{C}{n^2}\sum_{j_1}^{p}\sum_{j_2}^{p}\Big(\sum_{i=1}^{n}\frac{1}{n}S_{ij_1}\Big)^{\gamma-1}\Big(\sum_{i=1}^{n}\frac{1}{n}S_{ij_2}\Big)^{\gamma-1}\sum_{i=1}^{n}((\mu_{0i}-\hat{\mu}_{0i})X_{ij_1})\sum_{i=1}^{n}((\mu_{0i}-\hat{\mu}_{0i})X_{ij_2})\Big] \\
&= \frac{C}{n^2}\sum_{j_1,j_2}\sum_{i_1,i_2,i_3,i_4} E\Big[X_{i_1 j_1}X_{i_2 j_2}\epsilon_{0i_3}b_{i_1 i_3}\epsilon_{0i_4}b_{i_2 i_4} \times \big(\sum_{l\in\{i_1,i_2,i_3,i_4\}}\frac{\epsilon_{0l}X_{lj_1}}{n} + \sum_{l\notin\{i_1,i_2,i_3,i_4\}}\frac{\epsilon_{0l}X_{lj_1}}{n}\big)^{\gamma-1} \\
&\qquad \times \big(\sum_{l\in\{i_1,i_2,i_3,i_4\}}\frac{\epsilon_{0l}X_{lj_2}}{n} + \sum_{l\notin\{i_1,i_2,i_3,i_4\}}\frac{\epsilon_{0l}X_{lj_2}}{n}\big)^{\gamma-1}\Big].
\end{aligned}
$$

By Binomial theorem, we have

$$
\begin{aligned}
\big(\sum_{l}\frac{\epsilon_{0l}X_{lj_1}}{n}\big)^{\gamma-1} \le{}& \big(\sum_{l\notin\{i_1,\dots,i_4\}}\frac{\epsilon_{0l}X_{lj_1}}{n}\big)^{\gamma-1} + C\sum_{l\in\{i_1,\dots,i_4\}}\frac{\epsilon_{0l}X_{lj_1}}{n}\big(\sum_{l\notin\{i_1,\dots,i_4\}}\frac{\epsilon_{0l}X_{lj_1}}{n}\big)^{\gamma-2} + \dots \\
&+ C\big(\sum_{l\in\{i_1,\dots,i_4\}}\frac{\epsilon_{0l}X_{lj_1}}{n}\big)^{\gamma-2}\sum_{l\notin\{i_1,\dots,i_4\}}\frac{\epsilon_{0l}X_{lj_1}}{n}\big) + C\big(\sum_{l\in\{i_1,\dots,i_4\}}\frac{\epsilon_{0l}X_{lj_1}}{n}\big)^{\gamma-1}.
\end{aligned}
$$

Then

$$
\begin{aligned}
E[(T_{\gamma 1})^2] &= \sum_{k_1=1}^{\gamma}\sum_{k_2=1}^{\gamma}\frac{C}{n^2}\sum_{j_1,j_2}\sum_{i_1,i_2,i_3,i_4} E\Big[X_{i_1 j_1}X_{i_2 j_2}\epsilon_{0i_3}b_{i_1 i_3}\epsilon_{0i_4}b_{i_2 i_4}\big(\sum_{l\in\{i_1,\dots,i_4\}}\frac{\epsilon_{0l}X_{lj_1}}{n}\big)^{k_1-1} \\
&\qquad \times \big(\sum_{l\in\{i_1,\dots,i_4\}}\frac{\epsilon_{0l}X_{lj_2}}{n}\big)^{k_2-1}\big(\sum_{l\notin\{i_1,i_2,i_3,i_4\}}\frac{\epsilon_{0l}X_{lj_1}}{n}\big)^{\gamma-k_1}\big(\sum_{l\notin\{i_1,i_2,i_3,i_4\}}\frac{\epsilon_{0l}X_{lj_2}}{n}\big)^{\gamma-k_2}\Big] \\
&= \sum_{k_1=1}^{\gamma}\sum_{k_2=1}^{\gamma} T_{\gamma 1 k_1 k_2}, \qquad \text{say.}
\end{aligned}
$$

To prove the order of $|T_{\gamma 1}|$ is ignorable, we discuss two situations: $k_1+k_2 \le 6$ and $k_1+k_2 > 6$. First, we focus on the situation with $k_1 + k_2 \le 6$ and discuss the order of each term individually. By Lemma 2, we have

$$
\begin{aligned}
&T_{\gamma 111}\\
=&\frac{C}{n^2}\sum_{j_1,j_2}\sum_{i_1,i_2,i_3,i_4} E\Big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_3}b_{i_1i_3}\epsilon_{0i_4}b_{i_2i_4}\big(\sum_{l\notin\{i_1,i_2,i_3,i_4\}}\frac{\epsilon_{0l}X_{lj_1}}{n}\big)^{\gamma-1}\big(\sum_{l\notin\{i_1,i_2,i_3,i_4\}}\frac{\epsilon_{0l}X_{lj_2}}{n}\big)^{\gamma-1}\Big]\\
=&\frac{C}{n^2}\sum_{j_1,j_2}\sum_{i_1,i_2,i_3,i_4} E\big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_3}b_{i_1i_3}\epsilon_{0i_4}b_{i_2i_4}\big]E\Big[\big(\sum_{l\notin\{i_1,\dots,i_4\}}\frac{\epsilon_{0l}X_{lj_1}}{n}\big)^{\gamma-1}\big(\sum_{l\notin\{i_1,\dots,i_4\}}\frac{\epsilon_{0l}X_{lj_2}}{n}\big)^{\gamma-1}\Big]\\
=&O(n^{-2})\sum_{j_1,j_2}\sum_{i_1,i_2,i_3,i_4} E\big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_3}b_{i_1i_3}\epsilon_{0i_4}b_{i_2i_4}\big]\times O(n^{-(\gamma-1)})\\
=&O(n^{-2})\sum_{j_1,j_2}\sum_{i_1,i_2,i_3,i_4} E\big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_3}b_{i_1i_3}\epsilon_{0i_4}b_{i_2i_4}\big]\times O(n^{-(\gamma-1)}).
\end{aligned}
$$

Note that

$$
\begin{aligned}
&\sum_{j_1,j_2}\sum_{i_1,i_2,i_3,i_4} E\big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_3}b_{i_1i_3}\epsilon_{0i_4}b_{i_2i_4}|\mathbb{Z}\big]\times O(n^{-(\gamma-1)})\\
\le&\sum_{j_1,j_2}\sum_{i_1,i_2,i_3,i_4} E\big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_3}\epsilon_{0i_4}|\mathbb{Z}\big]\times O(n^{-(\gamma+1)})\\
\le&\sum_{j_1,j_2}\sum_{i_1,i_2,i_3} E\big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_3}^2|\mathbb{Z}\big]\times O(n^{-(\gamma+1)})\\
&+\sum_{j_1,j_2}\sum_{i_1,i_2} E\big[X_{i_1j_1}\epsilon_{0i_1}X_{i_2j_2}\epsilon_{0i_2}|\mathbb{Z}\big]\times O(n^{-(\gamma+1)})\\
&+\sum_{j_1,j_2}\sum_{i_1,i_2} E\big[X_{i_1j_1}\epsilon_{0i_1}^2X_{i_2j_2}|\mathbb{Z}\big]\times O(n^{-(\gamma+1)}).
\end{aligned}
$$

We discuss each term separately. By a similar discussion of $E[T_{11}|\mathbb{Z}]$, for the first term in $T_{\gamma 111}$, we have

$$
\begin{aligned}
&\sum_{j_1,j_2}\sum_{i_1,i_2,i_3} E\big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_3}^2|\mathbb{Z}\big]\times O(n^{-(\gamma+1)})\\
\le&\sum_{j_1,j_2}\sum_{i_1,i_2} E\big[X_{i_1j_1}X_{i_2j_2}|\mathbb{Z}\big]\times O(n^{-\gamma})\\
=&O(pn+p_0^2n^2)\times O(n^{-\gamma})=o(pn^{-\gamma+2}).
\end{aligned}
$$

For the second term in $T_{\gamma 111}$, we have

$$
\begin{aligned}
&\sum_{j_1,j_2}\sum_{i_1,i_2} E\big[X_{i_1j_1}\epsilon_{0i_1}X_{i_2j_2}\epsilon_{0i_2}|\mathbb{Z}\big]\times O(n^{-(\gamma+1)})\\
\le&O(p^2n)\times O(n^{-(\gamma+1)})=O(p^2n^{-\gamma})=o(pn^{-\gamma+2}).
\end{aligned}
$$

For the third term in $T_{\gamma 111}$, we have

$$\begin{aligned}
&\sum_{j_1,j_2}\sum_{i_1,i_2} E\big[X_{i_1j_1}\epsilon_{0i_1}^2 X_{i_2j_2}|\mathbb{Z}\big] \times O(n^{-(\gamma+1)})\\
\leq & O(p^2n + p_0pn^2 + p_0^2n^2) \times O(n^{-(\gamma+1)})\\
= & O(p^2n^{-\gamma} + p_0pn^{-\gamma+1} + p_0^2n^{-\gamma+1}) = o(pn^{-\gamma+2}).
\end{aligned}$$

Combing the above equations, we have $T_{\gamma 111} = o(pn^{-\gamma})$. Similarly,

$$\begin{aligned}
& T_{\gamma 121}\\
= & \frac{C}{n^2}\sum_{j_1,j_2}\sum_{i_1,\ldots,i_4} E\Big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_3}b_{i_1i_3}\epsilon_{0i_4}b_{i_2i_4}\sum_{l\in\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_1}}{n}\\
& \qquad \times \big(\sum_{l\notin\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_1}}{n}\big)^{\gamma-2}\big(\sum_{l\notin\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_2}}{n}\big)^{\gamma-1}\Big]\\
= & \frac{C}{n^2}\sum_{j_1,j_2}\sum_{i_1,\ldots,i_4} E\Big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_3}b_{i_1i_3}\epsilon_{0i_4}b_{i_2i_4}\sum_{l\in\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_1}}{n}\Big]\\
& \qquad \times E\Big[\big(\sum_{l\notin\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_1}}{n}\big)^{\gamma-2}\big(\sum_{l\notin\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_2}}{n}\big)^{\gamma-1}\Big]\\
= & \frac{C}{n^3}\sum_{j_1,j_2}\sum_{i_1,i_2,i_3} E\big[X_{i_1j_1}^2X_{i_2j_2}\epsilon_{0i_1}b_{i_1i_3}b_{i_2i_3}\epsilon_{0i_3}^2 + X_{i_1j_1}X_{i_2j_2}b_{i_1i_3}b_{i_2i_3}\epsilon_{0i_3}^3X_{i_3j_1}\big] \times O(n^{-(\gamma-1)})\\
= & O(n^{-3})\sum_{j_1,j_2}\sum_{i_1,i_2,i_3}\Big(E\big[X_{i_1j_1}^2X_{i_2j_2}\epsilon_{0i_1}b_{i_1i_3}b_{i_2i_3}\epsilon_{0i_3}^2\big] + E\big[X_{i_1j_1}X_{i_2j_2}b_{i_1i_3}b_{i_2i_3}\epsilon_{0i_3}^3X_{i_3j_1}\big]\Big) \times O(n^{-(\gamma-1)}).
\end{aligned}$$

Similar to before, we discuss each term separately. Note that since $E[\epsilon|\mathbb{X},\mathbb{Z}] = 0$, we have $E[X_{ij}^2\epsilon_{0i}|\mathbb{Z}] = 0$. Thus

$$\begin{aligned}
&\sum_{j_1,j_2}\sum_{i_1,i_2,i_3} E\big[X_{i_1j_1}^2X_{i_2j_2}\epsilon_{0i_1}b_{i_1i_3}b_{i_2i_3}\epsilon_{0i_3}^2|\mathbb{Z}\big] \times O(n^{-(\gamma-1)})\\
= & \sum_{j_1,j_2}\sum_{i_1,i_2} E\big[X_{i_1j_1}^2X_{i_2j_2}\epsilon_{0i_1}^3|\mathbb{Z}\big] \times O(n^{-(\gamma+1)})\\
= & O(p^2n^2) \times O(n^{-(\gamma+1)}) = O(p^2n^{-(\gamma-1)}) = o(pn^{-(\gamma-3)}).
\end{aligned}$$

Similarly, for the second term in $T_{\gamma 121}$, we have

$$\begin{aligned}
&\sum_{j_1,j_2}\sum_{i_1,i_2,i_3} E\big[X_{i_1j_1}X_{i_2j_2}b_{i_1i_3}b_{i_2i_3}\epsilon_{0i_3}^3X_{i_3j_1}|\mathbb{Z}\big]\times O(n^{-(\gamma-1)})\\
=&\sum_{j_1,j_2}\sum_{i_1,i_2,i_3} E\big[X_{i_1j_1}X_{i_2j_2}|\mathbb{Z}\big]E\big[\epsilon_{0i_3}^3X_{i_3j_1}|\mathbb{Z}\big]\times O(n^{-(\gamma+1)})\\
=&\sum_{j_1,j_2}\sum_{i_1,i_2} E\big[X_{i_1j_1}X_{i_2j_2}|\mathbb{Z}\big]\times O(n^{-\gamma}).
\end{aligned}$$

By noting that (similar to the derivation of $E[T_{114}|\mathbb{Z}]$), we have

$$\sum_{j_1,j_2}\sum_{i_1,i_2} E\big[X_{i_1j_1}X_{i_2j_2}|\mathbb{Z}\big] = O(pn + p_0n + p_0^2n^2).$$

Then

$$\sum_{j_1,j_2}\sum_{i_1,i_2,i_3} E\big[X_{i_1j_1}X_{i_2j_2}b_{i_1i_3}b_{i_2i_3}\epsilon_{0i_3}^3X_{i_3j_1}|\mathbb{Z}\big]\times O(n^{-(\gamma-1)}) = o(pn^{-(\gamma-3)}).$$

Combining these two parts, we have $T_{\gamma 121} = o(pn^{-\gamma})$. For $T_{\gamma 122}$, we have

$$\begin{aligned}
&T_{\gamma 122}\\
=&\frac{C}{n^2}\sum_{j_1,j_2}\sum_{i_1,\ldots,i_4} E\Big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_3}b_{i_1i_3}\epsilon_{0i_4}b_{i_2i_4}\sum_{l\in\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_1}}{n}\sum_{l\in\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_2}}{n}\\
&\qquad\times\Big(\sum_{l\notin\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_1}}{n}\Big)^{\gamma-2}\Big(\sum_{l\notin\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_2}}{n}\Big)^{\gamma-2}\Big]\\
=&\frac{C}{n^2}\sum_{j_1,j_2}\sum_{i_1,\ldots,i_4} E\Big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_3}b_{i_1i_3}\epsilon_{0i_4}b_{i_2i_4}\sum_{l\in\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_1}}{n}\sum_{l\in\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_2}}{n}\Big]\\
&\qquad\times E\Big[\Big(\sum_{l\notin\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_1}}{n}\Big)^{\gamma-2}\Big(\sum_{l\notin\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_2}}{n}\Big)^{\gamma-2}\Big]\\
=&O(n^{-4})\sum_{j_1,j_2}\sum_{i_1,\ldots,i_4} E\big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_3}b_{i_1i_3}\epsilon_{0i_4}b_{i_2i_4}\sum_{l\in\{i_1,\ldots,i_4\}}\epsilon_{0l}X_{lj_1}\sum_{l\in\{i_1,\ldots,i_4\}}\epsilon_{0l}X_{lj_2}\big]\times O(n^{-\gamma+2}).
\end{aligned}$$

Note that

$$\begin{aligned}
&\sum_{j_1,j_2}\sum_{i_1,\dots,i_4} E\big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_3}b_{i_1i_3}\epsilon_{0i_4}b_{i_2i_4}\sum_{l\in\{i_1,\dots,i_4\}}\epsilon_{0l}X_{lj_1}\sum_{l\in\{i_1,\dots,i_4\}}\epsilon_{0l}X_{lj_2}|\mathbb{Z}\big]\times O(n^{-\gamma+2})\\
\le&\sum_{j_1,j_2}\sum_{i_1,i_2} E\big[X_{i_1j_1}^2X_{i_2j_2}^2\epsilon_{0i_1}^2\epsilon_{0i_2}^2+X_{i_1j_1}^2X_{i_1j_2}X_{i_2j_2}\epsilon_{0i_1}^4|\mathbb{Z}\big]\times O(n^{-\gamma})\\
&+\sum_{j_1,j_2}\sum_{i_1,i_2,i_3} E\big[X_{i_1j_1}^2\epsilon_{0i_1}X_{i_2j_2}^2\epsilon_{0i_2}\epsilon_{0i_3}^2|\mathbb{Z}\big]\times O(n^{-\gamma})\\
&+\sum_{j_1,j_2}\sum_{i_1,i_2,i_3} E\big[X_{i_1j_1}^2X_{i_1j_2}\epsilon_{0i_1}^2X_{i_2j_2}\epsilon_{0i_3}^2|\mathbb{Z}\big]\times O(n^{-\gamma})\\
&+\sum_{j_1,j_2}\sum_{i_1,i_2,i_3} E\big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_3}^4X_{i_3j_1}X_{i_3j_2}|\mathbb{Z}\big]\times O(n^{-\gamma})\\
&+\sum_{j_1,j_2}\sum_{i_1,i_2,i_3,i_4} E\big[X_{i_1j_1}X_{i_2j_2}X_{i_3j_1}\epsilon_{0i_3}^2X_{i_4j_2}\epsilon_{0i_4}^2|\mathbb{Z}\big]\times O(n^{-\gamma}).
\end{aligned}$$

Then we discuss each term separately. For the first term, we have

$$\begin{aligned}
&\sum_{j_1,j_2}\sum_{i_1,i_2} E\Big[X_{i_1j_1}^2X_{i_2j_2}^2\epsilon_{0i_1}^2\epsilon_{0i_2}^2+X_{i_1j_1}^2X_{i_1j_2}X_{i_2j_2}\epsilon_{0i_1}^4|\mathbb{Z}\Big]\times O(n^{-\gamma})\\
=&O(p^2n^2)\times O(n^{-\gamma})=o(pn^{-\gamma+4}).
\end{aligned}$$

For the second term, by noting that $E[X_{ij}^2\epsilon_{0i}|\mathbb{Z}]=0$, we have

$$\begin{aligned}
&\sum_{j_1,j_2}\sum_{i_1,i_2,i_3} E\big[X_{i_1j_1}^2\epsilon_{0i_1}X_{i_2j_2}^2\epsilon_{0i_2}\epsilon_{0i_3}^2|\mathbb{Z}\big]\times O(n^{-\gamma})\\
=&\sum_{j_1,j_2}\sum_{i_1,i_2,i_3} E\big[X_{i_1j_1}^2\epsilon_{0i_1}X_{i_2j_2}^2\epsilon_{0i_2}|\mathbb{Z}\big]\times E\big[\epsilon_{0i_3}^2|\mathbb{Z}\big]\times O(n^{-\gamma})\\
=&\sum_{j_1,j_2}\sum_{i_1,i_3} E\big[X_{i_1j_1}^2\epsilon_{0i_1}^2X_{i_1j_2}^2|\mathbb{Z}\big]\times E\big[\epsilon_{0i_3}^2|\mathbb{Z}\big]\times O(n^{-\gamma})\\
=&O(p^2n)\times O(n)\times O(n^{-\gamma})=o(pn^{-\gamma+4}).
\end{aligned}$$

For the third term, by noting that $E[X_{ij}|\mathbb{Z}]=0$ for $j\notin P_0$, we have

$$
\begin{aligned}
&\sum_{j_1,j_2}\sum_{i_1,i_2,i_3} E\big[X_{i_1j_1}^2X_{i_1j_2}\epsilon_{0i_1}^2X_{i_2j_2}\epsilon_{0i_3}^2|\mathbb{Z}\big]\times O(n^{-\gamma})\\
\le&\sum_{j_1,j_2}\sum_{i_1,i_2} E\big[X_{i_1j_1}^2X_{i_1j_2}\epsilon_{0i_1}^2X_{i_2j_2}|\mathbb{Z}\big]\times\sum_{i_3}E\big[\epsilon_{0i_3}^2|\mathbb{Z}\big]\times O(n^{-\gamma})\\
\le&\sum_{j_1,j_2}\sum_{i_1,i_2} E\big[X_{i_1j_1}^2X_{i_1j_2}\epsilon_{0i_1}^2X_{i_2j_2}|\mathbb{Z}\big]\times O(n^{-\gamma})\\
\le&\sum_{j_1\notin P_0,j_2\notin P_0}\sum_{i_1} E\big[X_{i_1j_1}^2X_{i_1j_2}^2\epsilon_{0i_1}^2|\mathbb{Z}\big]\times O(n^{-\gamma})\\
&+\sum_{j_1\in P_0,j_2\notin P_0}\sum_{i_1} E\big[X_{i_1j_1}^2X_{i_1j_2}^2\epsilon_{0i_1}^2|\mathbb{Z}\big]\times O(n^{-\gamma})\\
&+\sum_{j_1\notin P_0,j_2\in P_0}\sum_{i_1,i_2} E\big[X_{i_1j_1}^2X_{i_1j_2}\epsilon_{0i_1}^2X_{i_2j_2}|\mathbb{Z}\big]\times O(n^{-\gamma})\\
&+\sum_{j_1\in P_0,j_2\in P_0}\sum_{i_1,i_2} E\big[X_{i_1j_1}^2X_{i_1j_2}\epsilon_{0i_1}^2X_{i_2j_2}|\mathbb{Z}\big]\times O(n^{-\gamma})\\
\le&O(p^2n)\times O(n^{-\gamma})+O(pp_0n)\times O(n^{-\gamma})+O(pp_0n^2)\times O(n^{-\gamma})+O(p_0^2n^2)\times O(n^{-\gamma})\\
=&o(pn^{-\gamma+4}).
\end{aligned}
$$

For the fourth term, similar to the derivation of $E[T_{114}|\mathbb{Z}]$, we have

$$
\begin{aligned}
&\sum_{j_1,j_2}\sum_{i_1,i_2,i_3} E\big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_3}^4X_{i_3j_1}X_{i_3j_2}|\mathbb{Z}\big]\times O(n^{-\gamma})\\
=&\sum_{j_1,j_2}\sum_{i_1,i_2,i_3} E\big[X_{i_1j_1}X_{i_2j_2}|\mathbb{Z}\big]E\big[\epsilon_{0i_3}^4X_{i_3j_1}X_{i_3j_2}|\mathbb{Z}\big]\times O(n^{-\gamma})\\
=&\sum_{j_1,j_2}\sum_{i_1,i_2} E\big[X_{i_1j_1}X_{i_2j_2}|\mathbb{Z}\big]\times O(n)\times O(n^{-\gamma})\\
=&O(pn+p_0^2n^2)\times O(n^{-\gamma+1})=o(pn^{-\gamma+4}).
\end{aligned}
$$

For the last term, by the derivation of $E[T_{114}|\mathbb{Z}]$, we have

$$
\begin{aligned}
&\sum_{j_1,j_2}\sum_{i_1,i_2,i_3,i_4} E\Big[X_{i_1j_1}X_{i_2j_2}X_{i_3j_1}\epsilon_{0i_3}^2X_{i_4j_2}\epsilon_{0i_4}^2|\mathbb{Z}\Big]\times O(n^{-\gamma})\\
=&\sum_{j_1,j_2}\sum_{i_1,i_2} E[X_{i_1j_1}X_{i_2j_2}|\mathbb{Z}]\times O(n^2)\times O(n^{-\gamma})\\
=&\sum_{j_1,j_2}\sum_{i_1,i_2} E[X_{i_1j_1}X_{i_2j_2}|\mathbb{Z}]=O(pn+p_0n+p_0^2n^2)\times O(n^{-\gamma+2})=o(pn^{-\gamma+4}).
\end{aligned}
$$

Combining the above equations, we have $T_{\gamma122}=o(pn^{-\gamma})$. Then we discuss the order of

$T_{\gamma 131}$, we have

$$\begin{aligned}
T_{\gamma 131} = \frac{C}{n^2} \sum_{j_1,j_2} \sum_{i_1,\ldots,i_4} & E\Big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_3}b_{i_1i_3}\epsilon_{0i_4}b_{i_2i_4}\big(\sum_{l\in\{i_1,\ldots,i_4\}} \frac{\epsilon_{0l}X_{lj_1}}{n}\big)^2 \\
& \times \big(\sum_{l\notin\{i_1,\ldots,i_4\}} \frac{\epsilon_{0l}X_{lj_1}}{n}\big)^{\gamma-3}\big(\sum_{l\notin\{i_1,\ldots,i_4\}} \frac{\epsilon_{0l}X_{lj_2}}{n}\big)^{\gamma-1}\Big] \\
= O(n^{-2}) \sum_{j_1,j_2} \sum_{i_1,\ldots,i_4} & E\Big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_3}b_{i_1i_3}\epsilon_{0i_4}b_{i_2i_4}\big(\sum_{l\in\{i_1,\ldots,i_4\}} \frac{\epsilon_{0l}X_{lj_1}}{n}\big)^2\Big] \\
& \times E\Big[\big(\sum_{l\notin\{i_1,\ldots,i_4\}} \frac{\epsilon_{0l}X_{lj_1}}{n}\big)^{\gamma-3}\big(\sum_{l\notin\{i_1,\ldots,i_4\}} \frac{\epsilon_{0l}X_{lj_2}}{n}\big)^{\gamma-1}\Big] \\
= O(n^{-2}) \sum_{j_1,j_2} \sum_{i_1,\ldots,i_4} & E\Big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_3}b_{i_1i_3}\epsilon_{0i_4}b_{i_2i_4}\big(\sum_{l\in\{i_1,\ldots,i_4\}} \frac{\epsilon_{0l}X_{lj_1}}{n}\big)^2\Big] \times O(n^{-\gamma+2}).
\end{aligned}$$

$(\sum_{l\in\{i_1,\ldots,i_4\}} \epsilon_{0l}X_{lj_1}/n)^2$ and $(\sum_{l\in\{i_1,\ldots,i_4\}} \epsilon_{0l}X_{lj_1}/n) \times (\sum_{l\in\{i_1,\ldots,i_4\}} \epsilon_{0l}X_{lj_2}/n)$ have the same effect to the order. Similar to the discussion of $T_{\gamma 122}$, we have $T_{\gamma 131} = o(pn^{-\gamma})$. Next, we discuss the order of $T_{\gamma 132}$:

$$\begin{aligned}
& T_{\gamma 132} \\
= & \frac{C}{n^2} \sum_{j_1,j_2} \sum_{i_1,\ldots,i_4} E\Big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_3}b_{i_1i_3}\epsilon_{0i_4}b_{i_2i_4}\big(\sum_{l\in\{i_1,\ldots,i_4\}} \frac{\epsilon_{0l}X_{lj_1}}{n}\big)^2 \sum_{l\in\{i_1,\ldots,i_4\}} \frac{\epsilon_{0l}X_{lj_2}}{n} \\
& \qquad \times \big(\sum_{l\notin\{i_1,\ldots,i_4\}} \frac{\epsilon_{0l}X_{lj_1}}{n}\big)^{\gamma-3}\big(\sum_{l\notin\{i_1,\ldots,i_4\}} \frac{\epsilon_{0l}X_{lj_2}}{n}\big)^{\gamma-2}\Big] \\
= & O(n^{-5}) \sum_{j_1,j_2} \sum_{i_1,\ldots,i_4} E\Big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_3}b_{i_1i_3}\epsilon_{0i_4}b_{i_2i_4}\big(\sum_{l\in\{i_1,\ldots,i_4\}} \epsilon_{0l}X_{lj_1}\big)^2 \sum_{l\in\{i_1,\ldots,i_4\}} \epsilon_{0l}X_{lj_2}\Big] \times O(n^{-\gamma+2}).
\end{aligned}$$

Note that $\sum_{j_1,j_2}\sum_{i_1,\ldots,i_4} E[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_3}b_{i_1i_3}\epsilon_{0i_4}b_{i_2i_4}(\sum_{l\in\{i_1,\ldots,i_4\}} \epsilon_{0l}X_{lj_1})^2 \sum_{l\in\{i_1,\ldots,i_4\}} \epsilon_{0l}X_{lj_2}]$ contains 5 ϵ_{0i} and $E[\epsilon_{0i}|\mathbb{X},\mathbb{Z}] = 0$, for a fixed j_1 and j_2. Then we at most have n^2 terms with non-zero expectation. Thus by further noting that $b_{ij} = O(1/n)$, we have

$$T_{\gamma 132} = O(n^{-5}) \times O(p^2) \times O(n^{-\gamma+2}) = O(p^2n^{-\gamma-3}) = o(pn^{-\gamma}).$$

Similarly, we can prove $T_{\gamma 141} = o(pn^{-\gamma})$. For $T_{\gamma 133}$, we have

$$
\begin{aligned}
&T_{\gamma 133}\\
=&\frac{C}{n^2}\sum_{j_1,j_2}\sum_{i_1,\ldots,i_4} E\Big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_3}b_{i_1i_3}\epsilon_{0i_4}b_{i_2i_4}\big(\sum_{l\in\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_1}}{n}\big)^2\big(\sum_{l\in\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_2}}{n}\big)^2\\
&\qquad\qquad\times\big(\sum_{l\notin\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_1}}{n}\big)^{\gamma-3}\big(\sum_{l\notin\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_2}}{n}\big)^{\gamma-3}\Big]\\
=&O(n^{-6})\sum_{j_1,j_2}\sum_{i_1,\ldots,i_4} E\Big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_3}b_{i_1i_3}\epsilon_{0i_4}b_{i_2i_4}\big(\sum_{l\in\{i_1,\ldots,i_4\}}\epsilon_{0l}X_{lj_1}\big)^2\big(\sum_{l\in\{i_1,\ldots,i_4\}}\epsilon_{0l}X_{lj_2}\big)^2\Big]\times O(n^{-\gamma+3}).
\end{aligned}
$$

By noting that $\sum_{i_1,\ldots,i_4} E[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_3}b_{i_1i_3}\epsilon_{0i_4}b_{i_2i_4}(\sum_{l\in\{i_1,\ldots,i_4\}}\epsilon_{0l}X_{lj_1})^2(\sum_{l\in\{i_1,\ldots,i_4\}}\epsilon_{0l}X_{lj_2})^2]$ contains 6 ϵ_{0i} and $E[\epsilon_{0i}|\mathbb{X},\mathbb{Z}] = 0$, for a fixed j_1 and j_2, we at most have n^3 terms with non-zero expectation. Further noting that $b_{ij} = O(1/n)$, we have

$$T_{\gamma 133} = O(n^{-6})\times O(p^2 n)\times O(n^{-\gamma+3}) = O(p^2 n^{-\gamma-2}) = o(pn^{-\gamma}).$$

Similarly, we can prove $T_{\gamma 1k_1k_2} = o(pn^{-\gamma})$ for $k_1 + k_2 = 6$.

For $k_1 + k_2 \geq 7$, we have

$$
\begin{aligned}
&\frac{C}{n^2}\sum_{j_1,j_2}\sum_{i_1,\ldots,i_4} E\Big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_3}b_{i_1i_3}\epsilon_{0i_4}b_{i_2i_4}\\
&\quad\times\big(\sum_{l\in\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_1}}{n}\big)^{k_1-1}\big(\sum_{l\in\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_2}}{n}\big)^{k_2-1}\big(\sum_{l\notin\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_1}}{n}\big)^{\gamma-k_1}\big(\sum_{l\notin\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_2}}{n}\big)^{\gamma-k_2}\Big]\\
=&O(n^{-2})\sum_{j_1,j_2}\sum_{i_1,\ldots,i_4} E\Big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_3}b_{i_1i_3}\epsilon_{0i_4}b_{i_2i_4}\big(\sum_{l\in\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_1}}{n}\big)^{k_1-1}\big(\sum_{l\in\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_2}}{n}\big)^{k_2-1}\Big]\\
&\quad\times E\Big[\big(\sum_{l\notin\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_1}}{n}\big)^{\gamma-k_1}\big(\sum_{l\notin\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_2}}{n}\big)^{\gamma-k_2}\Big]\\
=&O(n^{-2})\sum_{j_1,j_2}\sum_{i_1,\ldots,i_4} E\Big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_3}b_{i_1i_3}\epsilon_{0i_4}b_{i_2i_4}\big(\sum_{l\in\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_1}}{n}\big)^{k_1-1}\big(\sum_{l\in\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_2}}{n}\big)^{k_2-1}\Big]\\
&\quad\times O(n^{-\gamma+\lfloor(k_1+k_2)/2\rfloor})\\
=&O(n^{-2})O(p^2\times n^2\times n^{-(k_1+k_2-2)})\times O(n^{-\gamma+\lfloor(k_1+k_2)/2\rfloor})\\
=&O(pn^{-\gamma})\times O(pn^{-(k_1+k_2-2)+\lfloor(k_1+k_2)/2\rfloor}) = o(pn^{-\gamma}).
\end{aligned}
$$

By noting that $p = o(n^2)$ and for $k_1 + k_2 \geq 7$, $-(k_1+k_2-2) + \lfloor(k_1+k_2)/2\rfloor \geq 2$, we can derive the last equation. In summary, we have $|T_{\gamma 1}| = o_p(n^{-\gamma/2}\sqrt{p})$.

When $1 < v \leq \gamma$, by Minkowski's inequality, we have

$$E[|T_{\gamma v}|] \leq C \sum_{j=1}^{p} E\Big[\Big|\Big(\frac{1}{n}\sum_{i=1}^{n} S_{ij}\Big)^{\gamma-v}\Big(\frac{1}{n}\sum_{i=1}^{n}(\mu_{0i}-\hat{\mu}_{0i})X_{ij}\Big)^{v}\Big|\Big].$$

Then by Cauchy-Schwarz inequality,

$$\begin{aligned} E[|T_{\gamma v}|] &\leq C \sum_{j=1}^{p} E\Big[\big(\frac{1}{n}\sum_{i=1}^{n} S_{ij}\big)^{2(\gamma-v)}\Big]^{1/2} E\Big[\big(\frac{1}{n}\sum_{i=1}^{n}(\mu_{0i}-\hat{\mu}_{0i})X_{ij}\big)^{2v}\Big]^{1/2} \\ &\leq O(n^{-(\gamma-v)/2}) \sum_{j=1}^{p} E\Big[\big(\frac{1}{n}\sum_{i=1}^{n}(\mu_{0i}-\hat{\mu}_{0i})X_{ij}\big)^{2v}\Big]^{1/2}. \end{aligned}$$

Next, we derive the order of $T_{2v,j} = \big(\frac{1}{n}\sum_{i=1}^{n}(\mu_{0i}-\hat{\mu}_{0i})X_{ij}\big)^{2v}$ for any positive integer v. For $j \in P_0$, we have

$$\begin{aligned} E[T_{2v,j}|\mathbb{Z}] &= \frac{1}{n^{2v}} \sum_{i_1,i_2,\ldots,i_{4v}} E[X_{i_1 j} b_{i_1 i_{2v+1}} \epsilon_{0 i_{2v+1}} X_{i_2 j} b_{i_2 i_{2v+2}} \epsilon_{0 i_{2v+1}} \times \cdots \times X_{i_{2v} j} b_{i_{2v} i_{4v}} \epsilon_{0 i_{4v}} |\mathbb{Z}] \\ &\leq \frac{C}{n^{4v}} \sum_{i_1,i_2,\ldots,i_{3v}} E[X_{i_1 j} X_{i_2 j} \times \cdots \times X_{i_{2v} j}^2 \times \epsilon_{0 i_{2v+1}}^2 \epsilon_{0 i_{2v+2}}^2 \times \cdots \times \epsilon_{0 i_{3v}}^2 |\mathbb{Z}] \\ &= O(n^{-v}). \end{aligned}$$

Note that for $j \notin P_0$, $E[X_{ij}|\mathbb{Z}] = 0$. Then for For $j \notin P_0$,

$$\begin{aligned} E[T_{2v,j}|\mathbb{Z}] &= \frac{1}{n^{2v}} \sum_{i_1,i_2,\ldots,i_{4v}} E[X_{i_1 j} b_{i_1 i_{2v+1}} \epsilon_{0 i_{2v+1}} X_{i_2 j} b_{i_2 i_{2v+2}} \epsilon_{0 i_{2v+1}} \times \cdots \times X_{i_{2v} j} b_{i_{2v} i_{4v}} \epsilon_{0 i_{4v}} |\mathbb{Z}] \\ &\leq \frac{C}{n^{4v}} \sum_{i_1,i_2,\ldots,i_{2v}} E[X_{i_1 j}^2 X_{i_2 j}^2 \times \cdots \times X_{i_v j}^2 \times \epsilon_{0 i_{v+1}}^2 \epsilon_{0 i_{v+2}}^2 \times \cdots \times \epsilon_{0 i_{2v}}^2 |\mathbb{Z}] \\ &= O(n^{-2v}). \end{aligned}$$

In summary, $E[T_{2v,j}] = O(n^{-v})$ if $j \in P_0$ and $E[T_{2v,j}] = O(n^{-2v})$ if $j \notin P_0$. Then we have $\sum_{j=1}^{p} E\big[\big(\frac{1}{n}\sum_{i=1}^{n}(\mu_{0i}-\hat{\mu}_{0i})X_{ij}\big)^{2v}\big]^{1/2} = O(p_0 n^{-v/2} + pn^{-v})$. This leads to

$$\begin{aligned} E[|T_{\gamma v}|] &\leq O(n^{-(\gamma-v)/2}) \times O(p_0 n^{-v/2} + pn^{-v}) \\ &= O(p_0 n^{-\gamma/2} + \sqrt{p} n^{-\gamma/2} \sqrt{p} n^{-v/2}). \end{aligned}$$

Note that $p = o(n^2)$, $v \geq 2$, and $p_0 = O(p^{1/2-\delta})$ for a small positive δ, we have $E[|T_{\gamma v}|] =$

$o(\sqrt{p}n^{-\gamma/2})$, leading to $|T_{\gamma v}| = o_p(\sqrt{p}n^{-\gamma/2})$. In summary, we have proved for any finite γ,

$$[\{L(\gamma, \hat{\mu}_0) - \mu(\gamma)\}/\sigma(\gamma)]^{\intercal}_{\gamma\in\Gamma'} = [\{L(\gamma, \mu_0) - \mu(\gamma)\}/\sigma(\gamma)]^{\intercal}_{\gamma\in\Gamma'} + o_p(1).$$

(ii) Define

$$\tilde{V}_{ij} = (Y_i - \hat{\mu}_{0i})X_{ij}/\sqrt{\sigma_{jj}}, \qquad 1 \le i \le n, 1 \le j \le p.$$

Let $\tilde{W}_j = \sum_{i=1}^n \tilde{V}_{ij}/\sqrt{n}$. W_j and $\hat{W}_j$ are defined in the proof of Theorem 1. We discuss two cases: $j \in P_0$ and $j \notin P_0$.

For the first case, we define ϵ is a small constant. Note that

$$\begin{aligned}
&Pr\big(\max_{j\in P_0} \tilde{W}_j^2 > \epsilon \log p\big)\\
\le &Pr\big(\max_{j\in P_0} |\tilde{W}_j| > \epsilon(\log p)^{1/2}\big)\\
\le &Pr\Big(\max_{j\in P_0} \Big|\frac{\sum_{i=1}^n (Y_i - \mu_{0i} + \mu_{0i} - \hat{\mu}_{0i})X_{ij}}{\sqrt{\sigma_{jj}n}}\Big| > (\epsilon \log p)^{1/2}\Big)\\
\le &Pr\Big(\max_{j\in P_0} \Big|\frac{\sum_{i=1}^n (Y_i - \mu_{0i})X_{ij}}{\sqrt{\sigma_{jj}n}}\Big| > \frac{(\epsilon \log p)^{1/2}}{2}\Big) + Pr\Big(\max_{j\in P_0} \Big|\frac{\sum_{i=1}^n (\mu_{0i} - \hat{\mu}_{0i})X_{ij}}{\sqrt{\sigma_{jj}n}}\Big| > \frac{(\epsilon \log p)^{1/2}}{2}\Big).
\end{aligned}$$

For the first term, we have

$$Pr\Big(\max_{j\in P_0} \Big|\frac{\sum_{i=1}^n (Y_i - \mu_{0i})X_{ij}}{\sqrt{\sigma_{jj}n}}\Big| > \frac{(\epsilon \log p)^{1/2}}{2}\Big) \le p_0 Pr\Big(\Big|\frac{\sum_{i=1}^n S_{ij}}{\sqrt{\sigma_{jj}n}}\Big| > \frac{(\epsilon \log p)^{1/2}}{2}\Big).$$

Note that S_{ij} follows a sub-Gaussian distribution (C5) and $S_{i_1 j}$ and $S_{i_2 j}$ are independent for $i_1 \neq i_2$. Then using a Chernoff bound, we have

$$\begin{aligned}
&Pr\Big(\max_{j\in P_0} \Big|\frac{\sum_{i=1}^n (Y_i - \mu_{0i})X_{ij}}{\sqrt{\sigma_{jj}n}}\Big| > \frac{(\epsilon \log p)^{1/2}}{2}\Big)\\
\le &p_0 \times 2\exp\Big(-\frac{\epsilon \log p/4}{2}\Big) = 2p_0 p^{-\epsilon/8} = o(1).
\end{aligned}$$

By noting that $p_0 = p^{\eta}$, where η is a small constant, we have the last equation.

Remark: Suppose $S_{1j}, \ldots, S_{nj}$ be n independent random variables such that S_{ij} follows sub-Gaussian distribution subG$(0, \sigma^2)$. Then for any $a \in \mathbb{R}^n$, using a Chernoff bound, we have $Pr\left(|\sum_{i=1}^n a_i S_{ij}| > t\right) \le 2\exp\left(-t^2/(2\sigma^2||a||_2^2)\right)$.

For the second term, we have

$$
\begin{aligned}
&Pr\Big(\max_{j\in P_0}\Big|\frac{\sum_{i=1}^{n}(\mu_{0i}-\hat{\mu}_{0i})X_{ij}}{\sqrt{\sigma_{jj}n}}\Big|>\frac{(\epsilon\log p)^{1/2}}{2}\Big)\\
\le &\operatorname{Pr}\Big(\max_{j\in P_0}\Big|\frac{\sum_{i_1,i_2}X_{i_1j}\epsilon_{0i_2}b_{i_1i_2}}{\sqrt{\sigma_{jj}n}}\Big|>\frac{(\epsilon\log p)^{1/2}}{2}\Big)\\
\le &Pr\Big(\max_{j\in P_0}\big|\sum_{i_2=1}^{n}\epsilon_{0i_2}\big(\sum_{i_1}X_{i_1j}b_{i_1i_2}/\sqrt{\sigma_{jj}n}\big)\big|>\frac{(\epsilon\log p)^{1/2}}{2}\big|\max_{i_2}\sum_{i_1}X_{i_1j}b_{i_1i_2}/\sqrt{\sigma_{jj}n}<\frac{C}{\sqrt{\sigma_{jj}n}}\Big)\\
&+Pr\Big(\max_{i_2}\sum_{i_1}X_{i_1j}b_{i_1i_2}/\sqrt{\sigma_{jj}n}\ge\frac{C}{\sqrt{\sigma_{jj}n}}\Big).
\end{aligned}
$$

We discuss these two terms separately. For the first term, we have

$$
\begin{aligned}
&Pr\Big(\max_{j\in P_0}\big|\sum_{i_2=1}^{n}\epsilon_{0i_2}\big(\sum_{i_1}X_{i_1j}b_{i_1i_2}/\sqrt{\sigma_{jj}n}\big)\big|>\frac{(\epsilon\log p)^{1/2}}{2}\big|\max_{i_2}\sum_{i_1}X_{i_1j}b_{i_1i_2}/\sqrt{\sigma_{jj}n}<\frac{C}{\sqrt{\sigma_{jj}n}}\Big)\\
\le &p_0Pr\Big(\big|\sum_{i_2=1}^{n}\epsilon_{0i_2}\big(\sum_{i_1}X_{i_1j}b_{i_1i_2}/\sqrt{\sigma_{jj}n}\big)\big|>\frac{(\epsilon\log p)^{1/2}}{2}\big|\max_{i_2}\sum_{i_1}X_{i_1j}b_{i_1i_2}/\sqrt{\sigma_{jj}n}<\frac{C}{\sqrt{\sigma_{jj}n}}\Big)\\
\le &p_0E\Big[Pr\Big(\big|\sum_{i_2=1}^{n}\epsilon_{0i_2}\big(\sum_{i_1}X_{i_1j}b_{i_1i_2}/\sqrt{\sigma_{jj}n}\big)\big|>\frac{(\epsilon\log p)^{1/2}}{2}\big|\max_{i_2}\sum_{i_1}X_{i_1j}b_{i_1i_2}/\sqrt{\sigma_{jj}n}<\frac{C}{\sqrt{\sigma_{jj}n}},\mathbb{X},\mathbb{Z}\Big)\Big].
\end{aligned}
$$

Noting that ϵ_{0i} follows a sub-Gaussian distribution, we have

$$
\begin{aligned}
&Pr\Big(\max_{j\in P_0}\big|\sum_{i_2=1}^{n}\epsilon_{0i_2}\big(\sum_{i_1}X_{i_1j}b_{i_1i_2}/\sqrt{\sigma_{jj}n}\big)\big|>\frac{(\epsilon\log p)^{1/2}}{2}\big|\max_{i_2}\sum_{i_1}X_{i_1j}b_{i_1i_2}/\sqrt{\sigma_{jj}n}<\frac{C}{\sqrt{\sigma_{jj}n}},\mathbb{X},\mathbb{Z}\Big)\\
\le &p_0\times 2\exp\Big(-\frac{\epsilon\log p/4}{2C^2}\Big)=2p_0p^{-\epsilon/(8C^2)}=o(1).
\end{aligned}
$$

For the second term, we have

$$
\begin{aligned}
&Pr\Big(\max_{i_2}\sum_{i_1}X_{i_1j}b_{i_1i_2}/\sqrt{\sigma_{jj}n}\ge\frac{C}{\sqrt{\sigma_{jj}n}}\Big)\\
\le &E\Big[Pr\Big(\frac{\sum_{i_1}|X_{i_1j}|}{n}\ge C\big|\mathbb{Z}\Big)\Big]\\
\le &E\Big[Pr\Big(\frac{\sum_{i_1}|X_{i_1j}|-E[|X_{i_1j}|]}{n}\ge C-E[|X_{i_1j}|]\big|\mathbb{Z}\Big)\Big]=o(1).
\end{aligned}
$$

In summary, as $n,p\to\infty$, $Pr\big(\max_{j\in P_0}\tilde{W}_j^2>\epsilon\log p\big)=o(1)$. Then we focus on the second situation. Note that

$$
\begin{aligned}
&Pr\big(\max_{j\notin P_0}|\tilde{W}_j - W_j| \geq \frac{1}{\log p}\big)\\
&\leq np\max_{j\notin P_0} Pr(|V_{1j}| \geq \tau_n) + Pr\Big(\max_{j\notin P_0}\Big|\sum_{i=1}^{n}\frac{(\mu_{0i}-\hat{\mu}_{0i})X_{ij}}{\sqrt{\sigma_{jj}n}}\Big| \geq \frac{1}{\log p}\Big).
\end{aligned}
$$

From the proof of Theorem 1, the first term is $O(1/p+1/n)$ and thus we only need discuss the second term. By the Markov inequality and the Jensen's inequality,

$$
\begin{aligned}
&Pr\Big(\max_{j\notin P_0}\Big|\sum_{i=1}^{n}\frac{(\mu_{0i}-\hat{\mu}_{0i})X_{ij}}{\sqrt{\sigma_{jj}n}}\Big| \geq \frac{1}{\log p}\Big)\\
&\leq Pr\Big(\max_{j\notin P_0}\big(\sum_{i=1}^{n}\frac{(\mu_{0i}-\hat{\mu}_{0i})X_{ij}}{\sqrt{\sigma_{jj}n}}\big)^{16} \geq \frac{1}{(\log p)^{16}}\Big)\\
&\leq pPr\Big(\big(\sum_{i=1}^{n}\frac{(\mu_{0i}-\hat{\mu}_{0i})X_{ij}}{\sqrt{\sigma_{jj}n}}\big)^{16} \geq \frac{1}{(\log p)^{16}}\Big)\\
&\leq p\log pE\Big[\big(\sum_{i=1}^{n}\frac{(\mu_{0i}-\hat{\mu}_{0i})X_{ij}}{\sqrt{\sigma_{jj}n}}\big)^{16}\Big]\\
&\leq p\log p \times O(n^{-8}) = o(1).
\end{aligned}
$$

Thus, we have $Pr\big(\max_{1\leq j\leq p}|\tilde{W}_j - W_j| \geq 1/\log p\big) = o(1)$ as $n, p \to \infty$. Further note that

$$
|\max_{j\notin P_0} W_j^2 - \max_{j\notin P_0}\tilde{W}_j^2| \leq 2\max_{j\notin P_0}|W_j|\max_{j\notin P_0}|W_j-\tilde{W}_j| + \max_{j\notin P_0}|W_j-\tilde{W}_j|^2.
$$

The above two inequalities indicate that when $n, p \to \infty$, $|\max_{j\notin P_0} W_j^2 - \max_{j\notin P_0}\tilde{W}_j^2| \to 0$. By Cai et al. (2014), we have

$$
Pr\big(\max_{j\notin P_0}\tilde{W}_j^2 - 2\log p + \log\log p \leq x\big) \to \exp\{-\pi^{-1/2}\exp(-x/2)\}.
$$

Note that

$$
\max_{1\leq j\leq p}\tilde{W}_j^2 = \max\big(\max_{j\in P_0}\tilde{W}_j^2, \max_{j\notin P_0}\tilde{W}_j^2\big) = \max_{j\notin P_0}\tilde{W}_j^2.
$$

Thus,

$$
Pr\big(\max_{1\leq j\leq p}\tilde{W}_j^2 - 2\log p + \log\log p \leq x\big) \to \exp\{-\pi^{-1/2}\exp(-x/2)\}.
$$

(iii) By proof in (i) and (ii), we have

$$[\{L(\gamma, \hat{\mu}_0) - \mu(\gamma)\}/\sigma(\gamma)]^{\intercal}_{\gamma \in \Gamma'} = [\{L(\gamma, \mu_0) - \mu(\gamma)\}/\sigma(\gamma)]^{\intercal}_{\gamma \in \Gamma'} + o_p(1)$$

and $L(\infty, \hat{\mu}_0) = L(\infty, \mu_0) + o_p(1)$. By Theorem 1, $[\{L(\gamma, \mu_0) - \mu(\gamma)\}/\sigma(\gamma)]^{\intercal}_{\gamma \in \Gamma'}$ is asymptotically independent with $L(\infty, \mu_0)$. Note that $o_p(1)$ is asymptotic independent with $L(\infty, \mu_0)$ and $[\{L(\gamma, \mu_0) - \mu(\gamma)\}/\sigma(\gamma)]^{\intercal}_{\gamma \in \Gamma'}$, thus $[\{L(\gamma, \hat{\mu}_0) - \mu(\gamma)\}/\sigma(\gamma)]^{\intercal}_{\gamma \in \Gamma'}$ is asymptotically independent with $L(\infty, \hat{\mu}_0)$.

This completes the proof. ■

2 Supplementary Tables and Figures

We put extensive simulation results in this section and the simulation settings are described in the subsection 4.1.

Table S1: Empirical type I errors and powers (%) of various tests in a simulation with $n = 200$, $p = 4000$. The sparsity parameter was $s = 0.1$, leading to 400 non-zero elements in β with a constant value c. The results outside and inside parentheses were calculated from asymptotics and parametric bootstrap based method, respectively.

c	0	0.01	0.03	0.05	0.07	0.1	0.15
SPU(1)	5 (5)	12 (11)	51 (51)	75 (76)	85 (85)	90 (90)	93 (93)
SPU(2)	7 (5)	9 (7)	26 (21)	48 (41)	60 (54)	71 (65)	79 (73)
SPU(3)	4 (4)	8 (9)	45 (47)	69 (70)	81 (82)	87 (88)	92 (92)
SPU(4)	2 (5)	4 (6)	14 (17)	34 (37)	47 (50)	60 (61)	67 (68)
SPU(5)	2 (4)	5 (7)	26 (30)	46 (51)	59 (62)	68 (70)	74 (77)
SPU(6)	2 (5)	1 (5)	7 (12)	18 (26)	29 (35)	37 (44)	46 (52)
SPU(infty)	5 (4)	5 (5)	6 (7)	9 (11)	10 (12)	12 (15)	15 (20)
aSPU	4 (4)	7 (7)	38 (43)	66 (69)	80 (82)	87 (89)	93 (94)
HDGLM	7 (5)	9 (7)	25 (21)	48 (41)	59 (53)	70 (64)	78 (72)
GT	5	7	21	42	54	65	73

Table S2: Empirical type I errors and powers (%) of various tests in a simulation with $n = 200$, $p = 4000$. The sparsity parameter was $s = 0.05$, leading to 200 non-zero elements in β with a constant value c. The results outside and inside parentheses were calculated from asymptotics and parametric bootstrap based method, respectively.

c	0	0.03	0.06	0.1	0.15	0.2	0.3
SPU(1)	5 (5)	18 (18)	40 (39)	55 (55)	65 (65)	67 (68)	70 (70)
SPU(2)	7 (5)	14 (11)	38 (32)	63 (56)	74 (67)	78 (72)	81 (75)
SPU(3)	4 (4)	16 (17)	38 (40)	61 (62)	76 (76)	80 (81)	82 (83)
SPU(4)	2 (5)	8 (12)	26 (31)	52 (54)	66 (67)	72 (72)	78 (79)
SPU(5)	2 (4)	8 (11)	28 (32)	50 (53)	66 (69)	71 (73)	74 (76)
SPU(6)	2 (5)	4 (10)	15 (24)	35 (43)	49 (54)	56 (61)	60 (65)
SPU(∞)	5 (4)	6 (6)	11 (12)	14 (17)	17 (23)	19 (25)	20 (27)
aSPU	4 (4)	14 (15)	37 (42)	64 (67)	80 (82)	84 (86)	88 (88)
HDGLM	7 (5)	14 (11)	37 (32)	62 (56)	72 (67)	77 (72)	80 (77)
GT	5	11	32	57	68	72	76

Table S3: Empirical type I errors and powers (%) of various tests in a simulation with $n = 200$, $p = 2000$. The sparsity parameter was $s = 0.001$, leading to 2 non-zero elements in β with a constant value c. The results outside and inside parentheses were calculated from asymptotics and parametric bootstrap based method, respectively.

c	0	0.1	0.2	0.3	0.4	0.5	0.6
SPU(1)	5 (5)	6 (5)	6 (6)	6 (6)	5 (5)	6 (6)	6 (5)
SPU(2)	6 (5)	8 (6)	11 (9)	13 (10)	16 (14)	21 (17)	28 (23)
SPU(3)	4 (5)	5 (6)	6 (7)	9 (9)	15 (15)	26 (28)	44 (45)
SPU(4)	4 (6)	4 (6)	8 (10)	18 (21)	43 (46)	73 (76)	91 (92)
SPU(5)	4 (5)	4 (6)	6 (9)	20 (24)	53 (56)	83 (85)	96 (96)
SPU(6)	3 (6)	3 (6)	7 (12)	26 (32)	64 (70)	91 (93)	98 (99)
SPU(∞)	5 (5)	5 (5)	11 (10)	33 (33)	75 (75)	96 (96)	100 (100)
aSPU	5 (5)	5 (6)	10 (10)	31 (29)	70 (69)	93 (93)	99 (99)
HDGLM	7 (5)	8 (7)	11 (9)	13 (10)	16 (14)	21 (17)	28 (23)
GT	5	6	9	10	14	18	24

Table S4: Empirical type I errors and powers (%) of various tests in a simulation with $n = 200$, $p = 2000$. The sparsity parameter was $s = 0.05$, leading to 100 non-zero elements in β with a constant value c. The results outside and inside parentheses were calculated from asymptotics and parametric bootstrap based method, respectively.

c	0	0.03	0.06	0.1	0.15	0.2	0.3
SPU(1)	5 (5)	12 (12)	27 (27)	47 (47)	58 (58)	64 (65)	68 (68)
SPU(2)	6 (5)	12 (9)	34 (29)	65 (61)	85 (82)	91 (89)	96 (94)
SPU(3)	4 (5)	11 (12)	32 (33)	66 (67)	84 (84)	90 (90)	95 (95)
SPU(4)	4 (6)	7 (9)	26 (29)	62 (63)	85 (86)	93 (93)	97 (97)
SPU(5)	4 (5)	7 (10)	24 (27)	60 (64)	82 (84)	90 (91)	94 (94)
SPU(6)	3 (6)	4 (7)	16 (22)	44 (50)	73 (77)	85 (87)	92 (93)
SPU(∞)	5 (5)	5 (5)	9 (9)	18 (21)	27 (31)	36 (42)	42 (50)
aSPU	5 (5)	9 (10)	33 (37)	67 (70)	89 (90)	96 (96)	99 (99)
HDGLM	7 (5)	12 (10)	35 (29)	65 (60)	85 (82)	91 (89)	96 (94)
GT	5	10	29	61	82	89	94

Table S5: Empirical type I errors and powers (%) of various tests in a simulation with $n = 200$, $p = 2000$. The sparsity parameter was $s = 0.7$, leading to 1400 non-zero elements in β with a constant value c. The results outside and inside parentheses were calculated from asymptotics and parametric bootstrap based method, respectively.

c	0	0.001	0.003	0.005	0.007	0.01	0.015
SPU(1)	5 (5)	6 (5)	19 (19)	47 (46)	73 (72)	95 (95)	100 (100)
SPU(2)	6 (6)	6 (6)	5 (5)	7 (7)	8 (8)	14 (14)	24 (23)
SPU(3)	4 (4)	5 (5)	15 (15)	36 (36)	59 (58)	87 (87)	99 (99)
SPU(4)	6 (6)	7 (6)	6 (5)	7 (6)	8 (7)	12 (11)	23 (20)
SPU(5)	5 (5)	5 (5)	9 (9)	19 (19)	34 (33)	56 (55)	83 (82)
SPU(6)	6 (5)	7 (6)	6 (5)	7 (5)	9 (7)	11 (9)	16 (14)
SPU(∞)	3 (4)	3 (6)	3 (4)	3 (4)	4 (5)	5 (7)	6 (7)
aSPU	5 (5)	7 (6)	11 (12)	27 (29)	49 (53)	85 (88)	99 (99)
HDGLM	7 (5)	7 (6)	7 (5)	9 (7)	12 (9)	19 (15)	29 (26)
GT	5	6	5	7	9	15	26

Table S6: Empirical type I errors and powers (%) of various tests in a simulation with $n = 200$, $p = 2000$. The sparsity parameter was $s = 0.001$, leading to 2 non-zero elements in β with a constant value c. The covariance is Non-sparse structure. The results outside and inside parentheses were calculated from asymptotics and parametric bootstrap based method, respectively.

c	0	0.1	0.2	0.3	0.4	0.5	0.6
SPU(1)	6 (5)	4 (4)	5 (5)	5 (5)	5 (5)	5 (6)	6 (5)
SPU(2)	6 (5)	7 (5)	14 (11)	19 (16)	28 (24)	37 (33)	42 (39)
SPU(3)	4 (5)	5 (6)	10 (11)	28 (29)	45 (47)	60 (61)	68 (68)
SPU(4)	4 (5)	5 (6)	24 (26)	52 (52)	68 (69)	78 (78)	83 (83)
SPU(5)	4 (5)	5 (7)	25 (27)	54 (55)	71 (73)	80 (81)	85 (86)
SPU(6)	3 (5)	5 (7)	29 (32)	58 (59)	73 (75)	83 (84)	88 (88)
SPU(∞)	4 (4)	6 (5)	41 (41)	81 (81)	95 (96)	99 (99)	100 (100)
aSPU	5 (5)	7 (6)	37 (35)	76 (74)	94 (93)	98 (98)	100 (100)
HDGLM	7 (5)	7 (5)	14 (11)	20 (16)	28 (24)	37 (33)	43 (39)
GT	5	5	11	16	25	34	40

Table S7: Empirical type I errors and powers (%) of various tests in a simulation with $n = 200$, $p = 2000$. The sparsity parameter was $s = 0.05$, leading to 100 non-zero elements in β with a constant value c. The covariance is Non-sparse structure. The results outside and inside parentheses were calculated from asymptotics and parametric bootstrap based method, respectively.

c	0	0.03	0.06	0.1	0.15	0.2	0.3
SPU(1)	6 (5)	18 (17)	41 (40)	55 (55)	62 (63)	67 (67)	69 (69)
SPU(2)	6 (5)	17 (15)	51 (46)	77 (73)	90 (87)	92 (89)	94 (92)
SPU(3)	4 (5)	17 (19)	45 (47)	77 (78)	88 (88)	92 (92)	94 (94)
SPU(4)	4 (5)	12 (13)	43 (43)	75 (75)	90 (89)	92 (92)	94 (94)
SPU(5)	4 (5)	10 (12)	35 (37)	67 (69)	82 (82)	87 (87)	90 (90)
SPU(6)	3 (5)	8 (10)	29 (33)	59 (61)	75 (76)	81 (81)	85 (86)
SPU(∞)	4 (4)	7 (6)	14 (16)	21 (28)	33 (40)	41 (48)	45 (53)
aSPU	5 (5)	17 (18)	52 (53)	82 (84)	92 (92)	96 (95)	96 (96)
HDGLM	7 (5)	17 (15)	51 (46)	78 (73)	90 (87)	92 (89)	94 (92)
GT	5	15	46	74	87	89	92

Table S8: Empirical type I errors and powers (%) of various tests in a simulation with $n = 200$, $p = 2000$. The sparsity parameter was $s = 0.1$, leading to 200 non-zero elements in β with a constant value c. The covariance is Non-sparse structure. The results outside and inside parentheses were calculated from asymptotics and parametric bootstrap based method, respectively.

c	0	0.03	0.06	0.1	0.15	0.2	0.3
SPU(1)	6 (5)	12 (12)	51 (50)	75 (75)	85 (85)	89 (89)	93 (92)
SPU(2)	6 (5)	8 (6)	32 (28)	59 (54)	78 (74)	88 (85)	93 (90)
SPU(3)	4 (5)	8 (9)	41 (44)	73 (74)	87 (87)	91 (91)	95 (95)
SPU(4)	4 (5)	4 (5)	22 (24)	47 (48)	67 (66)	80 (79)	88 (87)
SPU(5)	4 (5)	5 (6)	21 (25)	46 (50)	63 (65)	75 (77)	83 (84)
SPU(6)	3 (5)	3 (5)	11 (15)	30 (34)	45 (47)	56 (59)	65 (67)
SPU(∞)	4 (4)	4 (4)	10 (12)	15 (16)	17 (19)	21 (26)	25 (31)
aSPU	5 (5)	7 (8)	42 (44)	72 (75)	86 (89)	94 (94)	97 (97)
HDGLM	7 (5)	8 (6)	32 (28)	59 (54)	78 (74)	88 (85)	92 (90)
GT	5	6	28	54	74	86	91

Table S9: Empirical type I errors and powers (%) of various tests in a simulation with $n = 200$, $p = 2000$. The sparsity parameter was $s = 0.7$, leading to 1400 non-zero elements in β with a constant value c. The covariance is Non-sparse structure. The results outside and inside parentheses were calculated from asymptotics and parametric bootstrap based method, respectively.

c	0	0.001	0.003	0.005	0.007	0.01	0.015
SPU(1)	6 (5)	8 (7)	35 (36)	78 (78)	96 (96)	100 (100)	100 (100)
SPU(2)	6 (5)	6 (5)	6 (5)	9 (8)	13 (11)	24 (20)	48 (43)
SPU(3)	4 (5)	6 (6)	22 (23)	53 (55)	80 (81)	97 (97)	100 (100)
SPU(4)	4 (5)	4 (4)	4 (5)	5 (7)	9 (10)	17 (18)	31 (32)
SPU(5)	4 (5)	4 (6)	9 (11)	19 (23)	32 (36)	58 (64)	85 (88)
SPU(6)	3 (5)	3 (4)	3 (4)	3 (5)	6 (8)	10 (13)	16 (19)
SPU(∞)	4 (4)	4 (3)	4 (5)	6 (5)	5 (5)	7 (8)	9 (11)
aSPU	5 (5)	5 (5)	19 (22)	54 (57)	84 (88)	98 (99)	100 (100)
HDGLM	7 (5)	6 (5)	7 (4)	10 (8)	13 (10)	24 (20)	48 (43)
GT	5	5	5	8	11	20	43

Table S10: Empirical type I errors and powers (%) of various tests in a simulation with $n = 200$, $p = 2000$. The sparsity parameter was $s = 0.001$, leading to 2 non-zero elements in β with a constant value c. The covariance is Block diagonal structure. The results outside and inside parentheses were calculated from asymptotics and parametric bootstrap based method, respectively.

c	0	0.1	0.2	0.3	0.4	0.5	0.6
SPU(1)	5 (5)	4 (4)	4 (4)	4 (4)	5 (5)	6 (6)	5 (5)
SPU(2)	6 (5)	8 (7)	14 (12)	24 (21)	34 (29)	42 (38)	45 (42)
SPU(3)	4 (5)	5 (7)	18 (19)	38 (40)	56 (57)	67 (68)	75 (75)
SPU(4)	4 (5)	7 (8)	33 (33)	61 (61)	74 (75)	83 (83)	88 (89)
SPU(5)	4 (5)	7 (9)	35 (36)	63 (64)	78 (79)	85 (86)	90 (90)
SPU(6)	3 (5)	7 (10)	40 (43)	67 (68)	80 (81)	87 (88)	92 (92)
SPU(∞)	4 (4)	8 (7)	55 (56)	89 (89)	98 (98)	100 (100)	100 (100)
aSPU	5 (5)	9 (9)	51 (49)	87 (85)	97 (97)	100 (100)	100 (100)
HDGLM	6 (5)	8 (6)	15 (12)	25 (21)	34 (30)	43 (39)	45 (42)
GT	5	7	12	21	30	39	42

Table S11: Empirical type I errors and powers (%) of various tests in a simulation with $n = 200$, $p = 2000$. The sparsity parameter was $s = 0.05$, leading to 100 non-zero elements in β with a constant value c. The covariance is Block diagonal structure. The results outside and inside parentheses were calculated from asymptotics and parametric bootstrap based method, respectively.

c	0	0.03	0.06	0.1	0.15	0.2	0.3
SPU(1)	5 (5)	17 (17)	37 (36)	57 (57)	62 (62)	68 (68)	70 (70)
SPU(2)	6 (5)	16 (13)	39 (35)	65 (60)	78 (74)	85 (83)	87 (84)
SPU(3)	4 (5)	13 (14)	43 (43)	70 (72)	82 (82)	87 (87)	89 (89)
SPU(4)	4 (5)	10 (11)	31 (32)	61 (62)	78 (77)	84 (83)	88 (87)
SPU(5)	4 (5)	8 (10)	29 (31)	55 (57)	72 (74)	79 (80)	82 (83)
SPU(6)	3 (5)	8 (10)	20 (24)	46 (47)	62 (63)	71 (71)	75 (76)
SPU(∞)	4 (4)	7 (7)	10 (13)	19 (24)	26 (31)	30 (40)	34 (44)
aSPU	5 (5)	15 (15)	42 (44)	71 (72)	85 (85)	90 (91)	93 (93)
HDGLM	6 (5)	16 (14)	39 (35)	65 (61)	78 (74)	85 (82)	87 (84)
GT	5	14	35	60	75	83	84

Table S12: Empirical type I errors and powers (%) of various tests in a simulation with $n = 200$, $p = 2000$. The sparsity parameter was $s = 0.1$, leading to 200 non-zero elements in β with a constant value c. The covariance is Block diagonal structure. The results outside and inside parentheses were calculated from asymptotics and parametric bootstrap based method, respectively.

c	0	0.01	0.03	0.05	0.07	0.1	0.15
SPU(1)	5 (5)	11 (11)	47 (46)	71 (71)	83 (83)	89 (89)	92 (93)
SPU(2)	6 (5)	8 (6)	25 (22)	47 (43)	64 (60)	78 (74)	85 (82)
SPU(3)	4 (5)	8 (9)	35 (36)	63 (65)	80 (80)	87 (87)	91 (91)
SPU(4)	4 (5)	5 (6)	17 (18)	36 (36)	53 (54)	69 (69)	78 (77)
SPU(5)	4 (5)	5 (7)	16 (20)	36 (39)	50 (54)	63 (65)	70 (72)
SPU(6)	3 (5)	3 (5)	10 (13)	21 (24)	33 (36)	47 (48)	53 (54)
SPU(∞)	4 (4)	5 (6)	8 (9)	10 (12)	14 (17)	16 (21)	19 (24)
aSPU	5 (5)	8 (7)	34 (37)	64 (66)	82 (84)	91 (93)	94 (95)
HDGLM	6 (5)	8 (6)	26 (22)	48 (43)	65 (59)	78 (74)	85 (82)
GT	5	6	22	43	60	74	83

Table S13: Empirical type I errors and powers (%) of various tests in a simulation with $n = 200$, $p = 2000$. The sparsity parameter was $s = 0.7$, leading to 1400 non-zero elements in β with a constant value c. The covariance is Block diagonal structure. The results outside and inside parentheses were calculated from asymptotics and parametric bootstrap based method, respectively.

c	0	0.001	0.003	0.005	0.007	0.01	0.015
SPU(1)	5 (5)	6 (6)	35 (35)	72 (72)	93 (94)	100 (100)	100 (100)
SPU(2)	6 (5)	5 (5)	6 (4)	8 (7)	11 (9)	21 (17)	37 (33)
SPU(3)	4 (5)	6 (6)	18 (20)	45 (47)	71 (72)	95 (95)	100 (100)
SPU(4)	4 (5)	4 (5)	4 (5)	4 (5)	8 (9)	13 (14)	25 (26)
SPU(5)	4 (5)	4 (6)	8 (10)	15 (19)	27 (32)	48 (55)	75 (79)
SPU(6)	3 (5)	3 (5)	3 (5)	4 (6)	5 (7)	8 (11)	14 (17)
SPU(∞)	4 (4)	4 (4)	4 (4)	4 (4)	5 (5)	6 (7)	7 (9)
aSPU	5 (5)	4 (4)	17 (21)	47 (53)	80 (83)	98 (99)	100 (100)
HDGLM	6 (5)	6 (4)	6 (4)	8 (7)	11 (10)	21 (17)	37 (33)
GT	5	4	5	7	10	17	33

Table S14: Empirical type I errors and powers (%) of various tests in a simulation with $n = 200$, $p = 2000$. The sparsity parameter was $s = 0.1$, leading to 200 non-zero elements in β with a constant value c. The covariance is Block diagonal structure. The results outside and inside parentheses were calculated from asymptotics and parametric bootstrap based method, respectively. The outcome is continuous.

c	0	0.01	0.03	0.05	0.07	0.1	0.15
SPU(1)	5 (5)	11 (11)	47 (46)	71 (71)	83 (83)	89 (89)	92 (93)
SPU(2)	6 (5)	8 (6)	25 (22)	47 (43)	64 (60)	78 (74)	85 (82)
SPU(3)	4 (5)	8 (9)	35 (36)	63 (65)	80 (80)	87 (87)	91 (91)
SPU(4)	4 (5)	5 (6)	17 (18)	36 (36)	53 (54)	69 (69)	78 (77)
SPU(5)	4 (5)	5 (7)	16 (20)	36 (39)	50 (54)	63 (65)	70 (72)
SPU(6)	3 (5)	3 (5)	10 (13)	21 (24)	33 (36)	47 (48)	53 (54)
SPU(∞)	4 (4)	5 (6)	8 (9)	10 (12)	14 (17)	16 (21)	19 (24)
aSPU	5 (5)	8 (7)	34 (37)	64 (66)	82 (84)	91 (93)	94 (95)
HDGLM	6 (5)	8 (6)	26 (22)	48 (43)	65 (59)	78 (74)	85 (82)
GT	5	6	22	43	60	74	83

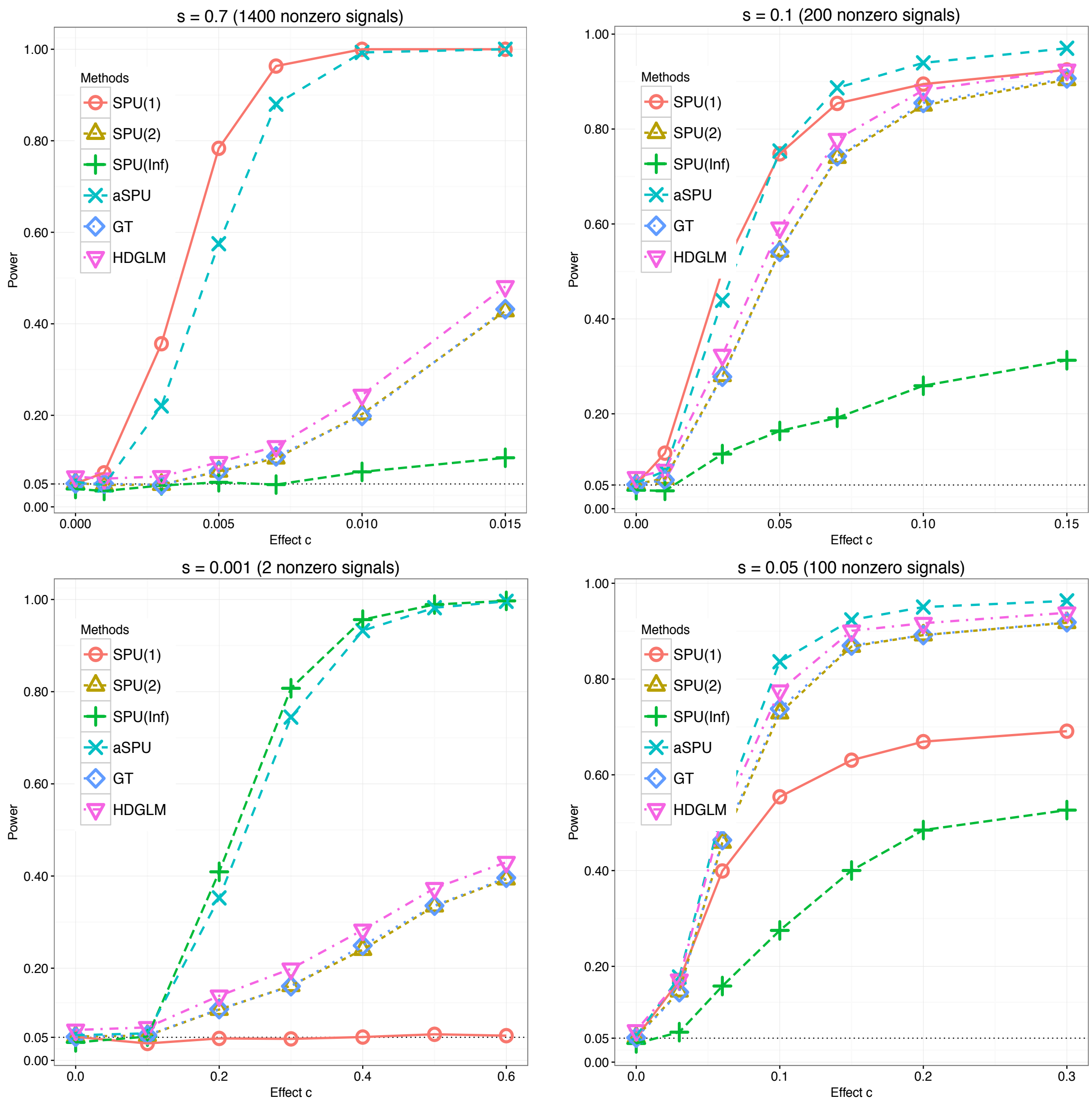

Figure S1: Empirical powers of SPU(1), SPU(2), SPU(∞), aSPU, GT (Goeman et al., 2011), and HDGLM (Guo and Chen, 2016). The signal sparsity parameter s varies from 0.001 to 0.7. We set $n = 200$ and $p = 2000$, respectively. The covariance matrix structure is block diagonal structure.

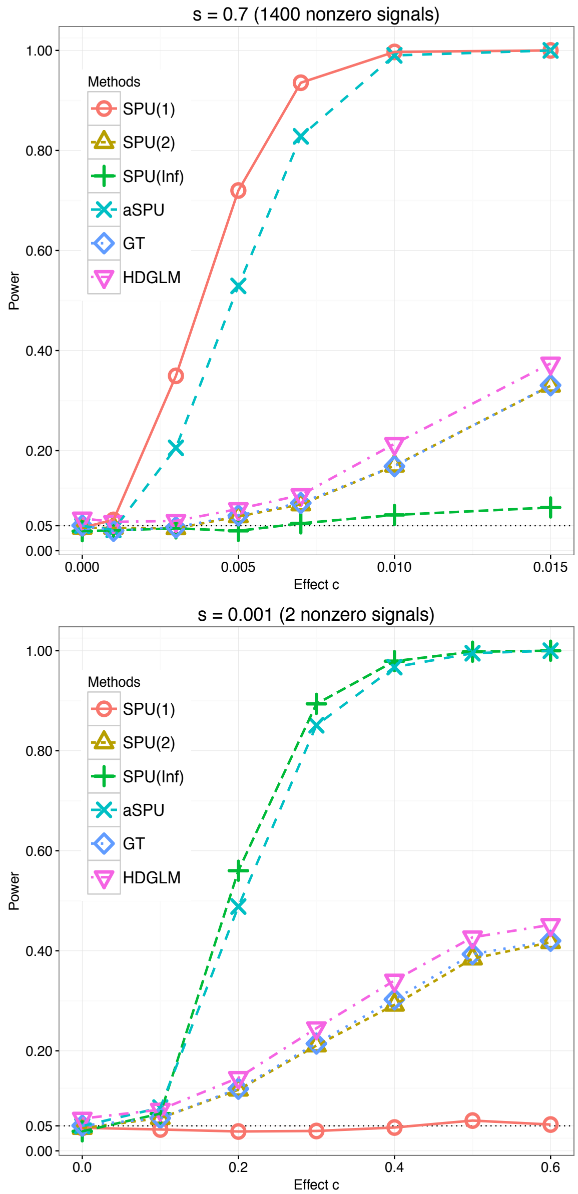

Figure S2: Empirical powers of SPU(1), SPU(2), SPU(∞), aSPU, GT (Goeman et al., 2011), and HDGLM (Guo and Chen, 2016). The signal sparsity parameter s varies from 0.001 to 0.7. We set $n = 200$ and $p = 2000$, respectively. The covariance matrix structure is non-sparse structure.

References

Cai, T. T., W. Liu, and Y. Xia (2014). Two-sample test of high dimensional means under dependence. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 76*(2), 349–372.

Goeman, J. J., H. C. Van Houwelingen, and L. Finos (2011). Testing against a high-dimensional alternative in the generalized linear model: asymptotic type i error control. *Biometrika 98*(2), 381–390.

Guo, B. and S. X. Chen (2016). Tests for high dimensional generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 78*(5), 1079–1102.

Guyon, X. (1995). *Random fields on a network: modeling, statistics, and applications.* Springer Science & Business Media.

Hsing, T. (1995). A note on the asymptotic independence of the sum and maximum of strongly mixing stationary random variables. *The Annals of Probability*, 938–947.

Ibragimov, I. A. (1971). Independent and stationary sequences of random variables.

Kim, T. Y. (1994). Moment bounds for non-stationary dependent sequences. *Journal of Applied Probability*, 731–742.

Le Cessie, S. and J. Van Houwelingen (1991). A goodness-of-fit test for binary regression models, based on smoothing methods. *Biometrics*, 1267–1282.

Xu, G., L. Lin, P. Wei, and W. Pan (2016). An adaptive two-sample test for high-dimensional means. *Biometrika 103*(3), 609–624.