# JOINT VARIABLE SCREENING
# IN ACCELERATED FAILURE TIME MODELS

YIXiN FANG[†] and Jinfeng XU[‡,*]

[†]*AbbVie,* [‡]*The University of Hong Kong*

## Supplementary Material

In the following, some technical results for our proposed methods are provided which include the proofs of Theorem 1, Theorem 2, Proposition 1 and Proposition 2.

Recall that $\boldsymbol{Y} = (Y_1, \cdots, Y_n)^\mathsf{T}$ and $\widehat{\boldsymbol{Y}} = (\Delta_1 Y_1/\widehat{G}(Y_1), \cdots, \Delta_n Y_n/\widehat{G}(Y_n))^\mathsf{T}$. Replacing $\widehat{G}$ by $G$ in $\widehat{\boldsymbol{Y}}$, we define $\widetilde{\boldsymbol{Y}} = (\Delta_1 Y_1/G(Y_1), \cdots, \Delta_n Y_n/G(Y_n))^\mathsf{T}$. Then, we can decompose the estimator $\widehat{\boldsymbol{\beta}}$ in (2.3) as

$$\begin{aligned}
\widehat{\boldsymbol{\beta}} &= \mathbf{X}^\mathsf{T}(\mathbf{X}\mathbf{X}^\mathsf{T})^{-1}\widetilde{\boldsymbol{Y}} + \mathbf{X}^\mathsf{T}(\mathbf{X}\mathbf{X}^\mathsf{T})^{-1}(\widehat{\boldsymbol{Y}} - \widetilde{\boldsymbol{Y}}) \\
&= \mathbf{X}^\mathsf{T}(\mathbf{X}\mathbf{X}^\mathsf{T})^{-1}\left(\frac{\Delta_i Y_i}{G(Y_i)}\right)_{i=1}^{n} + \mathbf{X}^\mathsf{T}(\mathbf{X}\mathbf{X}^\mathsf{T})^{-1}\left(\frac{\Delta_i Y_i}{G(Y_i)}\left[\frac{G(Y_i)}{\widehat{G}(Y_i)} - 1\right]\right)_{i=1}^{n} \\
&= \widehat{\boldsymbol{\beta}}^{(1)} + \widehat{\boldsymbol{\beta}}^{(2)}.
\end{aligned}$$

We can further decompose $\widehat{\boldsymbol{\beta}}^{(1)}$ as

$$\begin{aligned}\widehat{\boldsymbol{\beta}}^{(1)} &= \mathbf{X}^{\mathsf{T}}(\mathbf{X}\mathbf{X}^{\mathsf{T}})^{-1}\mathbf{X}\boldsymbol{\beta}_{\star} + \mathbf{X}^{\mathsf{T}}(\mathbf{X}\mathbf{X}^{\mathsf{T}})^{-1}\left(\left[\frac{\Delta_i}{G(Y_i)} - 1\right]\boldsymbol{X}_i^{\mathsf{T}}\boldsymbol{\beta}_{\star}\right)_{i=1}^{n} + \mathbf{X}^{\mathsf{T}}(\mathbf{X}\mathbf{X}^{\mathsf{T}})^{-1}\left(\frac{\Delta_i\varepsilon_i}{G(Y_i)}\right)_{i=1}^{n} \\ &= \widehat{\boldsymbol{\beta}}^{(1,1)} + \widehat{\boldsymbol{\beta}}^{(1,2)} + \widehat{\boldsymbol{\beta}}^{(1,3)},\end{aligned}$$

and decompose $\widehat{\boldsymbol{\beta}}^{(2)}$ as

$$\begin{aligned}\widehat{\boldsymbol{\beta}}^{(2)} &= \mathbf{X}^{\mathsf{T}}(\mathbf{X}\mathbf{X}^{\mathsf{T}})^{-1}\left(\frac{\Delta_i\boldsymbol{X}_i^{\mathsf{T}}\boldsymbol{\beta}_{\star}}{G(Y_i)}\left[\frac{G(Y_i)}{\widehat{G}(Y_i)} - 1\right]\right)_{i=1}^{n} + \mathbf{X}^{\mathsf{T}}(\mathbf{X}\mathbf{X}^{\mathsf{T}})^{-1}\left(\frac{\Delta_i\varepsilon_i}{G(Y_i)}\left[\frac{G(Y_i)}{\widehat{G}(Y_i)} - 1\right]\right)_{i=1}^{n} \\ &= \widehat{\boldsymbol{\beta}}^{(2,1)} + \widehat{\boldsymbol{\beta}}^{(2,2)}.\end{aligned}$$

## A.1 Property of $\widehat{\boldsymbol{\beta}}^{(1,1)}$

Consider the singular value decomposition of $\mathbf{Z}$ as $\mathbf{Z} = VDU^{\mathsf{T}}$, where $V \in \mathcal{O}(n)$, $U \in V_{n,p_n}$ and $D$ is an $n \times n$ diagonal matrix. Here $\mathcal{O}(n)$ is the set of all $n \times n$ orthogonal matrices and $V_{n,p_n} = \{U \in \mathcal{R}^{p_n \times n} : U^{\mathsf{T}}U = \mathbf{I}_n\}$. This gives $\mathbf{X} = VDU^{\mathsf{T}}\Sigma^{1/2}$. Hence the projection matrix can be written as

$$\mathbf{X}^{\mathsf{T}}(\mathbf{X}\mathbf{X}^{\mathsf{T}})^{-1}\mathbf{X} = HH^{\mathsf{T}},$$

where $H = \Sigma^{1/2}U(U^{\mathsf{T}}\Sigma U)^{-1/2}$ satisfying $H^{\mathsf{T}}H = \mathbf{I}_n$. Therefore, $\widehat{\boldsymbol{\beta}}^{(1,1)} = HH^{\mathsf{T}}\boldsymbol{\beta}$. Let $\boldsymbol{e}_i = (0, \cdots, 0, 1, 0, \cdots, 0)^{\mathsf{T}}$ denote the $i^{\text{th}}$ natural base in the $p_n$ dimension space. Following the proofs of Lemmas 4 and 5 in Wang and Leng (2016) respectively, we derive the following two lemmas.

**Lemma 1.** *Under Assumptions A1-A3, for any $M' > 0$ and for any fixed vector $\boldsymbol{v}$ with $\|\boldsymbol{v}\| = 1$, there exist constants $m_1'$ and $m_2'$ with $0 < m_1' < 1 <$*

*$m_2'$ such that*

$$P\left(\boldsymbol{v}^\top HH^\top \boldsymbol{v} < m_1' n^{1-\tau}/p_n \text{ or } \boldsymbol{v}^\top HH^\top \boldsymbol{v} > m_2' n^{1+\tau}/p_n\right) < 4\exp(-M'n).$$

*In particular for $\boldsymbol{v} = \boldsymbol{\beta}_\star$, whose norm is not 1 though, a similar inequality holds for one side with $m_2' > 1$ (same as previous $m_2'$; if not, the maximum of the two is used in both inequalities) as*

$$P\left(\boldsymbol{\beta}_\star^\top HH^\top \boldsymbol{\beta}_\star > m_2' n^{1+\tau}/p_n\right) < 2\exp(-M'n).$$

**Lemma 2.** *Under Assumptions A1-A3, for any $M' > 0$, there exist some positive constants $m_3'$ and $m_4'$ such that for any $i \in \mathcal{M}_\star$,*

$$P\left(|\boldsymbol{e}_i^\top HH^\top \boldsymbol{\beta}_\star| < m_3' \frac{n^{1-\tau-\kappa}}{p_n}\right) \leqslant O\left\{\exp\left(\frac{-M' n^{1-5\tau-2\kappa-\nu}}{2\log n}\right)\right\},$$

*and for any $i \notin \mathcal{M}_\star$,*

$$P\left(|\boldsymbol{e}_i^\top HH^\top \boldsymbol{\beta}_\star| > \frac{m_4'}{\sqrt{\log n}} \frac{n^{1-\tau-\kappa}}{p_n}\right) \leqslant O\left\{\exp\left(\frac{-M' n^{1-5\tau-2\kappa-\nu}}{2\log n}\right)\right\}.$$

Applying Lemma 1 and Lemma 2 to all $i \in \mathcal{M}_\star$, we have

$$P\left(\min_{i\in\mathcal{M}_\star} |\widehat{\beta}_i^{(1,1)}| < m_3' \frac{n^{1-\tau-\kappa}}{p_n}\right) = O\left\{s_n \exp\left(\frac{-M' n^{1-5\tau-2\kappa-\nu}}{2\log n}\right)\right\}. \quad \text{(A.1)}$$

## A.2 Property of $\widehat{\boldsymbol{\beta}}^{(1,2)}$

Let $\widetilde{\boldsymbol{\epsilon}} = \left(\frac{\Delta_i}{G(Y_i)} - 1\right)_{i=1}^{n}$ and $W = \mathrm{diag}\{\boldsymbol{X}_1^{\mathsf{T}}\boldsymbol{\beta}, \cdots, \boldsymbol{X}_n^{\mathsf{T}}\boldsymbol{\beta}\}$. Then $\widehat{\beta}_i^{(1,2)} = \boldsymbol{e}_i^{\mathsf{T}}\widehat{\boldsymbol{\beta}}^{(1,2)} = \boldsymbol{e}_i^{\mathsf{T}}\mathbf{X}^{\mathsf{T}}(\mathbf{X}\mathbf{X}^{\mathsf{T}})^{-1}W\widetilde{\boldsymbol{\epsilon}}$. If we define

$$\boldsymbol{a} = \boldsymbol{e}_i^{\mathsf{T}}\mathbf{X}^{\mathsf{T}}(\mathbf{X}\mathbf{X}^{\mathsf{T}})^{-1}W/\|\boldsymbol{e}_i^{\mathsf{T}}\mathbf{X}^{\mathsf{T}}(\mathbf{X}\mathbf{X}^{\mathsf{T}})^{-1}W\|_2,$$

then $\widehat{\beta}_i^{(1,2)} = \|\boldsymbol{e}_i^{\mathsf{T}}\mathbf{X}^{\mathsf{T}}(\mathbf{X}\mathbf{X}^{\mathsf{T}})^{-1}W\|_2 \cdot \boldsymbol{a}^{\mathsf{T}}\widetilde{\boldsymbol{\epsilon}}$.

First we investigate the bound of squared norm $\|\boldsymbol{e}_i^{\mathsf{T}}\mathbf{X}^{\mathsf{T}}(\mathbf{X}\mathbf{X}^{\mathsf{T}})^{-1}W\|_2^2$, which equals $\boldsymbol{e}_i^{\mathsf{T}}\mathbf{X}^{\mathsf{T}}(\mathbf{X}\mathbf{X}^{\mathsf{T}})^{-1/2}[(\mathbf{X}\mathbf{X}^{\mathsf{T}})^{-1/2}W^2(\mathbf{X}\mathbf{X}^{\mathsf{T}})^{-1/2}](\mathbf{X}\mathbf{X}^{\mathsf{T}})^{-1/2}\mathbf{X}\boldsymbol{e}_i$. Thus,

$$\|\boldsymbol{e}_i^{\mathsf{T}}\mathbf{X}^{\mathsf{T}}(\mathbf{X}\mathbf{X}^{\mathsf{T}})^{-1}W\|_2^2 \leqslant \lambda_{\max}\{(\mathbf{X}\mathbf{X}^{\mathsf{T}})^{-1/2}W^2(\mathbf{X}\mathbf{X}^{\mathsf{T}})^{-1/2}\} \cdot \boldsymbol{e}_i^{\mathsf{T}}HH^{\mathsf{T}}\boldsymbol{e}_i. \quad \text{(A.2)}$$

Note that $\lambda_{\max}\{(\mathbf{X}\mathbf{X}^{\mathsf{T}})^{-1/2}W^2(\mathbf{X}\mathbf{X}^{\mathsf{T}})^{-1/2}\} \leqslant \lambda_{\max}(W^2)[\lambda_{\min}(\mathbf{Z}\Sigma\mathbf{Z}^{\mathsf{T}})]^{-1}$. S-ince the trace of $\Sigma$ is $p_n$, $\lambda_{\max}(\Sigma) \geqslant 1$. By Assumption A3,

$$\lambda_{\min}(\Sigma) \geqslant \frac{\lambda_{\min}(\Sigma)}{\lambda_{\max}(\Sigma)} > \frac{1}{m_4 n^{\tau}}.$$

Then, by Assumption A1, we have $P(\lambda_{\min}(p_n^{-1}\mathbf{Z}\mathbf{Z}^{\mathsf{T}}) < 1/m_1) \leqslant \exp(-M_1 n)$. By Assumption 5, for any $\varsigma \in (0, 1/2 - 2\tau - \kappa)$,

$$P(|\boldsymbol{\beta}_\star^{\mathsf{T}}\boldsymbol{X}| > n^{\varsigma}) \leqslant 2\exp(-M_2 n^{\varsigma}). \quad \text{(A.3)}$$

In addition, because $\lambda_{\max}(W^2) = \max_{1\leqslant i\leqslant n}(\boldsymbol{X}_i^{\mathsf{T}}\boldsymbol{\beta}_\star)^2$ and by (A.3) , we have $P(\lambda_{\max}(W^2) > n^{2\varsigma}) \leqslant 2n\exp(-M_2 n^{\varsigma})$. Therefore,

$$P\left(\lambda_{\max}\{(\mathbf{X}\mathbf{X}^{\mathsf{T}})^{-1/2}W^2(\mathbf{X}\mathbf{X}^{\mathsf{T}})^{-1/2}\} > \frac{m_1 m_4 n^{\tau+2\varsigma}}{p_n}\right) \leqslant \exp(-M_1 n) + 2n\exp(-M_2 n^{\varsigma}).$$

Combine this result and Lemma 1, we have

$$P\left(\|\boldsymbol{e}_i^\top \mathbf{X}^\top(\mathbf{X}\mathbf{X}^\top)^{-1}W\|_2^2 > \frac{m_1 m_2' m_4 n^{1+2\tau+2\varsigma}}{p_n^2}\right) \leqslant 3\exp(-M_1 n)+2n\exp(-M_2 n^{\varsigma}).$$

Next we consider $\boldsymbol{a}^\top\widetilde{\boldsymbol{\epsilon}}$. Note that condition on $\mathbf{X}=\mathbf{x}$, $|\widetilde{\epsilon}_i| \leqslant 1+1/\delta_1$, which is independent of $\mathbf{x}$. By the General Hoeffding's inequality, there exists $M_3$, which is independent of $\mathbf{x}$, such that, for any $t>0$,

$$P\left(|\boldsymbol{a}^\top\widetilde{\boldsymbol{\epsilon}}| > t \mid \mathbf{X}=\mathbf{x}\right) \leqslant 2\exp\left\{-M_3 t^2/(1+1/\delta_1)\right\}.$$

Therefore, taking expectation on $\mathbf{X}$ and taking $t=\sqrt{M'}n^{1/2-2\tau-\kappa-\varsigma}/\sqrt{\log n}$ for some constant $M'>0$, we have

$$P\left(|\boldsymbol{a}^\top\widetilde{\boldsymbol{\epsilon}}| > \frac{\sqrt{M'}n^{1/2-2\tau-\kappa-\varsigma}}{\sqrt{\log n}}\right) \leqslant 2\exp\left\{\frac{-M'M_3 n^{1-4\tau-2\kappa-2\varsigma}}{(1+1/\delta_1)\log n}\right\}.$$

Combining the above two final results, taking the union bound, we have

$$P\left(|\widehat{\beta}_i^{(1,2)}| > \frac{\sqrt{M' m_1 m_2' m_4}}{\sqrt{\log n}}\frac{n^{1-\tau-\kappa}}{p_n}, \Omega_x\right) < 2\exp\left\{\frac{-M'M_3 n^{1-4\tau-2\kappa-2\varsigma}}{(1+1/\delta_1)\log n}\right\}+3\exp(-M_1 n),$$

where $\Omega_x=\left\{\omega : \max_{1\leqslant i\leqslant n}|\boldsymbol{X}_i^\top\boldsymbol{\beta}| \leqslant n^{\varsigma}\right\}$ with $P(\Omega_x) > 1-2n\exp(-M_2 n^{\varsigma})$.

### A.3 Property of $\widehat{\boldsymbol{\beta}}^{(1,3)}$

Let $\boldsymbol{\epsilon}=\left(\frac{\Delta_i\varepsilon_i}{G(Y_i)}\right)_{i=1}^n$. Then $\widehat{\beta}_i^{(1,3)}=\boldsymbol{e}_i^\top\widehat{\boldsymbol{\beta}}^{(1,3)}=\boldsymbol{e}_i^\top\mathbf{X}^\top(\mathbf{X}\mathbf{X}^\top)^{-1}\boldsymbol{\epsilon}$. If we define

$$\boldsymbol{b}=\boldsymbol{e}_i^\top\mathbf{X}^\top(\mathbf{X}\mathbf{X}^\top)^{-1}/\|\boldsymbol{e}_i^\top\mathbf{X}^\top(\mathbf{X}\mathbf{X}^\top)^{-1}\|_2,$$

then we have $\widehat{\beta}_i^{(1,3)}=\|\boldsymbol{e}_i^\top\mathbf{X}^\top(\mathbf{X}\mathbf{X}^\top)^{-1}\|_2\cdot\boldsymbol{b}^\top\boldsymbol{\epsilon}$.

First we investigate the bound of squared norm $\|\boldsymbol{e}_i^\top \mathbf{X}^\top(\mathbf{X}\mathbf{X}^\top)^{-1}\|_2^2$, which equals $\boldsymbol{e}_i^\top \mathbf{X}^\top(\mathbf{X}\mathbf{X}^\top)^{-1/2}(\mathbf{X}\mathbf{X}^\top)^{-1}(\mathbf{X}\mathbf{X}^\top)^{-1/2}\mathbf{X}\boldsymbol{e}_i$. Thus,

$$\|\boldsymbol{e}_i^\top \mathbf{X}^\top(\mathbf{X}\mathbf{X}^\top)^{-1}W\|_2^2 \leqslant \lambda_{\max}\{(\mathbf{X}\mathbf{X}^\top)^{-1}\} \cdot \boldsymbol{e}_i^\top HH^\top \boldsymbol{e}_i. \tag{A.4}$$

Using the same arguments as those in Section A.2 (replacing $W$ by $\mathbf{I}_n$),

$$P\left(\lambda_{\max}\{(\mathbf{X}\mathbf{X}^\top)^{-1}\} > \frac{m_1 m_4 n^\tau}{p_n}\right) \leqslant \exp(-M_1 n).$$

Combine this result and Lemma 1, we have

$$P\left(\|\boldsymbol{e}_i^\top \mathbf{X}^\top(\mathbf{X}\mathbf{X}^\top)^{-1}\|_2^2 > \frac{m_1 m_2' m_4 n^{1+2\tau}}{p_n^2}\right) < 3\exp(-M_1 n).$$

Next we consider $\boldsymbol{b}^\top \boldsymbol{\epsilon}$. By Assumption A2, we have

$$P\left(|\boldsymbol{b}^\top \boldsymbol{\epsilon}/\sigma| > \frac{\sqrt{M'}n^{1/2-2\tau-\kappa}}{\sqrt{\log n}}\right) \leqslant \exp\left\{1 - q\left(\frac{\sqrt{M'}n^{1/2-2\tau-\kappa}}{\sqrt{\log n}}\right)\right\}.$$

Combining the above two final results, taking the union bound, we have

$$P\left(|\widehat{\beta}_i^{(1,3)}| > \frac{\sqrt{M' m_1 m_2' m_4}}{\sqrt{\log n}}\frac{n^{1-\tau-\kappa}}{p_n}, \Omega_z\right) < \exp\left\{1 - q\left(\frac{\sqrt{M'}n^{1/2-2\tau-\kappa}}{\sqrt{\log n}}\right)\right\},$$

where $\Omega_z = \{\omega : \lambda_{\min}(p_n^{-1}\mathbf{Z}\mathbf{Z}^\top) \geqslant 1/m_1\} \bigcap \{\boldsymbol{\beta}_\star^\top HH^\top \boldsymbol{\beta}_\star \leqslant m_2' n^{1+\tau}/p_n\}$ with $P(\Omega_z) > 1 - 3\exp(-M_1 n)$.

## A.4 Properties of $\widehat{\boldsymbol{\beta}}^{(2,1)}$ and $\widehat{\boldsymbol{\beta}}^{(2,2)}$

**Lemma 3.** (Bitouze 1999; Theorem 1) *Let $\{T_i\}_{i=1}^n$ and $\{C_i\}_{i=1}^n$ be independent sequences of independently identically distributed nonnegative random* *variables with distribution functions $F_1$ and $F_2$, respectively. Let $\widehat{F}_1$ be the Kaplan-Meier estimator of the distribution function $F_1$. There exists a positive constant, $D$, such that for any positive constant $\lambda$,*

$$P\left(n^{1/2}\|(1-F_2)(\widehat{F}_1-F_1)\|_\infty > \lambda\right) \leqslant 2.5\exp(-2\lambda^2+D\lambda).$$

Using Lemma 3 and following the proof of Lemma A3 in Song et al. (2014), we derive the following lemma.

**Lemma 4.** *Let $D$ be the constant in Lemma 3. For any $\lambda > 0$, when $n^{1/2} > D\lambda^{-1}(1-\delta_2)^{-1}/\delta_1$, we have*

$$P\left(\max_{1\leqslant i\leqslant n}\left|\frac{G(V_i)}{\widehat{G}(V_i)}-1\right| \geqslant \lambda\right) \leqslant 2.5\exp\left(-n(1-\delta_2)^2/\delta_1^2\lambda^2\right),$$

*where $V_i = Y_i \wedge \log(C_i)$ and $\delta_1$ and $\delta_2$ are defined in Assumption A4.*

Let $\theta = 0.25 + (\varsigma + 2\tau + \kappa)/2$ and consider $\lambda = n^{-\theta}$ in Lemma 4. Let $\boldsymbol{\zeta} = \left(\frac{\Delta_i \boldsymbol{X}_i^\top \boldsymbol{\beta}_\star}{G(Y_i)}\left[\frac{G(Y_i)}{\widehat{G}(Y_i)}-1\right]\right)_{i=1}^n$. Then $\widehat{\beta}_i^{(2,1)} = \boldsymbol{e}_i^\top \widehat{\boldsymbol{\beta}}^{(2,1)} = \boldsymbol{e}_i^\top \mathbf{X}^\top(\mathbf{X}\mathbf{X}^\top)^{-1}\boldsymbol{\zeta}$. Using $\boldsymbol{b}$ defined in Section A.3, we have $\widehat{\beta}_i^{(2,1)} = \|\boldsymbol{e}_i^\top \mathbf{X}^\top(\mathbf{X}\mathbf{X}^\top)^{-1}\|_2 \cdot \boldsymbol{b}^\top\boldsymbol{\zeta}$.

We have investigated the bound of squared norm $\|\boldsymbol{e}_i^\top \mathbf{X}^\top(\mathbf{X}\mathbf{X}^\top)^{-1}\|_2^2$ in Section A.3. Now we consider $\boldsymbol{b}^\top\boldsymbol{\zeta}$. Let $\Omega_g = \left\{\omega : \max_{1\leqslant i\leqslant n}\left|\frac{G(V_i)}{\widehat{G}(V_i)}-1\right| \leqslant \lambda\right\}$. By Lemma 4, $P(\Omega_g) > 1 - 2.5\exp\left(-n(1-\delta_2)^2/\delta_1^2\lambda^2\right)$. That is, $P(\Omega_g) > 1 - 2.5\exp\left\{-(1-\delta_2)^2/\delta_1^2 n^{1/2-2\tau-\kappa-\varsigma}\right\}$. On $\Omega_g \cap \Omega_x$, by Cauchy-Schwartz inequality, we have

$$|\boldsymbol{b}^\top \boldsymbol{\zeta}| \leqslant \max_{1\leqslant i\leqslant n} \left| \frac{G(V_i)}{\widehat{G}(V_i)} - 1 \right| \sqrt{\sum_{i=1}^{n} (\boldsymbol{X}_i^\top \boldsymbol{\beta}_\star)^2/(1-\delta)^2} \leqslant n^{1/2+\varsigma-\theta}.$$

By the definition of $\theta$, we can verify that $1/2+\varsigma-\theta < 1/2-2\tau-\kappa$. Thus, on $\Omega_g \cap \Omega_x$, $|\boldsymbol{b}^\top \boldsymbol{\zeta}| < \sqrt{M'} n^{1/2-2\tau-\kappa}/\sqrt{\log n}$. Using the bound on the norm of $\boldsymbol{b}$ in Section A.3, we have, on $\Omega_z \cap \Omega_g \cap \Omega_x$,

$$|\widehat{\beta}_i^{(2,1)}| < \frac{\sqrt{M' m_1 m_2' m_4}}{\sqrt{\log n}} \frac{n^{1-\tau-\kappa}}{p_n}.$$

Finally, we consider $\widehat{\boldsymbol{\beta}}^{(2,2)}$. On $\Omega_g$, $\max_{1\leqslant i\leqslant n} \left| \frac{G(V_i)}{\widehat{G}(V_i)} - 1 \right| \leqslant n^{-\theta}$. Comparing $\widehat{\boldsymbol{\beta}}^{(2,2)}$ with $\widehat{\boldsymbol{\beta}}^{(1,3)}$, we see that for any $i$, $\widehat{\beta}_i^{(2,2)} = \widehat{\beta}_i^{(1,3)} O(n^{-\theta})$ on $\Omega_g$. In other words, $\widehat{\boldsymbol{\beta}}^{(2,2)}$ is dominated by $\widehat{\boldsymbol{\beta}}^{(1,3)}$ on $\Omega_g$. Therefore,

$$P\left( |\widehat{\beta}_i^{(2,2)}| > \frac{\sqrt{M' m_1 m_2' m_4}}{\sqrt{\log n}} \frac{n^{1-\tau-\kappa}}{p_n}, \Omega_z \cap \Omega_g \right) < \exp\left\{ 1 - q\left( \frac{\sqrt{M'} n^{1/2-2\tau-\kappa}}{\sqrt{\log n}} \right) \right\}.$$

### A.5 Proof of Theorem 1

*Proof.* Let $\boldsymbol{\xi} = \widehat{\boldsymbol{\beta}}^{(1,1)}$ and $\boldsymbol{\eta} = \widehat{\boldsymbol{\beta}}^{(1,2)} + \widehat{\boldsymbol{\beta}}^{(1,3)} + \widehat{\boldsymbol{\beta}}^{(2,1)} + \widehat{\boldsymbol{\beta}}^{(2,2)}$. Using the final result obtained in Section A.1, for any $i \in \mathcal{M}_\star$, we have

$$P\left( \min_{i\in\mathcal{M}_\star} |\xi_i| < m_3' \frac{n^{1-\tau-\kappa}}{p_n} \right) = O\left\{ s_n \exp\left( \frac{-M' n^{1-5\tau-2\kappa-\nu}}{2\log n} \right) \right\}.$$

Combining the final results obtained in Sections A.2 - A.4,

$$\begin{aligned}
& P\left(\max_{i\in\mathcal{M}_\star}|\eta_i| > \frac{4\sqrt{M'm_1m_2'm_4}}{\sqrt{\log n}}\frac{n^{1-\tau-\kappa}}{p_n}\right) \\
< \; & P\left(\max_{i\in\mathcal{M}_\star}|\widehat{\beta}_i^{(1,2)}|, \max_{i\in\mathcal{M}_\star}|\widehat{\beta}_i^{(1,3)}|, \max_{i\in\mathcal{M}_\star}|\widehat{\beta}_i^{(2,1)}| \text{ or } \max_{i\in\mathcal{M}_\star}|\widehat{\beta}_i^{(2,2)}| > \frac{\sqrt{M'm_1m_2'm_4}}{\sqrt{\log n}}\frac{n^{1-\tau-\kappa}}{p_n}\right) \\
< \; & 2s_n\exp\left\{1-q\left(\frac{\sqrt{M'}n^{1/2-2\tau-\kappa}}{\sqrt{\log n}}\right)\right\} + 2s_n\exp\left\{-M'M_3\frac{n^{1-4\tau-2\kappa-2\varsigma}}{\log n}\right\} \\
& +s_n\exp\left\{1-(1-\delta_2)^2/\delta_1 n^{1/2-2\tau-\kappa-\varsigma}\right\} + 3s_n\exp(-M_1n) + 2n\exp(-M_2n^{\varsigma}).
\end{aligned}$$

Noting that $\varsigma$ defined in (A.3) is any constant in $(0, 1/2-2\tau-\kappa)$, we take $\varsigma = 1/4-\tau-\kappa/2$, leading to $1/2-2\tau-\kappa-\varsigma = \varsigma$. In addition, by Assumption A3, $s_n = m_3n^{\nu}$ with $v < 1$. Therefore, we have

$$\begin{aligned}
& P\left(\max_{i\in\mathcal{M}_\star}|\eta_i| > \frac{4\sqrt{M'm_1m_2'm_4}}{\sqrt{\log n}}\frac{n^{1-\tau-\kappa}}{p_n}\right) \\
< \; & 2s_n\exp\left\{1-q\left(\frac{\sqrt{M}n^{1/2-2\tau-\kappa}}{\sqrt{\log n}}\right)\right\} + O\left\{\exp\left(-Mn^{1/4-\tau-\kappa/2}\right)\right\},
\end{aligned}$$

for some constant $M$.

Moreover, again because $s_n = m_3n^{\nu}$, if $M$ large enough, we have

$$P\left(\min_{i\in\mathcal{M}_\star}|\xi_i| < m_4'\frac{n^{1-\tau-\kappa}}{p_n}\right) = O\left\{\exp\left(\frac{-Mn^{1-5\tau-2\kappa-\nu}}{\log n}\right)\right\}.$$

Therefore, if we choose $\gamma_n$ such that

$$\frac{p_n\gamma_n}{n^{1-\tau-\kappa}} \to 0, \quad \text{and} \quad \frac{p_n\gamma_n\sqrt{\log n}}{n^{1-\tau-\kappa}} \to \infty,$$

then we have

$$
\begin{aligned}
& P\left(\min_{i\in\mathcal{M}_\star}|\widehat{\beta}_i|<\gamma_n\right)=P\left(\min_{i\in\mathcal{M}_\star}|\xi_i+\eta_i|<\gamma_n\right)\\
\leqslant\ & P\left(\min_{i\in\mathcal{M}_\star}|\xi_i|<m_3'\frac{n^{1-\tau-\kappa}}{p_n}\right)+P\left(\max_{i\in\mathcal{M}_\star}|\eta_i|>\frac{4\sqrt{M'm_1m_2'm_4}}{\sqrt{\log n}}\frac{n^{1-\tau-\kappa}}{p_n}\right)\\
=\ & O\left\{\exp\left(\frac{-Mn^{1-5\tau-2\kappa-\nu}}{\log n}\right)\right\}+\varpi(n).
\end{aligned}
$$

This completes the proof of Theorem 1. □

### A.6 Proof of Theorem 2

*Proof.* Following Lemma 2, for any $i\neq\mathcal{M}_\star$ and any $M'>0$, there exists a $m_4'$ such that

$$
P\left(|\boldsymbol{e}_i^\top HH^\top\boldsymbol{\beta}|>\frac{m_4'}{\sqrt{\log n}}\frac{n^{1-\tau-\kappa}}{p_n}\right)\leqslant O\left\{\exp\left(\frac{-M'n^{1-5\tau-2\kappa-\nu}}{2\log n}\right)\right\}.
$$

With Bonferroni's inequality, we have

$$
P\left(\min_{i\notin\mathcal{M}_\star}|\xi_i|>\frac{m_4'}{\sqrt{\log n}}\frac{n^{1-\tau-\kappa}}{p_n}\right)<O\left\{p_n\exp\left(\frac{-M'n^{1-5\tau-2\kappa-\nu}}{2\log n}\right)\right\}.
$$

Also with Bonferroni's inequality, we have

$$
P\left(\max_i|\eta_i|>\frac{\sqrt{M'm_1m_2'm_4}}{\sqrt{\log n}}\frac{n^{1-\tau-\kappa}}{p_n}\right)<p_n\varpi(n).
$$

Now if $p_n$ satisfies

$$
\log p_n=o\left(\min\left\{\frac{n^{1-2\kappa-5\tau-\nu}}{\log n},n^{1/4-\tau-\kappa/2},q\left(\frac{\sqrt{M}n^{1/2-2\tau-\kappa}}{\sqrt{\log n}}\right)\right\}\right),
$$

we have

$$P\left(\min_{i\notin\mathcal{M}_\star}|\xi_i| > \frac{m_4'}{\sqrt{\log n}}\frac{n^{1-\tau-\kappa}}{p_n}\right) < O\left\{\exp\left(\frac{-Mn^{1-5\tau-2\kappa-\nu}}{\log n}\right)\right\},$$

$$P\left(\max_i|\eta_i| > \frac{4\sqrt{M'm_1m_2'm_4}}{\sqrt{\log n}}\frac{n^{1-\tau-\kappa}}{p_n}\right) < \varpi(n).$$

Now if $\gamma_n$ is chosen as the same as in Theorem 1, we have

$$P\left(\max_{i\notin\mathcal{M}_\star}|\widehat{\beta}_i| > \gamma_n\right) < O\left\{\exp\left(\frac{-Mn^{1-5\tau-2\kappa-\nu}}{\log n}\right) + \varpi(n)\right\}.$$

Together with Theorem 1 and the fact that $s_n < p_n$, we have

$$P\left(\max_{i\in\mathcal{M}_\star}|\widehat{\beta}_i| > \gamma_n > \max_{i\notin\mathcal{M}_\star}|\widehat{\beta}_i|\right) = 1 - O\left\{\exp\left(\frac{-Mn^{1-5\tau-2\kappa-\nu}}{\log n}\right) + \varpi(n)\right\}.$$

Furthermore, if we choose a submodel with size $d_n$, we have

$$P\left(\mathcal{M}_\star \subset \mathcal{M}_d\right) = 1 - O\left\{\exp\left(\frac{-Mn^{1-5\tau-2\kappa-\nu}}{\log n}\right) + \varpi(n)\right\}.$$

This completes the proof of Theorem 2. □

### A.7 Proof of Propositions 1 and 2

*Proof of Proposition 1.* Note that $P(\mathcal{M}_\star \subset \overline{\mathcal{M}}_d) = P\left(\bigcap_{b=1}^B\{\mathcal{M}_\star \subset \mathcal{M}_d^{(b)}\}\right)$ and $P\left(\bigcup_{b=1}^B\{\mathcal{M}_\star \subset \mathcal{M}_d^{(b)}\}^c\right) < \sum_{b=1}^B P\left(\{\mathcal{M}_\star \subset \mathcal{M}_d^{(b)}\}^c\right)$. By definition, $\pi(n) = 1 - P(\mathcal{M}_\star \subset \mathcal{M}_d^{(b)})$. Then $P(\mathcal{M}_\star \subset \overline{\mathcal{M}}_d) > 1 - B\pi(n)$. □

*Proof of Proposition 2.* Note that $P(\mathcal{M}_\star \subsetneqq \overline{\mathcal{M}}_d|\mathcal{D}) < \sum_{j\notin\mathcal{M}_\star} P(j \in \overline{\mathcal{M}}_d|\mathcal{D})$ and that for $j \notin \mathcal{M}_\star$, $P(j \in \overline{\mathcal{M}}_d|\mathcal{D}) = \prod_{b=1}^B P(j \in \mathcal{M}^{(b)}|D) < (\frac{d}{p_n-s_n})^B$. Then $P(\mathcal{M}_\star \subsetneqq \overline{\mathcal{M}}_d|\mathcal{D}) < (p_n - s_n)(\frac{d}{p_n-s_n})^B$. □