SUPPLEMENT TO "COVARIATE ASSISTED SCREENING AND ESTIMATION"

By Zheng Tracy Ke[*,†,‡], Jiashun Jin[†] and Jianqing Fan[*]

University of Chicago, Carnegie Mellon University and Princeton University

In this supplement we present the technical proofs for the main article Ke, Jin and Fan (2013). Equation and theorem references made to the main document do not contain letters.

APPENDIX A: PROOF OF THEOREM 2.2

As mentioned before, the success of CASE relies on two noteworthy properties: the Sure Screening (SS) property and the Separable After Screening (SAS) property. In this section, we discuss the two properties in detail, and illustrate how these properties enable us to decompose the original regression problem to many small-size regression problems which can be fit separately. We then use these properties to prove Theorem 2.2.

We start with the SS property. Recall that $\mathcal{U}_p^*$ is the set of all retained indices at the end of the PS-step. The following lemma is proved in Section B.

Lemma A.1 (SS). *Under the conditions of Theorem 2.2,*

$$\sum_{j=1}^{p} P\left(\beta_j \neq 0,\ j \notin \mathcal{U}_p^*\right) \leq L_p[p^{1-(m+1)\vartheta} + \sum_{j=1}^{p} p^{-\rho_j^*(\vartheta,r,G)}] + o(1).$$

This says that all but a negligible fraction of signals are retained in $\mathcal{U}_p^*$.

At the same time, we have the following lemma, which says that as a subgraph of $\mathcal{G}^+$, $\mathcal{U}_p^*$ splits into many disconnected components, and each component has a small size.

[*]Supported in part by National Science Foundation DMS-0704337, the National Institute of General Medical Sciences of the National Institutes of Health through Grant Numbers R01GM100474 and R01-GM072611.

[†]Supported in part by NSF CAREER award DMS-0908613.

[‡]The major work of this article was done when Z. Ke is a graduate student at Department of Operations Research and Financial Engineering, Princeton University.

LEMMA A.2 (SAS). *As $p \to \infty$, under the conditions of Theorem 2.2, there is a fixed integer $l_0 > 0$ such that with probability at least $1 - o(1/p)$, each component of $\mathcal{U}_p^*$ has a size $\leq l_0$.*

Together, these two properties enable us to decompose the original regression problem to many small-size regression problems. To see the point, let $\mathcal{I}$ be a component of $\mathcal{U}_p^*$, and $\mathcal{I}^{pe}$ be the associated patching set. Recall that $d \sim N(B\beta, H)$. If we limit our attention to nodes in $\mathcal{I}^{pe}$, then

$$d^{\mathcal{I}^{pe}} = (B\beta)^{\mathcal{I}^{pe}} + N(0, H^{\mathcal{I}^{pe}, \mathcal{I}^{pe}}). \tag{A.1}$$

Denote $V = \{1, \cdots, p\} \backslash \mathcal{U}_p^*$. Write

$$(B\beta)^{\mathcal{I}^{pe}} = B^{\mathcal{I}^{pe}, \mathcal{I}} \beta^{\mathcal{I}} + \xi_1 + \xi_2, \tag{A.2}$$

where

$$\xi_1 = \sum_{\mathcal{J}: \mathcal{J} \lhd \mathcal{U}_p^*, \mathcal{J} \neq \mathcal{I}} B^{\mathcal{I}^{pe}, \mathcal{J}} \beta^{\mathcal{J}}, \qquad \xi_2 = B^{\mathcal{I}^{pe}, V} \beta^V.$$

Now, first, by the SS property, V contains only a negligible number of signals, so we expect to see that $\|\xi_2\|_\infty$ to be negligibly small. Second, by the SAS property, for any $\mathcal{J} \lhd \mathcal{U}_p^*$ and $\mathcal{J} \neq \mathcal{I}$, nodes in $\mathcal{I}$ and $\mathcal{J}$ are not connected in $\mathcal{G}^+$. By the way $\mathcal{G}^+$ is defined, it follows that nodes in $\mathcal{I}^{pe}$ and $\mathcal{J}$ are not connected in the GOSD $\mathcal{G}^*$. Therefore, we expect to see that $\|\xi_1\|_\infty$ is negligibly small as well. These heuristics are validated in the proof of Theorem 2.2; see Section A.1 for details.

As a result,

$$d^{\mathcal{I}^{pe}} \approx N(B^{\mathcal{I}^{pe}, \mathcal{I}} \beta^{\mathcal{I}}, H^{\mathcal{I}^{pe}, \mathcal{I}^{pe}}), \tag{A.3}$$

where the right hand side is a small-size regression model. In other words, the original regression model decomposes into many small-size regression models, and each has a similar form to that of (A.3).

We now discuss how to fit Model (A.3). In our model $ARW(\vartheta, r, a, \mu)$, $\beta^{\mathcal{I}} = b^{\mathcal{I}} \circ \mu^{\mathcal{I}}$, and $P(\|\beta^{\mathcal{I}}\|_0 = k) \sim \epsilon_p^k$. At the same time, given a realization of $\beta^{\mathcal{I}}$, $d^{\mathcal{I}^{pe}}$ is (approximately) distributed as Gaussian as in (A.3). Combining these, for any eligible $|\mathcal{I}| \times 1$ vector θ, the log-likelihood for $\beta^{\mathcal{I}} = \theta$ associated with (A.3) is

$$- \left[\frac{1}{2} (d^{\mathcal{I}^{pe}} - B^{\mathcal{I}^{pe}, \mathcal{I}} \theta)' (H^{\mathcal{I}^{pe}, \mathcal{I}^{pe}})^{-1} (d^{\mathcal{I}^{pe}} - B^{\mathcal{I}^{pe}, \mathcal{I}} \theta) + \vartheta \log(p) \|\theta\|_0 \right]. \tag{A.4}$$

Note that θ is eligible if and only if its nonzero coordinates $\geq \tau_p$ in magnitude. Comparing (A.4) with (2.20), if the tuning parameters (u^{pe}, v^{pe}) are set as $u^{pe} = \sqrt{2\vartheta \log(p)}$ and $v^{pe} = \sqrt{2r\log(p)}$, then the PE-step is actually the MLE constrained in $\Theta_p(\tau_p)$. This explains the optimality of the PE-step.

The last missing piece of the puzzle is how the information leakage is patched. Consider the oracle situation first where $\beta^{\mathcal{I}^c}$ is known. In such a case, by $\tilde{Y} = X'Y \sim N(G\beta, G)$, it is easy to derive that

$$\tilde{Y}^{\mathcal{I}} - G^{\mathcal{I},\mathcal{I}^c}\beta^{\mathcal{I}^c} \sim N(G^{\mathcal{I},\mathcal{I}}\beta^{\mathcal{I}}, G^{\mathcal{I},\mathcal{I}}).$$

Comparing this with Model (A.3) and applying Lemma 2.2, we see that the information leakage associated with the component $\mathcal{I}$ is captured by the matrix $[U(U'(G^{\mathcal{J}^{pe},\mathcal{J}^{pe}})^{-1}U)^{-1}U']^{\mathcal{I},\mathcal{I}}$, where $\mathcal{J}^{pe} = \{1 \leq j \leq p : D(i,j) \neq 0, \text{for some } i \in \mathcal{I}^{pe}\}$ and U contains an orthonormal basis of $Null(\mathcal{I}^{pe}, \mathcal{J}^{pe})$. To patch the information leakage, we have to show that this matrix has a negligible influence. This is justified in the following lemma, which is proved in Section B.

Lemma A.3. *(Patching). Under the conditions of Theorem 2.2, for any $\mathcal{I} \trianglelefteq \mathcal{G}^+$ such that $|\mathcal{I}| \leq l_0$, and any $|\mathcal{J}^{pe}| \times (|\mathcal{J}^{pe}| - |\mathcal{I}^{pe}|)$ matrix U whose columns form an orthonormal basis of $Null(\mathcal{I}^{pe}, \mathcal{J}^{pe})$,*

$$\big\|[U(U'(G^{\mathcal{J}^{pe},\mathcal{J}^{pe}})^{-1}U)^{-1}U']^{\mathcal{I},\mathcal{I}}\big\| = o(1), \qquad p \to \infty.$$

We are now ready for proving Theorem 2.2.

A.1. Proof of Theorem 2.2 . For short, write $\hat{\beta} = \hat{\beta}^{case}$ and $\rho_j^* = \rho_j^*(\vartheta, r, G)$. For any $\mu \in \Theta_p^*(\tau_p, a)$, write

$$H_p(\hat{\beta}; \epsilon_p, \mu, G) = I + II,$$

where

$$I = \sum_{j=1}^p P\big(\beta_j \neq 0, j \notin \mathcal{U}_p^*\big), \qquad II = \sum_{j=1}^p P\big(j \in \mathcal{U}_p^*,\ \mathrm{sgn}(\hat{\beta}_j) \neq \mathrm{sgn}(\beta_j)\big). \tag{A.5}$$

Using Lemma A.1, $I \leq L_p[p^{1-(m+1)\vartheta} + \sum_{j=1}^p p^{-\rho_j^*}] + o(1)$. So it is sufficient to show

$$II \leq L_p[p^{1-(m+1)\vartheta} + \sum_{j=1}^p p^{-\rho_j^*}] + o(1). \tag{A.6}$$

View $\mathcal{U}_p^*$ as a subgraph of $\mathcal{G}^+$. By Lemma A.2, there is an event A_p and a fixed integer ℓ_0 such that $P(A_p^c) \leq o(1/p)$ and that over the event A_p, each component of $\mathcal{U}_p^*$ has a size $\leq \ell_0$. It is seen that

$$II \leq \sum_{j=1}^{p} P\big(j \in \mathcal{U}_p^*,\ \mathrm{sgn}(\hat{\beta}_j) \neq \mathrm{sgn}(\beta_j),\ A_p\big) + o(1).$$

Moreover, for each $1 \leq j \leq p$, there is a unique component $\mathcal{I}$ such that $j \in \mathcal{I} \lhd \mathcal{U}_p^*$, and that $|\mathcal{I}| \leq \ell_0$ over the event A_p (note that $\mathcal{I}$ depends on $\mathcal{U}_p^*$ and it is random). Since any realization of $\mathcal{I}$ must be a connected subgraph (but not necessarily a component) of $\mathcal{G}^+$,

$$II \leq \sum_{j=1}^{p} \sum_{\mathcal{I}: j \in \mathcal{I} \unlhd \mathcal{G}^+, |\mathcal{I}| \leq l_0} P\big(j \in \mathcal{I} \lhd \mathcal{U}_p^*,\ \mathrm{sgn}(\hat{\beta}_j) \neq \mathrm{sgn}(\beta_j),\ A_p\big) + o(1); \tag{A.7}$$

see Definition 1.8 for the difference between $\lhd$ and $\unlhd$. We stress that on the right hand side of (A.7), we have changed the meaning of $\mathcal{I}$ and use it to denote a fixed (non-random) connected subgraph of $\mathcal{G}^+$.

Next, let $\mathcal{E}(\mathcal{I}^{pe})$ be the set of nodes that are connected to $\mathcal{I}^{pe}$ by a length-1 path in $\mathcal{G}^*$:

$$\mathcal{E}(\mathcal{I}^{pe}) = \{k:\ \text{there is an edge between } k \text{ and } k' \text{ in } \mathcal{G}^* \text{ for some } k' \in \mathcal{I}^{pe}\}.$$

Heuristically, $S(\beta) \cap \mathcal{E}(\mathcal{I}^{pe})$ is the set of signals that have major effects on $d^{\mathcal{I}^{pe}}$. Let $E_{p,\mathcal{I}}$ be the event that $(S(\beta) \cap \mathcal{E}(\mathcal{I}^{pe})) \subset \mathcal{I}$ (note that $\mathcal{I}$ is non-random and the event is defined with respect to the randomness of β). From (A.7), we have

$$II \leq \sum_{j=1}^{p} \sum_{\mathcal{I}: j \in \mathcal{I} \unlhd \mathcal{G}^+, |\mathcal{I}| \leq l_0} P\big(j \in \mathcal{I} \lhd \mathcal{U}_p^*,\ \mathrm{sgn}(\hat{\beta}_j) \neq \mathrm{sgn}(\beta_j),\ A_p \cap E_{p,\mathcal{I}}\big) + rem, \tag{A.8}$$

where it is seen that

$$rem \leq \sum_{j=1}^{p} \sum_{\mathcal{I}: j \in \mathcal{I} \unlhd \mathcal{G}^+, |\mathcal{I}| \leq l_0} P\big(\mathcal{I} \lhd \mathcal{U}_p^*,\ A_p \cap E_{p,\mathcal{I}}^c\big). \tag{A.9}$$

The following lemma is proved in Section B.

LEMMA A.4. *Under the conditions of Theorem 2.2,*

$$\sum_{j=1}^{p} \sum_{\mathcal{I}: j \in \mathcal{I} \unlhd \mathcal{G}^+, |\mathcal{I}| \leq l_0} P\big(\mathcal{I} \lhd \mathcal{U}_p^*, A_p \cap E_{p,\mathcal{I}}^c\big) \leq L_p \sum_{j=1}^{p} P\big(\beta_j \neq 0,\ j \notin \mathcal{U}_p^*\big).$$

Combining (A.9) with Lemma A.4 and using Lemma A.1,

$$rem \le L_p[p^{1-(m+1)\vartheta} + \sum_{j=1}^{p} p^{-\rho_j^*}] + o(1). \tag{A.10}$$

Insert (A.10) into (A.8). To show (A.6), it suffices to show for each $1 \le j \le p$,

$$\sum_{\mathcal{I}: j\in\mathcal{I}\trianglelefteq\mathcal{G}^+, |\mathcal{I}|\le l_0} P\big(j \in \mathcal{I} \lhd \mathcal{U}_p^*,\ \mathrm{sgn}(\hat{\beta}_j) \ne \mathrm{sgn}(\beta_j),\ A_p \cap E_{p,\mathcal{I}}\big) \le L_p p^{-\rho_j^*}. \tag{A.11}$$

We now further reduce (A.11) to a simpler form using the sparsity of $\mathcal{G}^+$. Fix $1 \le j \le p$. The number of subgraphs $\mathcal{I}$ satisfying that $j \in \mathcal{I} \trianglelefteq \mathcal{G}^+$ and that $|\mathcal{I}| \le l_0$ is no more than $C(eK_p^+)^{l_0}$ (Frieze and Molloy, 1999), where K_p^+ is the maximum degree of $\mathcal{G}^+$. By Lemma B.1 and Lemma B.2 (to be stated in Section B), $K_p^+ \le C(\ell^{pe})^2 K_p$, where K_p is the maximum degree of $\mathcal{G}^*$, which is an L_p term. Therefore, $C(eK_p^+)^{l_0}$ is also an L_p term. In other words, the total number of terms in the summation of (A.11) is an L_p term. As a result, to show (A.11), it suffices to show for each fixed $\mathcal{I}$ such that $j \in \mathcal{I} \trianglelefteq \mathcal{G}^+$ and $|\mathcal{I}| \le l_0$,

$$P(j \in \mathcal{I} \lhd \mathcal{U}_p^*,\ \mathrm{sgn}(\hat{\beta}_j) \ne \mathrm{sgn}(\beta_j),\ A_p \cap E_{p,\mathcal{I}}) \le L_p p^{-\rho_j^*}. \tag{A.12}$$

Moreover, note that the left hand side of (A.12) is no more than

$$\sum_{V_0, V_1 \subset \mathcal{I}: j \in V_0\cup V_1} P\big(\mathrm{Supp}(\beta^{\mathcal{I}}) = V_0, \mathrm{Supp}(\hat{\beta}^{\mathcal{I}}) = V_1, \mathcal{I} \lhd \mathcal{U}_p^*, \mathrm{sgn}(\hat{\beta}_j) \ne \mathrm{sgn}(\beta_j), A_p \cap E_{p,\mathcal{I}}\big),$$

where V_0 and V_1 are any non-random subsets satisfying the restriction. Since $|\mathcal{I}| \le l_0$, there are only finite pairs (V_0, V_1) in the summation. Therefore, to show (A.12), it is sufficient to show for each fixed triplet $(\mathcal{I}, V_0, V_1)$ satisfying $\mathcal{I} \trianglelefteq \mathcal{G}^+$, $|\mathcal{I}| \le l_0$, $V_0, V_1 \subset \mathcal{I}$ and $j \in V_0 \cup V_1$ that

$$P\big(\mathrm{Supp}(\beta^{\mathcal{I}}) = V_0, \mathrm{Supp}(\hat{\beta}^{\mathcal{I}}) = V_1, \mathcal{I} \lhd \mathcal{U}_p^*, \mathrm{sgn}(\hat{\beta}_j) \ne \mathrm{sgn}(\beta_j), A_p \cap E_{p,\mathcal{I}}\big) \le L_p p^{-\rho_j^*}. \tag{A.13}$$

We now show (A.13). Fix $(\mathcal{I}, V_0, V_1)$, and write $d_1 = d^{\mathcal{I}^{pe}}$, $B_1 = B^{\mathcal{I}^{pe},\mathcal{I}}$ and $H_1 = H^{\mathcal{I}^{pe},\mathcal{I}^{pe}}$ for short. Define $\Theta_p(\mathcal{I}, a) = \{\theta \in \mathbb{R}^{|\mathcal{I}|} : \theta_j = 0 \text{ or } \tau_p \le |\theta_j| \le a\tau_p\}$ and $\Theta_p(\mathcal{I}) \equiv \Theta_p(\mathcal{I}, \infty)$. Since $u^{pe} = \sqrt{2\vartheta \log(p)}$ and $v^{pe} = \tau_p$, the objective function (2.20) in the PE-step is

$$\mathcal{L}(\theta) \equiv \frac{1}{2}(d_1 - B_1\theta)' H_1^{-1}(d_1 - B_1\theta) + \vartheta \log(p)\|\theta\|_0.$$

Over the event $\{\mathcal{I} \lhd \mathcal{U}_p^*\}$, $\hat{\beta}^{\mathcal{I}}$ minimizes the objective function, so

$$\mathcal{L}(\hat{\beta}^{\mathcal{I}}) \le \mathcal{L}(\beta^{\mathcal{I}}).$$

As a result, the left hand side of (A.13) is no greater than

$$P\big(\mathrm{Supp}(\beta^{\mathcal{I}}) = V_0, \mathrm{Supp}(\hat{\beta}^{\mathcal{I}}) = V_1, \mathcal{L}(\hat{\beta}^{\mathcal{I}}) \le \mathcal{L}(\beta^{\mathcal{I}}), \mathrm{sgn}(\hat{\beta}_j) \ne \mathrm{sgn}(\beta_j), A_p \cap E_{p,\mathcal{I}}\big). \tag{A.14}$$

We now calculate (A.14). Write for short $Q_1 = B_1' H_1^{-1} B_1$, $\widehat{\varpi} = \tau_p^{-2}(\hat{\beta}^{\mathcal{I}} - \beta^{\mathcal{I}})' Q_1 (\hat{\beta}^{\mathcal{I}} - \beta^{\mathcal{I}})$, and define

$$\varpi_j(V_0, V_1, \mathcal{I}) \equiv \frac{1}{\tau_p^2} \min_{(\beta^{(0)}, \beta^{(1)})} (\beta^{(1)} - \beta^{(0)})' Q_1 (\beta^{(1)} - \beta^{(0)}),$$

where the minimum is taken over $(\beta^{(0)}, \beta^{(1)})$ such that $\mathrm{sgn}(\beta_j^{(0)}) \ne \mathrm{sgn}(\beta_j^{(1)})$ and $\beta^{(k)} \in \Theta_p(\mathcal{I})$, $\mathrm{Supp}(\beta^{(k)}) = V_k$, $k = 0, 1$. Introduce

$$\rho_j(V_0, V_1; \mathcal{I}) = \max\{|V_0|, |V_1|\}\vartheta + \frac{1}{4}\left[\left(\sqrt{\varpi_j(V_0, V_1; \mathcal{I}) r} - \frac{|(|V_1| - |V_0|)|)\vartheta}{\sqrt{\varpi_j(V_0, V_1; \mathcal{I}) r}}\right)_+\right]^2. \tag{A.15}$$

Over the event $\{\mathrm{Supp}(\beta^{\mathcal{I}}) = V_0, \mathrm{Supp}(\hat{\beta}^{\mathcal{I}}) = V_1\}$, $\mathcal{L}(\hat{\beta}^{\mathcal{I}}) \le \mathcal{L}(\beta^{\mathcal{I}})$ implies

$$-(d_1 - B_1\beta^{\mathcal{I}})' H_1^{-1} B_1 (\hat{\beta}^{\mathcal{I}} - \beta^{\mathcal{I}}) \ge \frac{1}{2}(\hat{\beta}^{\mathcal{I}} - \beta^{\mathcal{I}})' B_1' H_1^{-1} B_1 (\hat{\beta}^{\mathcal{I}} - \beta^{\mathcal{I}}) + (|V_1| - |V_0|)\vartheta \log(p). \tag{A.16}$$

With the notation $\widehat{\varpi}$, the right hand side of (A.16) is equal to

$$\frac{1}{2}\widehat{\varpi}\tau_p^2 + (|V_1| - |V_0|)\vartheta \log(p). \tag{A.17}$$

To simplify the left hand side of (A.16), we need the following lemma, which is proved in Section B.

Lemma A.5. *For any fixed $\mathcal{I}$ such that $|\mathcal{I}| \le l_0$, and any realization of β over the event $E_{p,\mathcal{I}}$,*

$$(B\beta)^{\mathcal{I}^{pe}} = \zeta + B^{\mathcal{I}^{pe}, \mathcal{I}} \beta^{\mathcal{I}},$$

for some ζ satisfying $\|\zeta\| \le C(\ell^{pe})^{1/2}[\log(p)]^{-(1-1/\alpha)}\tau_p$.

Using Lemma A.5, we can write $d_1 - B_1\beta^{\mathcal{I}} = \zeta + H_1^{1/2}\tilde{z}$, where ζ is as in Lemma A.5 and $\tilde{z} \sim N(0, I_{|\mathcal{I}^{pe}|})$. It follows that the left hand side of (A.16) is equal to

$$-\zeta' H_1^{-1} B_1 (\hat{\beta}^{\mathcal{I}} - \beta^{\mathcal{I}}) + \tilde{z}' H_1^{-1/2} B_1 (\hat{\beta}^{\mathcal{I}} - \beta^{\mathcal{I}}). \tag{A.18}$$

First, by Cauchy-Schwartz inequality, the second term in (A.18) is no larger than $\|\tilde{z}\|\sqrt{\widehat{\varpi}\tau_p^2}$. Second, we argue that the first term in (A.18) is $o(\|\hat{\beta}^{\mathcal{I}} - \beta^{\mathcal{I}}\|\tau_p)$. To see the point, it suffices to check $\|B_1' H_1^{-1}\zeta\| = o(\tau_p)$. In fact, note that since $B \in \mathcal{M}_p(\alpha, A_0)$, $\|B_1\| \leq \|B\| \leq C$; in addition, by RCB, $\|H_1^{-1}\| \leq c_1^{-1}|\mathcal{I}^{pe}|^{\kappa} = O((\ell^{pe})^{\kappa})$. Applying Lemma A.5 and noticing that $\ell^{pe} = (\log(p))^{\nu}$ with $\nu < (1 - 1/\alpha)/(\kappa + 1/2)$, we have $\|B_1' H_1^{-1}\zeta\| \leq \|B_1\|\|H_1^{-1}\|\|\zeta\| \leq C(\ell^{pe})^{\kappa+1/2}[\log(p)]^{-(1-1/\alpha)}\tau_p$, and the claim follows. Third, from Lemma 2.2 and Lemma A.3, $\|G^{\mathcal{I},\mathcal{I}} - Q_1\| = o(1)$ as p grows. So for sufficiently large p, $\lambda_{\min}(Q_1) \geq \frac{1}{2}\lambda_{\min}(G^{\mathcal{I},\mathcal{I}}) \geq C$ for some constant $C > 0$. It follows from the definition of $\widehat{\varpi}$ that $\sqrt{\widehat{\varpi}\tau_p^2} \geq C\|\hat{\beta}^{\mathcal{I}} - \beta^{\mathcal{I}}\|$. Combining these with (A.18), over the event $A_p \cap E_{p,\mathcal{I}}$, the left hand side of (A.16) is no larger than

$$\sqrt{\widehat{\varpi}\tau_p^2}\ \left(\|\tilde{z}\| + o(\tau_p)\right). \tag{A.19}$$

Inserting (A.17) and (A.19) into (A.16), we see that over the event $\{\mathrm{Supp}(\beta^{\mathcal{I}}) = V_0, \mathrm{Supp}(\hat{\beta}^{\mathcal{I}}) = V_1, \mathcal{L}(\hat{\beta}^{\mathcal{I}}) \leq \mathcal{L}(\beta^{\mathcal{I}}), A_p \cap E_{p,\mathcal{I}}\}$,

$$\|\tilde{z}\| \geq \frac{1}{2}\left(\sqrt{\widehat{\varpi}r} + \frac{(|V_1| - |V_0|)\vartheta}{\sqrt{\widehat{\varpi}r}}\right)_+ \sqrt{2\log(p)} + o(\sqrt{\log(p)}). \tag{A.20}$$

Introduce two functions defined over $(0, \infty)$: $J_1(x) = |V_0|\vartheta + \frac{1}{4}\big[\big(\sqrt{x} + \frac{(|V_1|-|V_0|)\vartheta}{\sqrt{x}}\big)_+\big]^2$ and $J_2(x) = \max\{|V_0|, |V_1|\}\vartheta + \frac{1}{4}\big[\big(\sqrt{x} - \frac{|(|V_1|-|V_0|)|\vartheta}{\sqrt{x}}\big)_+\big]^2$. By elementary calculations, $J_1(x) \geq J_2(y)$ for any $x \geq y > 0$. Now, by these notations, (A.20) can be written equivalently as $\|\tilde{z}\|^2 \geq [J_1(\widehat{\varpi}r) - |V_0|\vartheta] \cdot 2\log(p) + o(\log(p))$, and $\rho_j(V_0, V_1; \mathcal{I})$ defined in (A.15) reduces to $J_2(\varpi_j r)$, where $\varpi_j = \varpi_j(V_0, V_1; \mathcal{I})$ for short. Moreover, when $\mathrm{sgn}(\hat{\beta}_j^{\mathcal{I}}) \neq \mathrm{sgn}(\beta_j^{\mathcal{I}})$, $\widehat{\varpi} \geq \varpi_j$ by definition, and hence $J_1(\widehat{\varpi}r) \geq J_2(\varpi_j r)$. Combining these, it follows from (A.20) that over the event $\{\mathrm{Supp}(\beta^{\mathcal{I}}) = V_0, \mathrm{Supp}(\hat{\beta}^{\mathcal{I}}) = V_1, \mathcal{L}(\hat{\beta}^{\mathcal{I}}) \leq \mathcal{L}(\beta^{\mathcal{I}}), \mathrm{sgn}(\hat{\beta}_j^{\mathcal{I}}) \neq \mathrm{sgn}(\beta_j^{\mathcal{I}}), A_p \cap E_{p,\mathcal{I}}\}$,

$$\|\tilde{z}\|^2 \geq \big[\rho_j(V_0, V_1; \mathcal{I}) - |V_0|\vartheta\big] \cdot 2\log(p) + o(\log(p)),$$

where compared to (A.20), the right hand side is now non-random. It follows that the probability in (A.14)

$$\leq P\bigg(\mathrm{Supp}(\beta^{\mathcal{I}}) = V_0,\ \|\tilde{z}\|^2 \geq \big[\rho_j(V_0, V_1; \mathcal{I}) - |V_0|\vartheta\big] \cdot 2\log(p) + o(\log(p))\bigg). \tag{A.21}$$

Recall that $\beta^{\mathcal{I}} = b^{\mathcal{I}} \circ \mu^{\mathcal{I}}$, where b_j's are independent Bernoulli variables with surviving probability $\epsilon_p = p^{-\vartheta}$. It follows that $P(\mathrm{Supp}(\beta^{\mathcal{I}}) = V_0) =$

$L_p p^{-|V_0|\vartheta}$. Moreover, $\|\tilde{z}\|^2$ is independent of $\beta^{\mathcal{I}}$, and is distributed as χ^2 with degree of freedom $|\mathcal{I}^{ps}| \leq L_p$. From basic properties of the χ^2-distribution, $P(\|\tilde{z}\|^2 > 2C\log(p) + o(\log(p))) \leq L_p p^{-C}$ for any $C > 0$. Combining these, we find that the term in (A.21)

$$\leq L_p p^{-|V_0|\vartheta - [\rho_j(V_0,V_1;\mathcal{I}) - |V_0|\vartheta]} = L_p p^{-\rho_j(V_0,V_1;\mathcal{I})}. \tag{A.22}$$

The claim follows by combining (A.22) and the following lemma.

LEMMA A.6. *Under conditions of Theorem 2.2, for any $(j, V_0, V_1, \mathcal{I})$ satisfying $\mathcal{I} \trianglelefteq \mathcal{G}^+$, $|\mathcal{I}| \leq l_0$, $V_0, V_1 \subset \mathcal{I}$ and $j \in V_0 \cup V_1$,*

$$\rho_j(V_0, V_1; \mathcal{I}) \geq \rho_j^*(\vartheta, r, G) + o(1).$$

□

APPENDIX B: PROOFS OF OTHER THEOREMS AND LEMMAS

This section is organized as follows. In Section B.1, we state and prove three preliminary lemmas, which are useful for this section. In Sections B.2-B.12, we give the proofs of all the main theorems and lemmas stated in the preceding sections.

B.1. Preliminary lemmas.

We introduce Lemmas B.1-B.3, where Lemmas B.1-B.2 are proved below, and Lemma B.3 is proved in Lemma in 16 in Jin, Zhang and Zhang (2012).

Recall that $B = DG$ and $\mathcal{G}^*$ is the GOSD in Definition 2.3 with $\delta = 1/\log(p)$. Introduce the matrix B^{**} by

$$B^{**}(i,j) = B(i,j) \cdot 1\big\{ j \in \mathcal{E}(\{i\}) \big\}, \qquad 1 \leq i, j \leq p,$$

where for any set $V \subset \{1, \cdots, p\}$,

$$\mathcal{E}(V) = \big\{ k : \text{ there is an edge between } k \text{ and } k' \text{ in } \mathcal{G}^* \text{ for some } k' \in V \big\}.$$

Recall that $\mathcal{M}_p(\alpha, A_0)$ is the class of matrices defined in (2.9).

LEMMA B.1. *When $B \in \mathcal{M}_p(\alpha, A_0)$, $\mathcal{G}^*$ is K_p-sparse for $K_p \leq C[\log(p)]^{1/\alpha}$, and $\|B - B^{**}\|_\infty \leq C[\log(p)]^{-(1-1/\alpha)}$.*

PROOF. Consider the first claim. Since $B \in \mathcal{M}_p(\alpha, A_0)$ and $H(i,j) = \sum_{k=0}^h \eta_k B(i, j+k)$, there exists a constant $A_0' > 0$ such that $H \in \mathcal{M}_p(\alpha, A_0')$. Let K_p be the smallest integer satisfying

$$K_p \geq 2[\max(A_0, A_0') \log(p)]^{1/\alpha},$$

where it is seen that $K_p \leq C(\log(p))^{1/\alpha}$. At the same time, for any i, j such that $|i-j|+1 > K_p/2$, we have $|B(i,j)| < \delta$, $|B(j,i)| < \delta$ and $|H(i,j)| < \delta$. By definition, there is no edge between nodes i and j in $\mathcal{G}^*$. This proves that $\mathcal{G}^*$ is K_p-sparse, and the claim follows.

Consider the second claim. When $|B(i,j)| > \delta$, there is an edge between nodes i and j in $\mathcal{G}^*$, and it follows that $(B - B^{**})(i,j) = 0$. Therefore, for any $1 \leq i \leq p$,

$$\begin{aligned}\sum_{j=1}^{p} |(B - B^{**})(i,j)| &\leq \sum_{j:|j-i|+1>K_p/2} |B(i,j)| + \sum_{j:|j-i|+1\leq K_p/2, |B(i,j)|\leq\delta} |B(i,j)| \\ &\equiv I + II,\end{aligned}$$

where $I \leq 2A_0 \sum_{k+1>K_p/2} k^{-\alpha} \leq CK_p^{1-\alpha}$ and $II \leq K_p\delta = CK_p^{1-\alpha}$. Recalling $K_p \leq C[\log(p)]^{\alpha}$, $\|B - B^{**}\|_\infty \leq CK_p^{1-\alpha} \leq C[\log(p)]^{-(1-1/\alpha)}$, and the claim follows. □

Next, recall that $\mathcal{G}^+$ is an expanded graph of $\mathcal{G}^*$, given in Definition 2.7, and $\mathcal{I} \lhd \mathcal{G}$ denotes that $\mathcal{I}$ is a component of $\mathcal{G}$, as in Definition 2.8.

LEMMA B.2. *When $\mathcal{G}^*$ is K-sparse, $\mathcal{G}^+$ is $K(2\ell^{pe}+1)^2$-sparse. In addition, for any set $V \subset \{1, \cdots, p\}$, let $\mathcal{G}^+_V$ be the subgraph of $\mathcal{G}^+$ formed by nodes in V. Then for any $\mathcal{I} \lhd \mathcal{G}^+_V$, $(V \backslash \mathcal{I}) \cap \mathcal{E}(\mathcal{I}^{pe}) = \emptyset$.*

PROOF. Consider the first claim. It suffices to show that for any fixed $1 \leq i \leq p$, there are at most $K(2\ell^{pe}+1)^2$ different nodes j such that there is an edge between i and j in $\mathcal{G}^+$. Towards this end, note that $\{i\}^{pe}$ contains no more than $(2\ell^{pe}+1)$ nodes. Since $\mathcal{G}^*$ is K-sparse, for each $k \in \{i\}^{pe}$, there are no more than K nodes k' such that there is an edge between k and k' in $\mathcal{G}^*$. Again, for each such k', there are no more than $(2\ell^{pe}+1)$ nodes j such that $k' \in \{j\}^{pe}$. Combining these gives the claim.

Consider the second claim. Fix V and $\mathcal{I} \lhd \mathcal{G}^+_V$. Since $\mathcal{I}$ is a component, for any $i \in \mathcal{I}$ and $j \in V \backslash \mathcal{I}$, there is no edge between i and j in $\mathcal{G}^+_V$. By definition, this implies $\{j\}^{pe} \cap \mathcal{E}(\{i\}^{pe}) = \emptyset$, and especially $j \notin \mathcal{E}(\{i\}^{pe})$. Since this holds for all such i and j, using that $\mathcal{E}(\mathcal{I}^{pe}) = \cup_{i\in\mathcal{I}}\mathcal{E}(\{i\}^{pe})$, we have $(V \backslash \mathcal{I}) \cap \mathcal{E}(\mathcal{I}^{pe}) = \emptyset$, and the claim follows. □

Finally, recall the definition of $\rho_j^*(\vartheta, r, a, G)$ in (2.29) and that of $\psi(F, N)$ in (2.35).

LEMMA B.3. *When $a > a_g^*(G)$, $\rho_j^*(\vartheta, r, a, G)$ does not depend on a and $\rho_j^*(\vartheta, r, a, G) \equiv \rho_j^*(\vartheta, r, G) = \min_{(F,N): j\in F, F\cap N=\emptyset, F\neq\emptyset} \psi(F, N)$.*

B.2. Proof of Lemma 2.2. For preparation, note that the Fisher Information Matrix associated with model (2.15) is

$$Q \equiv (B^{\mathcal{I}^+,\mathcal{I}})'(H^{\mathcal{I}^+,\mathcal{I}^+})^{-1}(B^{\mathcal{I}^+,\mathcal{I}}).$$

Write $D_1 = D^{\mathcal{I}^+,\mathcal{J}^+}$ and $G_1 = G^{\mathcal{J}^+,\mathcal{J}^+}$ for short. It follows that $B^{\mathcal{I}^+,\mathcal{J}^+} = D_1 G_1$ and $H^{\mathcal{I}^+,\mathcal{I}^+} = D_1 G_1 D_1'$. Let $\mathcal{F}$ be the mapping from $\mathcal{J}^+$ to $\{1, \cdots, |\mathcal{J}^+|\}$ that maps each $j \in \mathcal{J}^+$ to its order in $\mathcal{J}^+$, and let $\mathcal{I}_1 = \mathcal{F}(\mathcal{I})$. By these notations, we can write

$$Q = Q_1^{\mathcal{I}_1,\mathcal{I}_1}, \qquad \text{where} \qquad Q_1 \equiv G_1 D_1'(D_1 G_1 D_1')^{-1} D_1 G_1. \tag{B.1}$$

Comparing (B.1) with the desired claim, it suffices to show

$$Q_1 = G_1 - U(U'G_1^{-1}U)^{-1}U'. \tag{B.2}$$

Let $R = D_1 G_1^{1/2}$ and $P_R = R'(RR')^{-1}R$. It is seen that

$$Q_1 = G_1^{1/2} P_R G_1^{1/2} = G_1 - G_1^{1/2}(I - P_R)G_1^{1/2}. \tag{B.3}$$

Now, we study the matrix $I - P_R$. Let $k = |\mathcal{J}^+|$, and denote $\mathcal{S}(R)$ the row space of R and $\mathcal{N}(R)$ the orthogonal complement of $\mathcal{S}(R)$ in $\mathbb{R}^k$. By construction, P_R is the orthogonal projection matrix from $\mathbb{R}^k$ to $\mathcal{S}(R)$. Hence, $I - P_R$ is the orthogonal projection matrix from $\mathbb{R}^k$ to $\mathcal{N}(R)$. By definition, $\mathcal{N}(R) = \{\eta \in \mathbb{R}^k : R\eta = 0\}$. Recall that $R = D_1 G_1^{1/2}$. Therefore, $R\eta = 0$ if and only if there exists $\xi \in \mathbb{R}^k$ such that $\eta = G_1^{-1/2}\xi$ and $D_1\xi = 0$. At the same time, $Null(\mathcal{I}^+, \mathcal{J}^+) = \{\xi \in \mathbb{R}^k : D_1\xi = 0\}$. Combining these, we have

$$\mathcal{N}(R) = \big\{G_1^{-1/2}\xi : \ \xi \in Null(\mathcal{I}^+, \mathcal{J}^+)\big\}. \tag{B.4}$$

Introduce a new matrix $V = G_1^{-1/2}U$. Since the columns of U form an orthonormal basis of $Null(\mathcal{I}^+, \mathcal{J}^+)$, it follows from (B.4) that the columns of V form a basis (but not necessarily an orthonormal basis) of $\mathcal{N}(R)$. Consequently,

$$I - P_R = V(V'V)^{-1}V' = G_1^{-1/2}U(U'G_1^{-1}U)^{-1}U'G_1^{-1/2}. \tag{B.5}$$

Plugging (B.5) into (B.3) gives (B.2). ☐

B.3. Proof of Lemma 2.4. Write $\rho_j^* = \rho_j^*(\vartheta, r, G)$ for short. It suffices to show for any $\log(p) \leq j \leq p - \log(p)$, there exists (V_0, V_1) such that

(B.6)
$$\rho(V_0, V_1) \leq \rho_j^* + o(1), \qquad j \in (V_0 \cup V_1) \subset \{j + i : -\log(p) \leq i \leq \log(p)\}.$$

In fact, once (B.6) is proved, then $d_p(\mathcal{G}^\diamond) \leq 2\log(p)+1$, and the claim follows directly.

We now construct (V_0, V_1) to satisfy (B.6) for any j such that $\log(p) \leq j \leq p - \log(p)$. The key is to construct a sequence of set pairs $(V_0^{(t)}, V_1^{(t)})$ recursively as follows. Let $V_0^{(1)} = V_{0j}^*$ and $V_1^{(1)} = V_{1j}^*$, where (V_{0j}^*, V_{1j}^*) are as defined in Section 2.8. For any integer $t \geq 1$, we update $(V_0^{(t)}, V_1^{(t)})$ as follows. If all inter-distance between the nodes in $V_0^{(t)} \cup V_1^{(t)}$ (assuming all nodes are sorted ascendingly) does not exceed $\log(p)/g$, then the process terminates. Otherwise, there are a pair of adjacent nodes i_1 and i_2 in $(V_0^{(t)} \cup V_1^{(t)})$ (again, assuming the nodes are sorted ascendingly) such that $i_2 > i_1 + \log(p)/g$. In our construction, it is not hard to see that $j \in V_0^{(t)} \cup V_1^{(t)}$. Therefore, we have either the case of $j \leq i_1$ or the case of $j \geq i_2$. In the first case, we let

$$N^{(t+1)} = N^{(t)} \cap \{i : i \leq i_1\}, \qquad F^{(t+1)} = F^{(t)} \cap \{i : i \leq i_1\},$$

and in the second case, we let

$$N^{(t+1)} = N^{(t)} \cap \{i : i \geq i_2\}, \qquad F^{(t+1)} = F^{(t)} \cap \{i : i \geq i_2\},$$

where $N^{(t)} = V_0^{(t)} \cap V_1^{(t)}$ and $F^{(t)} = (V_0^{(t)} \cup V_1^{(t)}) \setminus N^{(t)}$. We then update by defining

$$V_0^{(t+1)} = N^{(t+1)} \cup F', \qquad V_1^{(t+1)} = N^{(t+1)} \cup F''$$

where (F', F'') are constructed as follows: Write $F^{(t)} = \{j_1, j_2, \cdots, j_k\}$ where $j_1 < j_2 < \ldots < j_k$ and $k = |F^{(t)}|$. When k is even, let $F' = \{j_1, \cdots, j_{k/2}\}$ and $F'' = F^{(t)} \backslash F'$; otherwise, let $F' = \{j_1, \cdots, j_{(k-1)/2}\}$ and $F'' = F^{(t)} \backslash F'$.

Now, first, by the construction, $|F^{(t)} \cup N^{(t)}|$ is strictly decreasing in t. Second, by Lemma 12 in Jin, Zhang and Zhang (2012), $|F^{(1)} \cup N^{(1)}| \leq |V_{0j}^* \cup V_{1j}^*| \leq g$. As a result, the recursive process above terminates in finite rounds. Let T be the number of rounds when the process terminates, we construct (V_0, V_1) by

(B.7)
$$V_0 = V_0^{(T)}, \qquad V_1 = V_1^{(T)}.$$

Next, we justify (V_0, V_1) constructed in (B.7) satisfies (B.6). First, it is easy to see that $j \in V_0 \cup V_1$ and $|V_0 \cup V_1| \leq g$. Second, all pairs of adjacent nodes in $V_0 \cup V_1$ have an inter-distance $\leq \log(p)/g$ (assuming all nodes are sorted), so $(V_0 \cup V_1) \subset \{j - \log(p), \cdots, j + \log(p)\}$. As a result, all remains to show is

$$\rho(V_0, V_1) \leq \rho_j^* + o(1). \tag{B.8}$$

By similar argument as in Lemma 16 in Jin, Zhang and Zhang (2012) and definitions (i.e. (2.35) and (2.32) in Jin, Zhang and Zhang (2012)), if $a > a_g^*(G)$, then for any (V_0', V_1') such that $|V_0' \cup V_1'| \leq g$, we have $\rho(V_0', V_1') \geq \psi(F', N')$, where $N' = V_0' \cap V_1'$ and $F' = (V_0' \cup V_1') \backslash N'$. Moreover, the equality holds when $|V_0'| = |V_1'|$ in the case $|F'|$ is even, and $|V_0'| - |V_1'| = \pm 1$ in the case $|F'|$ is odd. Combining these with definitions,

$$\rho(V_0, V_1) = \psi(F^{(T)}, N^{(T)}), \quad \rho_j^* \equiv \rho(V_{0j}^*, V_{1j}^*) = \rho(V_0^{(1)}, V_1^{(1)}) \geq \psi(F^{(1)}, N^{(1)}).$$

Recall that T is a finite number. So to show (B.8), it suffices to show for each $1 \leq t \leq T-1$,

$$\psi(F^{(t+1)}, N^{(t+1)}) \leq \psi(F^{(t)}, N^{(t)}) + o(1). \tag{B.9}$$

Fixing $1 \leq t \leq T-1$, write for short $F = F^{(t)}$, $N = N^{(t)}$, $N_1 = N^{(t+1)}$ and $F_1 = F^{(t+1)}$. Let $\mathcal{I} = F \cup N$ and $\mathcal{I}_1 = F_1 \cup N_1$. With these notations, (B.9) reduces to

$$\psi(F_1, N_1) \leq \psi(F, N) + o(1). \tag{B.10}$$

By the way ψ is defined (i.e., (2.35)), it is sufficient to show

$$\omega(F_1, N_1) \leq \omega(F, N) + o(1). \tag{B.11}$$

In fact, once (B.11) is proved, (B.9) follows by noting that $|F_1| + 2|N_1| \leq |F| + 2|N| - 1$.

We now show (B.11). Letting $\Omega = \mathrm{diag}(G^{\mathcal{I}_1, \mathcal{I}_1}, G^{\mathcal{I} \backslash \mathcal{I}_1, \mathcal{I} \backslash \mathcal{I}_1})$, we write

$$\begin{aligned}
\omega(F, N) &= \min_{\theta \in \mathbb{R}^{|\mathcal{I}|}: |\theta_i| \geq 1, \forall i \in F} \theta' G^{\mathcal{I}, \mathcal{I}} \theta \\
&\geq \min_{\theta \in \mathbb{R}^{|\mathcal{I}|}: |\theta_i| \geq 1, \forall i \in F} \theta' \Omega \theta - \max_{\theta \in \mathbb{R}^{|\mathcal{I}|}: |\theta_i| \leq 2a, \forall i} |\theta' (G^{\mathcal{I}, \mathcal{I}} - \Omega) \theta| \\
&\geq \min_{\theta \in \mathbb{R}^{|\mathcal{I}_1|}: |\theta_i| \geq 1, \forall i \in F_1} \theta' G^{\mathcal{I}_1, \mathcal{I}_1} \theta - \max_{\theta \in \mathbb{R}^{|\mathcal{I}|}: |\theta_i| \leq 2a, \forall i} |\theta' (G^{\mathcal{I}, \mathcal{I}} - \Omega) \theta| \qquad \text{(B.12)} \\
&= \omega(F_1, N_1) - \max_{\theta \in \mathbb{R}^{|\mathcal{I}|}: |\theta_i| \leq 2a, \forall i} |\theta' (G^{\mathcal{I}, \mathcal{I}} - \Omega) \theta|,
\end{aligned}$$

where in the first and last equalities we use equivalent forms of $\omega(F, N)$, in the second inequality we use the fact that the constraints $|\theta_i| \geq 1$ can be replaced by $1 \le |\theta_i| \le a$ for any $a > a_g^*$ and the triangular inequality, and in the third inequality we use the definition of Ω.

Finally, note that for any $k \in \mathcal{I}_1$ and $k' \in \mathcal{I}\backslash\mathcal{I}_1$, $|k - k'| > \log(p)/g$ holds. In addition, G has polynomial off-diagonal decays with rate $\gamma > 0$. Together we find that $\|G^{\mathcal{I},\mathcal{I}} - \Omega\| \le C(\log(p)/g)^{-\gamma} = o(1)$. As a result, $\max_{\theta\in\mathbb{R}^{|\mathcal{I}|}:|\theta_i|\le 2a,\forall i} |\theta'(G^{\mathcal{I},\mathcal{I}} - \Omega)\theta| \le Ca^2 \cdot \|G^{\mathcal{I},\mathcal{I}} - \Omega\| \cdot |\mathcal{I}| = o(1)$. Inserting this into (B.12) gives (B.11). □

B.4. Proof of Theorem 2.3. First, we define $\rho^*_{lts}(\vartheta, r; f)$ as follows. For any spectral density function f, let $G^\infty = G^\infty(f)$ be the (infinitely dimensional) Toeplitz matrix generated by f: $G^\infty(i,j) = \hat{f}(|i-j|)$ for any $i, j \in \mathbb{Z}$, where $\hat{f}(k)$ is the k-th Fourier coefficient of f. In the definition of $\rho(V_0, V_1)$ in (2.27)-(2.28), replace G by G^∞ and call the new term $\rho^\infty(V_0, V_1)$. For any fixed j, let

$$\rho^*_{j,lts}(\vartheta, r; f) = \min_{(V_0,V_1):j\in V_0\cup V_1} \rho^\infty(V_0, V_1), \tag{B.13}$$

where V_0, V_1 are subsets of $\mathbb{Z}$. Due to the definition of Toeplitz matrices, $\rho^*_{j,lts}(\vartheta, r; f)$ does not depend on j, so we write it as $\rho^*_{lts}(\vartheta, r; f)$ for short. By (B.6), it is seen that

$$\rho^*_j(\vartheta, r, G) = \rho^*_{lts}(\vartheta, r; f) + o(1), \quad \text{for any } \log(p) \le j \le p - \log(p). \tag{B.14}$$

Now, to show the claim, it is sufficient to check the main conditions of Theorem 2.2. In detail, it suffices to check that

(a) $G \in \mathcal{M}^*_p(\gamma, g, c_0, A_1)$ with $\gamma = 1 - 2\phi > 0$, $A_1 > 0$ and $c_0 > 0$.
(b) $B \in \mathcal{M}_p(\alpha, A_0)$ with $\alpha = 2 - 2\phi > 1$ and $A_0 > 0$.
(c) Conditions RCA and RCB hold with $\kappa = 2 - 2\phi > 0$ and $c_1 > 0$.

To show these claims, we need some lemmas and results in elementary calculus. In detail, first, we have

$$|f'(\omega)| \le C|\omega|^{-(2\phi+1)}, \qquad |f''(\omega)| \le C|\omega|^{-(2\phi+2)}. \tag{B.15}$$

For a proof of (B.15), we rewrite $f(\omega) = f^*(\omega)/|2\sin(\omega/2)|^{2\phi}$, where by assumption $f^*(\omega)$ is a continuous function that is twice differentiable except at 0, and $|(f^*)''(\omega)| \le C|\omega|^{-2}$. It can be derived from basic properties in analysis that

$$|(f^*)''(\omega)| \le C|\omega|^{-2}, \quad |(f^*)'(\omega)| \le C|\omega|^{-1}, \quad \text{and} \quad |f^*(\omega)| \le C. \tag{B.16}$$

At the same time, by elementary calculation,

$$\begin{aligned} |f'(\omega)| &\le C|\omega|^{-(2\phi+1)}(|f^*(\omega)| + |\omega(f^*)'(\omega)|), \\ |f''(\omega)| &\le C|\omega|^{-(2\phi+2)}(|f^*(\omega)| + |\omega(f^*)'(\omega)| + |\omega^2(f^*)''(\omega)|), \end{aligned}$$

and (B.15) follows by plugging in (B.16).

Second, we need the following lemma, whose proof is a simple exercise of analysis and omitted.

LEMMA B.4. *Suppose g is a symmetric real function which is differentiable in $[-\pi, 0) \cup (0, \pi]$ and $|g'(\omega)| \leq C|\omega|^{-\alpha}$ for some $\alpha \in (1, 2)$. Then as $x \to \infty$, $\int_{-\pi}^{\pi} \cos(\omega x) g(\omega) d\omega = O\left(|x|^{-(2-\alpha)}\right)$.*

We now show (a)-(c). Consider (a) first. First, by (B.15) and Lemma B.4, $\int_{-\pi}^{\pi} \cos(k\omega) f(\omega) d\omega \leq Ck^{-(1-2\phi)}$ for large k, so that $|G(i,j)| \leq C(1 + |i - j|)^{-(1-2\phi)}$. Second, by well-known results on Toeplitz matrices, $\lambda_{\min}(G) \geq \min_{\omega \in [-\pi, \pi]} f(\omega) > 0$. Combining these, (a) holds with $\gamma = 1 - 2\phi$ and $c_0 = \min_{\omega \in [-\pi,\pi]} f(\omega)$.

Next, we consider (b). Recall that $B = DG$ where D is the first-order row-differencing matrix. So $B(i,j) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left[\cos(k\omega) - \cos((k+1)\omega)\right] f(\omega) d\omega$, where $k = i - j$. Without loss of generality, we only consider the case $k \geq 1$. Denote $g(\omega) = \omega f(\omega)$. By Fubini's theorem and integration by part,

$$
\begin{aligned}
B(i,j) &= \frac{1}{\pi} \int_0^{\pi} \left[\int_k^{k+1} \omega \sin(\omega x) dx\right] f(\omega) d\omega \\
&= \frac{1}{\pi} \int_k^{k+1} \left[\int_0^{\pi} g(\omega) \sin(\omega x) d\omega\right] dx \\
&= \frac{1}{\pi} \int_k^{k+1} \left[-g(\pi) \frac{\cos(\pi x)}{x} + \int_0^{\pi} \frac{\cos(\omega x)}{x} g'(\omega) d\omega\right] dx \\
&= -\frac{g(\pi)}{\pi} \int_k^{k+1} \frac{\cos(\pi x)}{x} dx + \frac{1}{2\pi} \int_k^{k+1} \frac{1}{x} \left[\int_{-\pi}^{\pi} \cos(\omega x) g'(\omega) d\omega\right] dx \\
&\equiv I_1 + I_2
\end{aligned}
$$

First, using integration by part, $|I_1| = \left|\pi^{-1} g(\pi) \int_k^{k+1} \frac{\sin(\pi x)}{\pi x^2} dx\right| = O(k^{-2})$. Second, similar to (B.15), we derive that $g''(\omega) = O(|\omega|^{-(1+2\phi)})$. Applying Lemma B.4 to g', we have $|\int_{-\pi}^{\pi} \cos(\omega x) g'(\omega) d\omega| \leq C|x|^{-(1-2\phi)}$, and so $|I_2| \leq \int_k^{k+1} Cx^{-(2-2\phi)} dx = O(k^{-(2-2\phi)})$. Combining these gives $|B(i,j)| \leq C(1 + |i - j|)^{-(2-2\phi)}$, and (b) holds with $\alpha = 2 - 2\phi$.

Last, we show (c). Since $\varphi_\eta(z) = 1 - z$, RCA holds trivially, and all remains is to check that RCB holds. Recall that $H = DGD'$, where D is the first-order row-differencing matrix. The goal is to show there exsits a constant $c_1 > 0$ such that for any triplet (k, b, V),

$$
b' H^{V,V} b \geq c_1 k^{-(2-2\phi)} \|b\|^2, \tag{B.17}
$$

where $1 \leq k \leq p$ is an integer, $b \in \mathbb{R}^k$ is a vector, and $V \subset \{1, 2, \ldots, p\}$ is a subset with $|V| = k$.

Towards this end, we introduce $f_1(\omega) = 4\sin^2(\omega/2) f(\omega)$, where we recall that f is the spectral density associated with G. Fixing a triplet (k, b, V), we write $b = (b_1, b_2, \ldots, b_k)'$ and $V = \{j_1, \cdots, j_k\}$ such that $j_1 < j_2 < \ldots < j_k$. By definitions and basic algebra,

$$
\begin{aligned}
H(i,j) &= G(i,j) - G(i+1,j) - G(i,j+1) + G(i+1,j+1) \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi} \left[2\cos(k\omega) - \cos((k+1)\omega) - \cos((k-1)\omega)\right] f(\omega) d\omega \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi} \cos(k\omega) f_1(\omega) d\omega, \qquad \text{where for short } k = i - j,
\end{aligned}
$$

which, together with direct calculations, implies that

$$
b' H^{V,V} b = \frac{1}{2\pi}\int_{-\pi}^{\pi} \sum_{s=1}^{k}\sum_{t=1}^{k} b_s b_t \cos((j_s - j_t)\omega) f_1(\omega) d\omega = \frac{1}{2\pi}\int_{-\pi}^{\pi} \Big| \sum_{s=1}^{k} b_s e^{\sqrt{-1} j_s \omega} \Big|^2 f_1(\omega) d\omega.
$$

At the same time, note that $f_1(\omega) \geq C|\omega|^{2-2\phi}$ for any $\omega \neq 0$ and $|\omega| \leq \pi$. Combining these with symmetry and monotonicity gives

(B.18)
$$
b' H^{V,V} b \geq \frac{C}{\pi}\int_{0}^{\pi} \Big| \sum_{s=1}^{k} b_s e^{\sqrt{-1} j_s \omega} \Big|^2 \omega^{2-2\phi} d\omega \geq \frac{C}{\pi}\int_{\pi/(2k)}^{\pi} \Big| \sum_{s=1}^{k} b_s e^{\sqrt{-1} j_s \omega} \Big|^2 \omega^{2-2\phi} d\omega.
$$

Next, we write

$$
\|b\|^2 = \frac{1}{\pi}\int_{0}^{\pi} \Big| \sum_{s=1}^{k} b_s e^{\sqrt{-1} j_s \omega} \Big|^2 d\omega = I + II, \tag{B.19}
$$

where I and II are the integration in the interval of $[0, \pi/(2k)]$ and $[\pi/(2k), \pi]$, respectively. By (B.18) and the monotonicity of the function $\omega^{2-2\phi}$ in $[\pi/(2k), \pi]$,

$$
b' H^{V,V} b \geq C k^{-(2-2\phi)} \cdot \frac{1}{\pi}\int_{\pi/(2k)}^{\pi} \Big| \sum_{s=1}^{k} b_s e^{\sqrt{-1} j_s \omega} \Big|^2 d\omega \equiv C k^{-(2-2\phi)} \cdot II. \tag{B.20}
$$

At the same time, by the Cauchy-Schwartz inequality, $|\sum_{s=1}^{k} b_s e^{\sqrt{-1} j_s \omega}|^2 \leq (\sum_{s=1}^{k} |e^{\sqrt{-1} j_s \omega}|^2)(\sum_{s=1}^{k} |b_s|^2) = k\|b\|^2$, and so $I \leq \frac{1}{\pi}\int_0^{\pi/(2k)} k\|b\|^2 d\omega \leq \|b\|^2/2$. Inserting this into (B.19) gives

$$
II \geq \|b\|^2 - \|b\|^2/2 = \|b\|^2/2, \tag{B.21}
$$

and (B.17) follows by combining (B.20) and (B.21). □

B.5. Proof of Lemma 2.5. First, we show $r^*_{lts}(\vartheta) \equiv r^*_{lts}(\vartheta; f)$ is a decreasing function of ϑ. Similarly to the proof of Theorem 2.3, in the definition of $\omega(F,N)$ and $\psi(F,N)$ (recall (2.35) and (2.36)), replace G by G^∞, and denote the new terms by $\omega^\infty(F,N) \equiv \omega^\infty(F,N;\vartheta,r,f)$ and $\psi^\infty(F,N) \equiv \psi^\infty(F,N;\vartheta,r,f)$, respectively. By similar argument in Lemma B.3,

$$\rho^*_{lts}(\vartheta, r; f) = \min_{(F,N): F\cap N=\emptyset, F\neq\emptyset} \psi^\infty(F,N).$$

For each pair of sets (F,N) and $\vartheta \in (0,1)$ let $r^*(\vartheta; F,N) \equiv r^*(\vartheta; F,N,f)$ be the minimum r such that $\psi^\infty(F,N;\vartheta,r,f) \geq 1$. It follows that

$$r^*_{lts}(\vartheta) = \max_{(F,N): F\cap N=\emptyset, F\neq\emptyset} r^*(\vartheta; F,N).$$

It is easy to see that $r^*(\vartheta; F,N)$ is a decreasing function of ϑ for each fixed (F,N). So $r^*_{lts}(\vartheta)$ is also a decreasing function of ϑ.

Next, we consider $\lim_{\vartheta\to 1} r^*_{lts}(\vartheta)$. In the special case of $F = \{j\}$ and $N = \emptyset$, $\omega^\infty(F,N) = 1$, $\lim_{\vartheta\to 1} r^*(\vartheta; F,N) = 1$, and so $\liminf_{\vartheta\to 1} r^*_{lts}(\vartheta) \geq 1$. At the same time, for any (F,N) such that $|F|+|N| > 1$, $\psi^\infty(F,N) \geq \vartheta$ and so $\lim_{\vartheta\to 1} r^*(\vartheta; F,N) \leq 1$. Hence, $\limsup_{\vartheta\to 1} r^*_{lts}(\vartheta) \leq 1$. Combining these gives the claim.

Last, we consider $\lim_{\vartheta\to 0} r^*_{lts}(\vartheta)$. First, since $\lim_{\vartheta\to 0} \psi^\infty(F,N) = \omega^\infty(F,N) r/4$ for any fixed (F,N), we have

$$\lim_{\vartheta\to 0} r^*_{lts}(\vartheta) = 4\left[\min_{(F,N): F\cap N=\emptyset, F\neq\emptyset} \omega^\infty(F,N)\right]^{-1}. \tag{B.22}$$

Second, by definitions,

$$\min_{(F,N): F\cap N=\emptyset, F\neq\emptyset} \omega^\infty(F,N) = \lim_{p\to\infty} \min_{(F,N): (F\cup N)\subset\{1,\cdots,p\}, F\cap N=\emptyset, F\neq\emptyset} \omega(F,N), \tag{B.23}$$

whenever the limit on the right hand side exists.

Third, note that (a) Given F, $\omega(F,N)$ decreases as N increases and (b) Given $F\cup N$, $\omega(F,N)$ decreases as N increases (the proofs are straightforward and we omit them). As a result, for all (F,N) such that $(F\cup N) \subset \{1,\cdots,p\}$, $\omega(F,N)$ is minimized at $F = \{j\}$ and $N = \{1,\cdots,p\}\backslash\{j\}$ for some j, with the minimal value equaling the reciprocal of the j-th diagonal of G^{-1}. In other words,

$$\lim_{p\to\infty} \min_{(F,N): (F\cup N)\subset\{1,\cdots,p\}, F\cap N=\emptyset, F\neq\emptyset} \omega(F,N) = \left[\lim_{p\to\infty} \max_{1\leq j\leq p} G^{-1}(j,j)\right]^{-1}. \tag{B.24}$$

Fourth, if we write $G = G_p$ to emphasize on the size of G, then by basic algebra and the Toeplitz structure of G, we have $(G_p^{-1})(j,j) \leq (G_{p+k}^{-1})(j+$

$k, j+k)$ for all $1 \leq k \leq p-j$ and $(G_p^{-1})(j,j) \leq (G_{p+k}^{-1})(j-k, j-k)$ for $1 \leq k \leq j-1$. Especially, if we take $k = \log(p)$, then it follows that

$$\lim_{p\to\infty} \max_{1\leq j\leq p} G^{-1}(j,j) = \lim_{p\to\infty} \max_{\log(p)\leq j\leq p-\log(p)} G^{-1}(j,j). \tag{B.25}$$

Last, we have the following lemma which is proved in Appendix C.

LEMMA B.5. *Under conditions of Lemma 2.5,*

$$\lim_{p\to\infty} \max_{\log(p)\leq j\leq p-\log(p)} G^{-1}(j,j) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f^{-1}(\omega)d\omega.$$

Combining (B.22)-(B.25) and using Lemma B.5,

$$\lim_{\vartheta\to 0} r_{lts}^*(\vartheta) = 4\cdot \big[\lim_{p\to\infty} \max_{\log(p)\leq j\leq p-\log(p)} G^{-1}(j,j)\big] = \frac{2}{\pi}\int_{-\pi}^{\pi} f^{-1}(\omega)d\omega.$$

□

B.6. Proof of Theorem 2.4. Write for short $\hat\beta = \hat\beta^{case}$ and $\rho_{cp}^* = \rho_{cp}^*(\vartheta, r)$. It suffices to show

$$\text{Hamm}_p^*(\vartheta, r, G) \geq L_p p^{1-\rho_{cp}^*}; \tag{B.26}$$

and for any $\mu \in \Theta_p^*(\tau_p, a)$,

$$H_p(\hat\beta; \epsilon_p, \mu, G) \equiv \sum_{j=1}^{p} P\big(\text{sgn}(\hat\beta_j) \neq \text{sgn}(\beta_j)\big) \leq L_p p^{1-\rho_{cp}^*} + o(1). \tag{B.27}$$

First, we show (B.26). The statement is similar to that of Theorem 2.1, but $d_p(\mathcal{G}^\diamond) \leq L_p$ does not hold. Therefore, we introduce a different graph $\mathcal{G}^\nabla$ as follows: Define a counter part of $\rho_j^*(\vartheta, r, G)$ as

$$\tilde\rho_j^*(\vartheta, r, G) = \min_{(V_0,V_1):\min(V_0\cup V_1)=j} \rho(V_0, V_1), \tag{B.28}$$

where $\min(V_0 \cup V_1) = j$ means j is the smallest node in $V_0 \cup V_1$. Let (V_{0j}^*, V_{1j}^*) be the minimizer of (B.28), and when there is a tie, pick the one that appears first lexicographically. Define the graph $\mathcal{G}^\nabla$ with nodes $\{1, \cdots, p\}$, and that there is an edge between nodes j and k whenever $(V_{0j}^* \cup V_{1j}^*) \cap (V_{0k}^* \cup V_{1k}^*) \neq \emptyset$.

Denote $d_p(\mathcal{G}^\nabla)$ the maximum degree of nodes in $\mathcal{G}^\nabla$. Similar to Theorem 2.1, as $p \to \infty$,

$$\text{Hamm}_p^*(\vartheta, r, G) \geq L_p [d_p(\mathcal{G}^\nabla)]^{-1} \sum_{j=1}^{p} p^{-\tilde\rho_j^*(\vartheta, r, G)}. \tag{B.29}$$

The proof is a trivial extension of Theorem 14 in Jin, Zhang and Zhang (2012) and we omit it. Moreover, the following lemma is proved below.

LEMMA B.6. *As* $p \to \infty$, $\max_{\log(p) \leq j \leq p-\log(p)} |\tilde{\rho}_j^*(\vartheta, r, G) - \rho_{cp}^*(\vartheta, r)| = o(1)$, *and* $d_p(\mathcal{G}^{\nabla}) \leq L_p$.

Combining (B.29) with Lemma B.6 gives (B.26).

Second, we show (B.27). The change-point model is an 'extreme' case and Theorem 2.2 does not apply directly. However, once we justify the following claims (a)-(c), (B.27) follows by similar arguments in Theorem 2.2.

(a) SS property:

$$\sum_{j=1}^{p} P\big(\beta_j \neq 0, j \notin \mathcal{U}_p^*\big) \leq L_p p^{1-\rho_{cp}^*} + o(1).$$

(b) SAS property: If we view $\mathcal{U}_p^*$ as a subgraph of $\mathcal{G}^+$, there is a fixed integer $l_0 > 0$ such that with probability at least $1 - o(1/p)$, each component of $\mathcal{U}_p^*$ has a size $\leq l_0$.

(c) A counter part of Lemma A.6: For any $\log(p) \leq j \leq p - \log(p)$, and fixed $\mathcal{I} \trianglelefteq \mathcal{G}^+$ such that $j \in \mathcal{I}$ and $|\mathcal{I}| \leq l_0$, suppose we construct $\{\mathcal{I}^{(k')}, \mathcal{I}^{(k'),pe}, 1 \leq k' \leq N\}$ using the process introduced in the *PE*-step, and $j \in \mathcal{I}^{(k)}$. Then for any pair of sets (V_0, V_1) such that $\mathcal{I}^{(k)} = V_0 \cup V_1$,

$$\rho_j(V_0, V_1; \mathcal{I}^{(k)}) \geq \rho_{cp}^* + o(1),$$

where $\rho_j(V_0, V_1; \mathcal{I}^{(k)})$ is defined in (A.15).

Consider (a) first. Following the proof of Lemma A.1 until (B.53), we find that for each $\log(p) \leq j \leq p - \log(p)$,

$$P\big(\beta_j \neq 0, j \notin \mathcal{U}_p^*\big) \leq \sum_{(\mathcal{I},F,N): j \in \mathcal{I} \trianglelefteq \mathcal{G}^*, |\mathcal{I}| \leq m, F \cup N = \mathcal{I}, F \cap N = \emptyset, F \neq \emptyset} L_p p^{-|\mathcal{I}|\vartheta - [(\sqrt{\omega_0 r} - \sqrt{q})_+]^2} + L_p p^{-(m+1)\vartheta} + o(1/p)$$

where $\omega_0 = \tau_p^{-2}(\beta^F)'[Q^{F,F} - Q^{F,N}(Q^{N,N})^{-1}Q^{N,F}]\beta^F$ and Q is defined as in (2.17). First, by the choice of m, $L_p p^{-(m+1)\vartheta} \leq L_p p^{-\rho_{cp}^*}$. Second, using similar arguments in Lemma A.1, the summation contains at most L_p terms. Third, by (2.37), $\omega_0 \geq \tilde{\omega}(F, N)$. Combining the above, it suffices to show for each triplet $(\mathcal{I}, F, N)$ in the summation,

$$|\mathcal{I}|\vartheta + [(\sqrt{\tilde{\omega}(F,N) r} - \sqrt{q})_+]^2 \geq \rho_{cp}^*. \tag{B.30}$$

The key to (B.30) is to show

$$\tilde{\omega}(F, N) \geq 1/2. \tag{B.31}$$

Once (B.31) is proved, since $q \leq \frac{r}{4}(\sqrt{2}-1)^2$,

$$|\mathcal{I}|\vartheta + [(\sqrt{\tilde{\omega}(F,N)r} - \sqrt{q})_+]^2 \geq |\mathcal{I}|\vartheta + r/4 \geq \rho_{cp}^*,$$

where in the last inequality we use the facts $\rho_{cp}^* \leq \vartheta + r/4$ and $|\mathcal{I}| \geq 1$. This gives (B.30).

All remains is to show (B.31). We argue that it suffices to consider those $(\mathcal{I}, F, N)$ where both $\mathcal{I}(= F \cup N)$ and F are formed by consecutive nodes. First, since G is tri-diagonal, the definition of $\mathcal{G}^*$ implies that any $\mathcal{I} \trianglelefteq \mathcal{G}^*$ is formed by consecutive nodes. Second, by (2.37) and basic algebra,

$$\tilde{\omega}(F,N) = \min_{\xi \in \mathbb{R}^{|F|}: |\xi_i| \geq 1} \xi'[(Q^{-1})^{F,F}]^{-1}\xi, \tag{B.32}$$

where Q is defined in (2.17). Note that B is an identity matrix and $\mathcal{I}^{ps} = \mathcal{I}$. So $Q^{-1} = H^{\mathcal{I},\mathcal{I}}$, which is a tri-diagonal matrix. It follows from (B.32) that if F is not formed by consecutive nodes, there exist $F_1 \subset F$ and $N_1 = \mathcal{I} \backslash F_1$ such that $\tilde{\omega}(F_1, N_1) \leq \tilde{\omega}(F, N)$. The argument then follows.

From now on, we focus on $(\mathcal{I}, F, N)$ such that both $\mathcal{I}$ and F are formed by consecutive nodes. Elementary calculation yields

$$[(Q^{-1})^{F,F}]^{-1} = (H^{F,F})^{-1} = \Omega^{(k)} - \frac{1}{k+1}\eta\eta', \tag{B.33}$$

where $k = |F|$, $\Omega^{(k)}$ is the $k \times k$ matrix defined by $\Omega^{(k)}(i,j) = \min\{i,j\}$ and $\eta = (1, \cdots, k)'$. We see that $\tilde{\omega}(F,N)$ only depends on k. When $k = 1$, $\tilde{\omega}(F,N) = 1/2$ by direct calculations following (B.32) and (B.33). When $k \geq 2$, from (B.32) and (B.33),

$$\tilde{\omega}(F,N) = \min_{\xi \in \mathbb{R}^{|F|}: |\xi_i| \geq 1} \left[\sum_{l=1}^{k} (\xi_l + \cdots + \xi_k)^2 - \frac{1}{k+1}(\xi_1 + 2\xi_2 + \cdots + k\xi_k)^2 \right].$$

Let $s_l = \sum_{j=l}^k \xi_j$. The above right hand side is lower bounded by $\sum_{l=1}^k s_l^2 - (\sum_{l=1}^k s_l)^2/k = \sum_{l<l'}(s_l - s_{l'})^2/k$, where $\sum_{l<l'}(s_l - s_{l'})^2 \geq \sum_{l=1}^{k-1}(s_{l+1} - s_l)^2 \geq k-1$. Therefore,

$$\tilde{\omega}(F,N) \geq (k-1)/k \geq 1/2.$$

This proves (B.31).

Next, consider (b). We check RCB, and the remaining proof is exactly the same as in Lemma A.2. Towards this end, the goal is to show there exists a constant $c_1 > 0$ such that for any (k, V) where $V \subset \{1, \cdots, p\}$ and $k = |V|$,

$$\lambda_{\min}(H^{V,V}) \geq c_1 k^{-2}. \tag{B.34}$$

Since H is tri-diagonal, it suffices to show that (B.34) holds when V is formed by consecutive nodes, i.e., $V = \{j, \cdots, j+k\}$ for some $1 \leq j \leq p-k$. In this case, we introduce a matrix $\Sigma^{(k)}$, which is 'smaller' than $H^{V,V}$ but much easier to analyse:

$$\Sigma^{(k)}(i,j) = 2 \cdot 1\{i=j\} - 1\{|i-j|=1\} - 1\{i=j=k\}, \quad 1 \leq i,j \leq k.$$

It is easy to see that $H^{V,V} - \Sigma^{(k)}$ is positive semi-definite. Hence,

$$\lambda_{\min}(H^{V,V}) \geq \lambda_{\min}(\Sigma^{(k)}). \tag{B.35}$$

Observing that $(\Sigma^{(k)})^{-1} = \Omega^{(k)}$, where $\Omega^{(k)}$ is as in (B.33), we have

$$\lambda_{\min}(\Sigma^{(k)}) = \left[\lambda_{\max}(\Omega^{(k)})\right]^{-1} \geq \left[\|\Omega^{(k)}\|_\infty\right]^{-1} = 2/(k^2+k). \tag{B.36}$$

Combining (B.35)-(B.36) gives (B.34).

Finally, consider (c). Fix $1 \leq j \leq p$ and the triplet $(\mathcal{I}^{(k)}, V_0, V_1)$, where $|\mathcal{I}^{(k)}| \leq l_0$. The goal is to show

$$\rho_j(V_0, V_1; \mathcal{I}^{(k)}) \geq \rho^*_{cp} + o(1). \tag{B.37}$$

Introduce the following quantities: From the PE-step and the choice $\ell^{pe} = 2\log(p)$, we can write

$$\mathcal{I}^{(k),pe} = \{j_1+1, \cdots, j_1+L\} \quad \text{and} \quad \mathcal{I}^{(k)} = \{j_1+M_1, \cdots, j_1+M_2\},$$

where the integers L, M_1 and M_2 staisfy

$$M_2 - M_1 \leq l_0+1, \quad M_1 \geq [\log(p)]^{1/(l_0+1)}, \quad (L-M_2)/M_1 \geq [\log(p)]^{1/(l_0+1)}. \tag{B.38}$$

Denote $K = M_2 - M_1 + 1$, $M_0 = M_1 - \frac{M_1^2}{L+1}$ and $\mathcal{I}'' = \{M_0, \cdots, M_0+K-1\}$. Let $\mathcal{F}$ be the one-to-one mapping from $\mathcal{I}^{(k)}$ to $\mathcal{I}''$ such that $\mathcal{F}(i) = i - (j_1 + M_1) + M_0$. Denote $V_0'' = \mathcal{F}(V_0)$ and $V_1'' = \mathcal{F}(V_1)$. Recall the definitions of $\varpi_j(V_0, V_1; \mathcal{I}^{(k)})$ and $\varpi^*(V_0, V_1)$ (see (A.15) and (2.27)). We claim that

$$\varpi_j(V_0, V_1; \mathcal{I}^{(k)}) \geq \varpi^*(V_0'', V_1'') + o(1). \tag{B.39}$$

Once we have (B.39), plug it into the definition $\rho_j(V_0, V_1; \mathcal{I}^{(k)})$ and use the monotonicity of the function $f(x) = [(x - a/x)_+]^2$ over $(0, \infty)$ when $a > 0$. It follows that

$$\rho_j(V_0, V_1; \mathcal{I}^{(k)}) \geq \max\{|V_0|, |V_1|\}\vartheta + \frac{1}{4}\left[\left(\sqrt{\varpi^* r} - \frac{\big||V_1| - |V_0|\big|\vartheta}{\sqrt{\varpi^* r}}\right)_+\right]^2 + o(1).$$

where ϖ^* is short for $\varpi^*(V_0'', V_1'')$. Compare the first term on the right hand side with (2.28) and recall that $|V_0''| = |V_0|$ and $|V_1''| = |V_1|$. It follows that

$$\rho_j(V_0, V_1; \mathcal{I}^{(k)}) \geq \rho(V_0'', V_1'') + o(1). \tag{B.40}$$

Moreover, since $M_0 = \min(V_0'', V_1'')$, by (B.28),

$$\rho(V_0'', V_1'') \geq \tilde{\rho}^*_{M_0}. \tag{B.41}$$

Note that (B.38) implies $M_0 \gtrsim M_1 \geq [\log(p)]^{1/(1+l_0)}$. By a trivial extension of Lemma B.6, we can derive that $\max_{(\log(p))^{1/(1+l_0)} \leq j \leq p-(\log(p))^{1/(1+l_0)}} |\tilde{\rho}^*_j - \rho^*_{cp}| = o(1)$. These together imply

$$\tilde{\rho}^*_{M_0} = \rho^*_{cp} + o(1). \tag{B.42}$$

Combining (B.40)-(B.42) gives (B.37).

What remains is to show (B.39). The proof is similar to that of (B.86). In detail, write for short $\varpi_j = \varpi_j(V_0, V_1; \mathcal{I}^{(k)})$, $\varpi^* = \varpi^*(V_0'', V_1'')$, $B_1 = B^{\mathcal{I}^{(k)},pe, \mathcal{I}^{(k)}}$, $H_1 = H^{\mathcal{I}^{(k)},pe, \mathcal{I}^{(k)},pe}$ and $Q_1 = B_1' H_1^{-1} B_1$. By similar arguments in (B.87), $\varpi_j \geq \min_{j \in \mathcal{I}} \varpi_j$, and there exists a constant $a_1 > 0$ such that

$$|\min_{j \in \mathcal{I}} \varpi_j - \varpi^*| \leq \max_{\xi \in \mathbb{R}^K : \|\xi\|_\infty \leq 2a_1} \left|\xi'(G^{\mathcal{I}'', \mathcal{I}''} - Q_1)\xi\right| \leq C\|G^{\mathcal{I}'', \mathcal{I}''} - Q_1\|.$$

Therefore, it suffices to show that

$$\|G^{\mathcal{I}'', \mathcal{I}''} - Q_1\| = o(1). \tag{B.43}$$

Note that Q_1 is the $(\mathcal{I}', \mathcal{I}')$-block of H_1^{-1}, where the index set $\mathcal{I}' = \{M_1, \cdots, M_2\}$. By (B.33), $H_1^{-1} = \Omega^{(L)} - \frac{1}{L+1}\eta\eta'$, where $\eta = (1, 2, \cdots, L)'$. It follows that

$$Q_1 = (M_1 - 1)1_K 1_K' + \Omega^{(K)} - \frac{1}{L+1}\xi\xi',$$

where 1_K is the K-dimensional vector whose elements are all equal to 1, and $\xi = (M_1, \cdots, M_2)'$. Define the $L \times L$ matrix Δ by $\Delta(i,j) = \frac{ij - M_1^2}{L+1}$, for $1 \leq i, j \leq L$ and let Δ_1 be the submatrix of Δ by restricting the rows and columns to $\mathcal{I}'$. By these notations,

$$Q_1 = (M_0 - 1)1_K 1_K' + \Omega^{(K)} - \Delta_1.$$

At the same time, we observe that

$$G^{\mathcal{I}'', \mathcal{I}''} = (M_0 - 1)1_K 1_K' + \Omega^{(K)}.$$

Combining the above yields that $G^{\mathcal{I}'', \mathcal{I}''} - Q_1 = \Delta_1$. Note that $|\Delta(i,j)| \leq \frac{M_2^2 - M_1^2}{L+1} \leq \frac{(l_0+1)(2M_1+l_0+1)}{L+1} = o(1)$ for all $i, j \in \mathcal{I}'$. Hence, $\|\Delta_1\| = o(1)$ and (B.43) follows directly. ∎

B.6.1. *Proof of Lemma B.6.* To show the claim, we need to introduce some quantities and lemmas. First, by a trivial extension of Lemma B.3,

$$\tilde{\rho}_j^*(\vartheta, r, G) = \min_{(F,N):\min(F\cup N)=j, F\cap N=\emptyset, F\neq\emptyset} \psi(F, N).$$

where $\psi(F, N) = \psi(F, N; \vartheta, r, G)$, defined in (2.35).

Second, let $\mathcal{R}_p$ denote the collection of all subsets of $\{1, \cdots, p\}$ that are formed by consecutive nodes. Define

$$\tilde{\tilde{\rho}}_j^*(\vartheta, r, G) = \min_{(F,N):\min(F\cup N)=j, F\cap N=\emptyset, F\neq\emptyset, F\cup N\in\mathcal{R}_p, F\in\mathcal{R}_p, |F|\leq 3, |N|\leq 2} \psi(F, N),$$

where we emphasize that the minimum is taken over finite pairs (F, N). The following lemma is proved in Appendix C.

LEMMA B.7. *As $p \to \infty$, $\max_{\log(p)\leq j\leq p-\log(p)} |\tilde{\rho}_j^*(\vartheta, r, G) - \tilde{\tilde{\rho}}_j^*(\vartheta, r, G)| = o(1)$.*

Third, for each dimension k, define the $k \times k$ matrix $\Sigma_*^{(k)}$ as

$$\Sigma_*^{(k)}(i, j) = 2 \cdot 1\{i = j\} - 1\{|i - j| = 1\}, \tag{B.44}$$

except that $\Sigma_*^{(k)}(1, 1) = \Sigma_*^{(k)}(k, k) = 1$, and the $k \times k$ matrix $\Omega_*^{(k)}$ as

$$\Omega_*^{(k)}(i, j) = \min\{i, j\} - 1. \tag{B.45}$$

Let

$$\omega^{(\infty)}(F, N) = \begin{cases} \min_{\xi\in\mathbb{R}^{|F|}:|\xi_j|\geq 1} \xi'[(\Sigma_*^{(k)})^{F,F}]^{-1}\xi, & |N| > 0 \\ \min_{\xi\in\mathbb{R}^{|\mathcal{I}|}:|\xi_j|\geq 1, 1'\xi=0} \xi'\Omega_*^{(k)}\xi, & |N| = 0,\ |F| > 1 \\ \infty, & |N| = 0,\ |F| = 1 \end{cases}$$

and define $\psi^{(\infty)}(F, N) = \psi^{(\infty)}(F, N; \vartheta, r, G)$, a counter part of $\psi(F, N)$, by replacing $\omega(F, N)$ by $\omega^{(\infty)}(F, N)$ in the definition (2.35). Let

$$\rho^{(\infty)}(\vartheta, r) = \min_{(F,N):\min(F\cup N)=1, F\cap N=\emptyset, F\neq\emptyset, F\in\mathcal{R}_p, F\cup N\in\mathcal{R}_p, |F|\leq 3, |N|\leq 2} \psi^{(\infty)}(F, N),$$

where we note that $\rho^{(\infty)}(\vartheta, r)$ does not depend on j. The following lemma is proved in Appendix C.

LEMMA B.8. *As $p \to \infty$, $\max_{\log(p)\leq j\leq p-\log(p)} |\tilde{\tilde{\rho}}_j^*(\vartheta, r, G) - \rho^{(\infty)}(\vartheta, r)| = o(1)$.*

Now, we show the claims. Write for short $\tilde{\rho}_j^* = \tilde{\rho}_j^*(\vartheta, r, G)$, and $\tilde{\tilde{\rho}}_j^*$, ρ_{cp}^* similarly. First, we show

$$d_p(\mathcal{G}^\triangledown) \leq L_p.$$

Denote (F_j^*, N_j^*) the minimum in defining $\tilde{\tilde{\rho}}_j^*$, and if there is a tie, we pick the one that appears first lexicographically. By definition and Lemma B.7, for any $\log(p) \leq j \leq p - \log(p)$,

$$\psi(F_j^*, N_j^*) \equiv \tilde{\tilde{\rho}}_j^* = \tilde{\rho}_j^* + o(1), \qquad \text{and} \qquad (F_j^* \cup N_j^*) \subset \{j, \cdots, j+4\}.$$

By the definition of $\mathcal{G}^\triangledown$, these imply that there is an edge between nodes j and k only when $|k - j| \leq 4$. So $d_p(\mathcal{G}^\triangledown) \leq C$.

Next, we show for all $\log(p) \leq j \leq p - \log(p)$,

$$\tilde{\rho}_j^* = \rho_{cp}^* + o(1).$$

By Lemma B.7 and Lemma B.8, it suffices to show

$$\rho^{(\infty)} = \rho_{cp}^*. \tag{B.46}$$

Introduce the function $\nu(\cdot; F, N)$ for each (F, N):

$$\nu(x; F, N) = \begin{cases} (|F| + 2|N|)/2 + \omega^{(\infty)} x/4, & |F| \text{ is even}, \\ (|F| + 2|N| + 1)/2 + [(\sqrt{\omega^{(\infty)} x} - 1/\sqrt{\omega^{(\infty)} x})_+]^2/4, & |F| \text{ is odd}, \end{cases}$$

where $\omega^{(\infty)}$ is short for $\omega^{(\infty)}(F, N)$. Then we can write

$$\psi^{(\infty)}(F, N; \vartheta, r, G) = \vartheta \cdot \nu(r/\vartheta; F, N).$$

Let $\nu^*(x) = \min_{(F,N)} \nu(x; F, N)$, where the minimum is taken over those (F, N) in defining $\rho^{(\infty)}$. It follows that

$$\rho^{(\infty)}(\vartheta, r) = \min_{(F,N)} \vartheta \cdot \nu(r/\vartheta; F, N) = \vartheta \cdot \nu^*(r/\vartheta). \tag{B.47}$$

Below, we compute the function $\nu^*(\cdot)$ by computing the functions $\nu(\cdot; F, N)$ for the finite pairs (F, N) in defining $\rho^{(\infty)}$. After excluding some obviously non-optimal pairs, all possible cases are displayed in Table 1. Using Table 1, we can further exclude the cases with $|F| = 3$. In the remaining, for each fixed value of $\omega^{(\infty)}$, we keep two pairs of (F, N) which minimize $|F| + 2|N|$ among those with $|F|$ odd and even respectively. The results are displayed in Table 2. Then $\nu^*(\cdot)$ is the lower envelope of the four functions listed. Direct calculations yield

$$\nu^*(x) = \begin{cases} 1 + x/4, & 0 < x \leq 6 + 2\sqrt{10}; \\ 3 + (\sqrt{x} - 2/\sqrt{x})^2/8, & x > 6 + 2\sqrt{10}. \end{cases}$$

Plugging this into (B.47) and comparing it with the definition of ρ_{cp}^*, we obtain (B.46). □

TABLE 1
Calculation of $\omega^{(\infty)}(F,N)$

F	N	$(\Sigma_*^{(k)})^{F,F}$	$\Omega_*^{(k)}$	ξ^*	$\omega^{(\infty)}(F,N)$
$\{1\}$	$\{2\}$	1	-	1	1
$\{2\}$	$\{1,3\}$	2	-	1	$\frac{1}{2}$
$\{1,2\}$	$\emptyset$	-	$\begin{pmatrix}0&0\\0&1\end{pmatrix}$	$(1,-1)'$	1
$\{1,2\}$	$\{3\}$	$\begin{pmatrix}1&-1\\-1&2\end{pmatrix}$	-	$(1,-1)'$	1
$\{2,3\}$	$\{1,4\}$	$\begin{pmatrix}2&-1\\-1&2\end{pmatrix}$	-	$(1,-1)'$	$\frac{2}{3}$
$\{1,2,3\}$	$\emptyset$	-	$\begin{pmatrix}0&0&0\\0&1&1\\0&1&2\end{pmatrix}$	$(1,-2,1)'$	2
$\{1,2,3\}$	$\{4\}$	$\begin{pmatrix}1&-1&0\\-1&2&-1\\0&-1&2\end{pmatrix}$	-	$(1,-\frac{3}{2},1)'$	$\frac{3}{2}$
$\{2,3,4\}$	$\{1,5\}$	$\begin{pmatrix}2&-1&0\\-1&2&-1\\0&-1&2\end{pmatrix}$	-	$(1,-1,1)'$	1

TABLE 2
Calculation of $\nu(x;F,N)$

$\omega^{(\infty)}$	$\lvert F\rvert$	$\lvert N\rvert$	$\nu(x;F,N)$	$\lVert\xi^*\rVert_\infty$
1	1	1	$2+\frac{1}{4}(\sqrt{x}-\frac{1}{\sqrt{x}})_+^2$	1
1	2	0	$1+\frac{x}{4}$	1
$\frac{1}{2}$	1	2	$3+\frac{1}{8}(\sqrt{x}-\frac{2}{\sqrt{x}})_+^2$	1
$\frac{2}{3}$	2	2	$3+\frac{x}{6}$	1

B.7. Proof of Lemma A.1. Fix ϑ and r. Write for short $\rho_j^*=\rho_j^*(\vartheta,r,G)$. To show the claim, it suffices to show for each $1\le j\le p$,

$$P\big(\beta_j\neq 0,\ j\notin\mathcal{U}_p^*\big)\le L_p[p^{-\rho_j^*}+p^{-(m+1)\vartheta}]+o(1/p). \tag{B.48}$$

Fix $1\le j\le p$. Recall that $\mathcal{G}_S^*$ is the subgraph of $\mathcal{G}^*$ by restricting the nodes into $S(\beta)$. Over the event $\{\beta_j\neq 0\}$, there is a unique component $\mathcal{I}$ such that $j\in\mathcal{I}\lhd\mathcal{G}_S^*$. By Frieze and Molloy (1999), $|\mathcal{I}|\le m$ except for a probability of at most $L_p p^{-(m+1)\vartheta}$, where the randomness comes from the law of β. Denote this event as $A_p=A_{p,j}$. To show (B.48), it suffices to show

$$P\big(\beta_j\neq 0,\ j\notin\mathcal{U}_p^*,\ A_p\big)\le L_p p^{-\rho_j^*}+o(1/p). \tag{B.49}$$

Note that $\mathcal{I}$ depends on β (and so is random), and also that over the event A_p, any realization of $\mathcal{I}$ is a connected subgraph in $\mathcal{G}^*$ with size $\le m$.

Therefore,

$$P\big(\beta_j \neq 0,\ j \notin \mathcal{U}_p^*,\ A_p\big) \leq \sum_{\mathcal{I}: j \in \mathcal{I} \trianglelefteq \mathcal{G}^*, |\mathcal{I}| \leq m} P\big(j \in \mathcal{I} \lhd \mathcal{G}_S^*,\ j \notin \mathcal{U}_p^*,\ A_p\big),$$

where on the right hand side, we have misused the notation slightly by denoting $\mathcal{I}$ as a fixed (non-random) connected subgraph of $\mathcal{G}^*$. Since $\mathcal{G}^*$ is K_p-sparse (see Lemma B.1), for any fixed j, there are no more than $C(eK_p)^m$ connected subgraph $\mathcal{I}$ such that $j \in \mathcal{I}$ and $|\mathcal{I}| \leq m$ (Frieze and Molloy, 1999). Noticing that $C(eK_p)^m \leq L_p$, to show (B.49), it is sufficient to show for any fixed $\mathcal{I}$ such that $j \in \mathcal{I} \trianglelefteq \mathcal{G}^*$ and $|\mathcal{I}| \leq m$,

$$P\big(j \in \mathcal{I} \lhd \mathcal{G}_S^*,\ j \notin \mathcal{U}_p^*,\ A_p\big) \leq L_p p^{-\rho_j^*} + o(1/p). \tag{B.50}$$

Fix such an $\mathcal{I}$. The subgraph (as a whole) has been screened in some sub-stage of the PS-step, say, sub-stage t. Let $\hat{N} = \mathcal{U}^{(t-1)} \cap \mathcal{I}$ and $\hat{F} = \mathcal{I} \backslash \hat{N}$ be as in the initial sub-step of the PS-step. By definitions, the event $\{j \notin \mathcal{U}_p^*\}$ is contained in the event that $\mathcal{I}$ fails to pass the χ^2-test in (2.19). As a result,

$$\begin{aligned} P\big(j \in \mathcal{I} \lhd \mathcal{G}_S^*, j \notin \mathcal{U}_p^*, A_p\big) &\leq P\big(j \in \mathcal{I} \lhd \mathcal{G}_S^*, T(d, \hat{F}, \hat{N}) \leq 2q(\hat{F}, \hat{N}) \log(p), A_p\big) \\ &\leq \sum_{(F,N): F \cup N = \mathcal{I}, F \cap N = \emptyset, F \neq \emptyset} P\big(j \in \mathcal{I} \lhd \mathcal{G}_S^*, T(d, F, N) \leq 2q(F, N) \log(p), A_p\big), \end{aligned}$$

where (F, N) are fixed (non-random) subsets, and $q = q(F, N)$ is either as in (2.33) or in (2.38). Since $|\mathcal{I}| \leq m$, the summation in the second line only involves at most finite terms. Therefore, to show (B.50), it suffices to show for each fixed triplet $(\mathcal{I}, F, N)$ satisfying $j \in \mathcal{I} \trianglelefteq \mathcal{G}^*$, $|\mathcal{I}| \leq m$, $F \cup N = \mathcal{I}$, $F \cap N = \emptyset$ and $F \neq \emptyset$,

$$P\big(j \in \mathcal{I} \lhd \mathcal{G}_S^*,\ T(d, F, N) \leq 2q(F, N) \log(p),\ A_p\big) \leq L_p p^{-\rho_j^*} + o(1/p). \tag{B.51}$$

Now, we show (B.51). The following lemma is proved below.

Lemma B.9. *For each fixed $(\mathcal{I}, F, N)$ such that $\mathcal{I} = F \cup N$, $F \cap N = \emptyset$, $F \neq \emptyset$ and $|\mathcal{I}| \leq m$, there exists a random variable T_0 such that with probability at least $1 - o(1/p)$, $|T(d, F, N) - T_0| \leq C(\log(p))^{1/\alpha}$, and conditioning on $\beta^{\mathcal{I}}$, T_0 has a non-central χ^2-distribution with the degree of freedom $k \leq |\mathcal{I}|$ and the non-centrality parameter*

$$\delta_0 = (\beta^F)' \left[Q^{F,F} - Q^{F,N} (Q^{N,N})^{-1} Q^{N,F}\right] \beta^F,$$

where Q is as defined in (2.17).

Fix a triplet $(\mathcal{I}, F, N)$ and let δ_0 be as in Lemma B.9. Then

$$\begin{aligned}
\text{(B.52)}\quad & P\big(\mathcal{I} \lhd \mathcal{G}_S^*,\ T(d,F,N) \le 2q(F,N)\log(p),\ A_{p,j}\big)\\
&\le P\big(\mathcal{I} \lhd \mathcal{G}_S^*,\ T_0 \le 2q(F,N)\log(p) + C(\log(p))^{1/\alpha}\big) + o(1/p)\\
&\le P\big(\mathcal{I} \lhd \mathcal{G}_S^*\big)\cdot P\big(T_0 \le 2q(F,N)\log(p) + C(\log(p))^{1/\alpha}\ \big|\beta^{\mathcal{I}}\big) + o(1/p).
\end{aligned}$$

Denote $\omega_0 = \tau_p^{-2}\delta_0$. By Lemma B.9, $(T_0|\beta^{\mathcal{I}}) \sim \chi_k^2(2r\omega_0\log(p))$, where $k \le m$. In addition, $(\log(p))^{1/\alpha} \ll \log(p)$ by recalling that $\alpha > 1$. Combining these and using the basic property of non-central χ^2-distributions,

$$P\big(T_0 \le 2q(F,N)\log(p) + C(\log(p))^{1/\alpha}\ \big|\beta^{\mathcal{I}}\big) \le L_p p^{-[(\sqrt{\omega_0 r} - \sqrt{q(F,N)})_+]^2}.$$

Inserting this into (B.52) and noting that $P\big(\mathcal{I} \lhd \mathcal{G}_S^*\big) \le L_p p^{-|\mathcal{I}|\vartheta}$, we have

$$P\big(\mathcal{I}\lhd\mathcal{G}_S^*,\ T(d,F,N) \le 2q(F,N)\log(p),\ A_p\big) \le L_p p^{-|\mathcal{I}|\vartheta - [(\sqrt{\omega_0 r} - \sqrt{q(F,N)})_+]^2} + o(1/p).$$

Comparing this with (B.51) and using the expression of ρ_j^* in Lemma B.3, it suffices to show

$$\text{(B.53)}\qquad |\mathcal{I}|\vartheta + [(\sqrt{\omega_0 r} - \sqrt{q(F,N)})_+]^2 \ge \psi(F,N).$$

Recall that $q = q(F,N)$ is chosen from either (2.33) or (2.38). In the former case, since $\omega_0 \ge \tilde{\omega}(F,N)$ by definition (see (2.37)), it follows immediately from (2.33) and (2.34) that (B.53) holds. Therefore, we only consider the latter, in which case $q(F,N) = \tilde{q}|F|$ and (B.53) reduces to

$$\text{(B.54)}\qquad |\mathcal{I}|\vartheta + [(\sqrt{\omega_0 r} - \sqrt{\tilde{q}|F|})_+]^2 \ge \psi(F,N).$$

By the expression of $\psi(F,N)$,

$$\psi(F,N) \le (|\mathcal{I}| - |F|/2)\vartheta + (\omega r/4 + \vartheta/2) \le |\mathcal{I}|\vartheta + \omega r/4,$$

where ω is a shorthand of $\omega(F,N)$. Therefore, to show (B.54), it suffices to check

$$\text{(B.55)}\qquad \big(\sqrt{\omega_0 r} - \sqrt{\tilde{q}|F|}\ \big)_+ \ge \sqrt{\omega r}/2.$$

Towards this end, recalling that $F \subset \mathcal{I}$, we let Σ and $\tilde{\Sigma}$ be the respective submatrices of $(G^{\mathcal{I},\mathcal{I}})^{-1}$ and Q^{-1} formed by restricting the rows and columns from $\mathcal{I}$ to F. Let $\xi^* = \tau_p^{-1}\beta^F$. By elementary calculation and noting that $a > a_g^*(G)$,

$$\omega = \min_{\xi \in \mathbb{R}^{|F|}: 1 \le |\xi_i| \le a} \xi'\Sigma^{-1}\xi, \qquad \omega_0 = (\xi^*)'\tilde{\Sigma}^{-1}\xi^*.$$

On one hand, since $G \in \mathcal{M}_p^*(\gamma, g, c_0, A_1)$ and $|\mathcal{I}| \leq m \leq g$,

$$\omega \geq |F| \cdot \lambda_{\min}(G^{\mathcal{I},\mathcal{I}}) \geq c_0 \cdot |F|. \tag{B.56}$$

On the other hand, noting that $\|\xi^*\|_\infty \leq a$,

$$|\omega - \omega_0| \leq \max_{\xi \in \mathbb{R}^{|F|}: 1 \leq |\xi_i| \leq a} |\xi'(\Sigma^{-1} - \tilde{\Sigma}^{-1})\xi| \leq (a^2 \cdot \|\Sigma^{-1} - \tilde{\Sigma}^{-1}\|)|F|. \tag{B.57}$$

We argue that $\|\Sigma^{-1} - \tilde{\Sigma}^{-1}\|$ can be taken to be sufficiently small by ℓ^{ps} sufficiently large. To see the point, note that $\|\Sigma^{-1} - \tilde{\Sigma}^{-1}\| \leq \|G^{\mathcal{I},\mathcal{I}}\|^2 \|Q^{-1}\|^2 \|G^{\mathcal{I},\mathcal{I}} - Q\|$. First, since $|\mathcal{I}| \leq m$, $\|G^{\mathcal{I},\mathcal{I}}\|^2 \leq C$. Second, note that Q is the Fisher Information Matrix associated with the model $d^{\mathcal{I}^{ps}} \sim N(B^{\mathcal{I}^{ps},\mathcal{I}}\beta^{\mathcal{I}}, H^{\mathcal{I}^{ps},\mathcal{I}^{ps}})$. Using Lemma 2.2 and (B.77), $\|G^{\mathcal{I},\mathcal{I}} - Q\| \leq C(\ell^{ps})^{-\gamma}$. Third, $\|(G^{\mathcal{I},\mathcal{I}})^{-1}\| \leq c_0^{-1}$, since $G \in \mathcal{M}_p^*(\gamma, g, c_0, A_1)$. Finally, $\|Q^{-1}\| \leq 2c_0^{-1}$ when $G^{\mathcal{I},\mathcal{I}}$ and Q are sufficiently close. Combining these gives that $\|\Sigma^{-1} - \tilde{\Sigma}^{-1}\| \leq C(\ell^{ps})^{-\gamma}$, for sufficiently large ℓ^{ps}, and the claim follows.

As a result, by taking ℓ^{ps} a sufficiently large constant integer, we have

$$a^2 \|\Sigma^{-1} - \tilde{\Sigma}^{-1}\| \leq \big(\frac{1}{2}\sqrt{c_0} - \sqrt{\tilde{q}/r}\big)^2, \tag{B.58}$$

where we note the right hand side is a fixed positive constant. Combining (B.56)-(B.58),

$$\sqrt{(\omega - \omega_0)_+} \leq \big(\frac{1}{2}\sqrt{c_0} - \sqrt{\tilde{q}/r}\big)\sqrt{|F|} \leq \frac{1}{2}\sqrt{\omega} - \sqrt{\tilde{q}|F|/r},$$

where the first inequality follows from (B.57) and (B.58), as well as the fact that $\tilde{q} < c_0 r/4$ (so that $\frac{1}{2}\sqrt{c_0} - \sqrt{\tilde{q}/r} > 0$); and the last inequality follows from (B.56). Combining this to the well known inequality that $\sqrt{a} + \sqrt{(b-a)_+} \geq \sqrt{b}$ for any $a, b \geq 0$, we have

$$\sqrt{\omega_0} \geq \sqrt{\omega} - \sqrt{(\omega - \omega_0)_+} \geq \sqrt{\omega} - \big(\frac{1}{2}\sqrt{\omega} - \sqrt{\tilde{q}|F|/r}\big) \geq \frac{1}{2}\sqrt{\omega} + \sqrt{\tilde{q}|F|/r},$$

and (B.55) follows directly. □

B.7.1. *Proof of Lemma B.9.* Recall that $T(d, F, N) = W'Q^{-1}W - W_N'(Q_{N,N})^{-1}W_N$ where $d = DY$, W and Q are defined in (2.17) which depend on $\mathcal{I} = F \cup N$, and W_N and $Q_{N,N}$ are defined in (2.18). Let $V = S(\beta) \backslash \mathcal{I}$. By definitions,

$$W = Q\beta^{\mathcal{I}} + \xi + u, \quad \text{where } \xi = (B^{\mathcal{I}^{ps},\mathcal{I}})'(H^{\mathcal{I}^{ps},\mathcal{I}^{ps}})^{-1}B^{\mathcal{I}^{ps},V}\beta^V \text{ and } u \sim N(0, Q).$$

Denote $\tilde{W} = Q\beta^{\mathcal{I}} + u$. Introduce a proxy of $T(d, F, N)$ by

$$T_0(d, F, N) = \tilde{W}'Q^{-1}\tilde{W} - (\tilde{W}^N)'(Q^{N,N})^{-1}\tilde{W}^N.$$

Write for short $T = T(d, F, N)$ and $T_0 = T_0(d, F, N)$. To show the claim, it is sufficient to show (a) $|T - T_0| \leq C(\log(p))^{1/\alpha}$ with probability at least $1 - o(1/p)$ and (b) $(T_0|\beta^{\mathcal{I}}) \sim \chi^2_k(\delta_0)$.

Consider (a) first. By direct calculations,

$$|T - T_0| \leq 2\|\xi\| \cdot \big(2\|\beta^{\mathcal{I}}\| + \|Q^{-1}\|\|\xi\| + 2\|Q^{-1/2}\|\|Q^{-1/2}u\|\big). \tag{B.59}$$

First, since $|\mathcal{I}| \leq m$ and $\|\beta\|_\infty \leq a\tau_p \leq C\sqrt{\log(p)}$, $\|\beta^{\mathcal{I}}\| \leq C\sqrt{\log(p)}$. Second, by definitions, $\max\{\|Q^{-1/2}\|, \|Q^{-1}\|\} \leq C$. Last, note that $Q^{-1/2}u \sim N(0, I_{|\mathcal{I}|})$ and so with probability at least $1 - o(1/p)$, $\|Q^{-1/2}u\| \leq C\sqrt{\log(p)}$. Inserting these into (B.59), we have that with probability at least $1 - o(1/p)$,

$$|T - T_0| \leq C\|\xi\|\big(\sqrt{\log(p)} + \|\xi\|\big).$$

We now study $\|\xi\|$. By definitions, it is seen that

$$\|\xi\| \leq \|B^{\mathcal{I}^{ps},\mathcal{I}}\| \cdot \|(H^{\mathcal{I}^{ps},\mathcal{I}^{ps}})^{-1}\| \cdot \|B^{\mathcal{I}^{ps},V}\beta^V\|.$$

First, we have $\|B^{\mathcal{I}^{ps},\mathcal{I}}\| \leq \|B\| \leq C$. Second, since $|\mathcal{I}^{ps}| \leq C$, by RCB, $\lambda_{\min}(H^{\mathcal{I}^{ps},\mathcal{I}^{ps}}) \geq c_1|\mathcal{I}^{ps}|^{-\kappa} \geq C > 0$, and so $\|(H^{\mathcal{I}^{ps},\mathcal{I}^{ps}})^{-1}\| \leq C$. Third, by basic algebra,

$$\|B^{\mathcal{I}^{ps},V}\beta^V\| \leq \sqrt{|\mathcal{I}^{ps}|} \cdot \|B^{\mathcal{I}^{ps},V}\beta^V\|_\infty \leq C\|B^{\mathcal{I}^{ps},V}\|_\infty \cdot \|\beta^V\|_\infty. \tag{B.60}$$

Here, we note that $\|B^{\mathcal{I}^{ps},V}\|_\infty \leq \|B - B^{**}\|_\infty$, where B^{**} is defined in Section B.1, and where by Lemma B.1, $\|B - B^{**}\|_\infty \leq C(\log(p))^{-(1-1/\alpha)}$. As a result, $\|B^{\mathcal{I}^{ps},V}\|_\infty \leq C(\log(p))^{-(1-1/\alpha)}$. Inserting this into (B.60) and recalling that $\|\beta^V\|_\infty \leq C\sqrt{\log(p)}$,

$$\|B^{\mathcal{I}^{ps},V}\beta^V\| \leq C(\log(p))^{-(1-1/\alpha)} \cdot \sqrt{\log(p)} = C(\log(p))^{1/\alpha-1/2}.$$

Combining these gives that $\|\xi\| \leq C(\log(p))^{1/\alpha-1/2}$. This, together with (B.59), implies that

$$|T-T_0| \leq C(\log(p))^{1/\alpha-1/2}[\sqrt{\log(p)}+(\log(p))^{1/\alpha-1/2}] \leq C[(\log(p))^{1/\alpha}+(\log(p))^{2/\alpha-1}],$$

and the claim follows by recalling $\alpha > 1$.

Next, consider (b). Write for short $R = (H^{\mathcal{I}^{ps},\mathcal{I}^{ps}})^{-1/2}B^{\mathcal{I}^{ps},\mathcal{I}}$. Also, recall that F and N are subsets of $\mathcal{I}$. We let R_F and R_N be the submatrices of R by restricting the columns to F and N, respectively (no restriction on the rows). By definitions, $Q = R'R$ and $u \sim N(0, Q)$, so that we can rewrite $u = R'\tilde{z}$ for some random vector $\tilde{z} \sim N(0, I_{|\mathcal{I}^{ps}|})$. With these notations, we can rewrite T_0 as

$$T_0 = (R\beta^{\mathcal{I}} + \tilde{z})'[R(R'R)^{-1}R' - R_N(R_N'R_N)^{-1}R_N'](R\beta^{\mathcal{I}} + \tilde{z}).$$

Therefore, $(T_0|\beta^{\mathcal{I}}) \sim \chi^2_k(\tilde{\delta}_0)$ (Lehmann and Casella, 1998), where $k = \mathrm{rank}(R) - \mathrm{rank}(R_N) \leq |\mathcal{I}|$, and

$$\tilde{\delta}_0 \equiv (R\beta^{\mathcal{I}})'[R(R'R)^{-1}R' - R_N(R_N'R_N)^{-1}R_N'](R\beta^{\mathcal{I}}).$$

By basic algebra, $\tilde{\delta}_0 = \delta_0$. This completes the proof. □

B.8. Proof of Lemma A.2. Viewing $\mathcal{U}_p^*$ as a subgraph of $\mathcal{G}^+$, we recall that $\mathcal{I} \lhd \mathcal{U}_p^*$ stands for that $\mathcal{I}$ is a component of $\mathcal{U}_p^*$. The assertion of Lemma A.2 is that there exists a constant integer l_0 such that

$$P\big(|\mathcal{I}| > l_0 \text{ for some } \mathcal{I} \lhd \mathcal{U}_p^*\big) = o(1/p). \tag{B.61}$$

The key to show the claim is the following lemma, which is proved below:

LEMMA B.10. *There is an event A_p and a constant $C_1 > 0$ such that $P(A_p^c) = o(1/p)$ and that over the event A_p, $\|d^{\mathcal{I}^{ps}}\|^2 \geq 5C_1|\mathcal{I}|\log(p)$ for all $\mathcal{I} \lhd \mathcal{U}_p^*$.*

By Lemma B.10, to show (B.61), it suffices to show

$$P\big(|\mathcal{I}| > l_0 \text{ for some } \mathcal{I} \lhd \mathcal{U}_p^*,\ A_p\big) = o(1/p). \tag{B.62}$$

Now, for each $1 \leq j \leq p$, there is a unique component $\mathcal{I}$ such that $j \in \mathcal{I} \lhd \mathcal{U}_p^*$. Such $\mathcal{I}$ is random, but any of its realization is a connected subgraph of $\mathcal{G}^+$. Therefore,

$$P\big(|\mathcal{I}| > l_0 \text{ for some } \mathcal{I} \lhd \mathcal{U}_p^*,\ A_p\big) \leq \sum_{j=1}^{p} \sum_{l=l_0+1}^{\infty} \sum_{\mathcal{I}: j \in \mathcal{I} \trianglelefteq \mathcal{G}^+, |\mathcal{I}|=l} P\big(j \in \mathcal{I} \lhd \mathcal{U}_p^*,\ A_p\big), \tag{B.63}$$

where on the right hand side we have changed the meaning of $\mathcal{I}$ to denote a fixed (non-random) connected subgraph of $\mathcal{G}^+$. We argue that

(a) for each (j,l), the third summation on the right of (B.63) sums over no more than L_p terms;
(b) there are constants $C_2, C_3 > 0$ such that for any $(j, \mathcal{I})$ satisfying $j \in \mathcal{I} \trianglelefteq \mathcal{G}^+$, $P\big(j \in \mathcal{I} \lhd \mathcal{U}_p^*,\ A_p\big) \leq L_p\big[p^{-C_2\sqrt{|\mathcal{I}|}} + p^{-C_3|\mathcal{I}|}\big]$.

Once (a) and (b) are proved, then it follows from (B.63) that

$$P\big(|\mathcal{I}| > l_0 \text{ for some } \mathcal{I} \lhd \mathcal{U}_p^*,\ A_p\big) \leq L_p\big[p^{1-C_2\sqrt{l_0}} + p^{1-C_3 l_0}\big],$$

and (B.62) follows by taking ℓ_0 sufficiently large.

It remains to show (a) and (b). Consider (a) first. Note that the number of connected subgraph $\mathcal{I}$ of size l such that $j \in \mathcal{I} \trianglelefteq \mathcal{G}^+$ is bounded by $C(eK_p^+)^l$ (Frieze and Molloy, 1999), where K_p^+ is the maximum degree of $\mathcal{G}^+$. At the same time, by Lemma B.1 and Lemma B.2, K_p^+ is an L_p term. Combining these gives (a).

Consider (b). Denote $V = \{j : B^{**}(i,j) \neq 0, \text{ for some } i \in \mathcal{I}^{ps}\}$, where B^{**} is defined in Section B.1. Write for short $d_1 = d^{\mathcal{I}^{ps}}$, $B_1 = B^{\mathcal{I}^{ps},V}$ and $H_1 = H^{\mathcal{I}^{ps},\mathcal{I}^{ps}}$. With these notations and by Lemma B.10, (b) reduces to

$$P\big(j \in \mathcal{I} \lhd \mathcal{U}_p^*,\ \|d_1\|^2 \geq 5C_1|\mathcal{I}|\log(p)\big) \leq L_p\big[p^{-C_2\sqrt{|\mathcal{I}|}} + p^{-C_3|\mathcal{I}|}\big]. \tag{B.64}$$

We now show (B.64). Note that $d_1 = B_1\beta^V + \xi + \tilde{z}$, where $\xi = [(B - B^{**})\beta]^{\mathcal{I}^{ps}}$ and $\tilde{z} \sim N(0, H_1)$. For preparation, we claim that

$$\|\xi\|^2 = |\mathcal{I}| \cdot o(\log(p)). \tag{B.65}$$

In fact, first since ℓ^{ps} is finite, $|\mathcal{I}^{ps}| \leq C|\mathcal{I}|$ and it follows that $\|\xi\|^2 \leq C|\mathcal{I}| \cdot \|\xi\|_\infty^2$. Second, by Lemma B.1, $\|B - B^{**}\|_\infty = o(1)$. Since $\|\beta\|_\infty \leq a\tau_p \leq C\sqrt{\log(p)}$, it follows that $\|\xi\|_\infty \leq \|B - B^{**}\|_\infty \|\beta\|_\infty = o(\sqrt{\log(p)})$. Combining these gives (B.65).

Now, combining (B.64) and (B.65) and using the well-known inequality $(a+b)^2 \leq 2a^2 + 2b^2$ for $a, b \in \mathbb{R}$, we find that for sufficiently large p,

$$\begin{aligned} & P\big(j \in \mathcal{I} \lhd \mathcal{U}_p^*,\ \|d_1\|^2 \geq 5C_1|\mathcal{I}|\log(p)\big) \\ \leq\ & P\big(j \in \mathcal{I} \lhd \mathcal{U}_p^*,\ \|B_1\beta^V + \tilde{z}\|^2 \geq 4C_1|\mathcal{I}|\log(p)\big) \\ \leq\ & P\big(j \in \mathcal{I} \lhd \mathcal{U}_p^*,\ \|B_1\beta^V\|^2 + \|\tilde{z}\|^2 \geq 2C_1|\mathcal{I}|\log(p)\big) \\ \leq\ & P\big(\|B_1\beta^V\|^2 \geq C_1|\mathcal{I}|\log(p)\big) + P\big(\|\tilde{z}\|^2 \geq C_1|\mathcal{I}|\log(p)\big) \equiv I + II. \end{aligned} \tag{B.66}$$

We now analyze I and II separately. Consider I first. We claim there is a constant $C_4 > 0$, not depending on $|\mathcal{I}|$, such that $\|B_1\beta^V\| \leq \sqrt{C_4\log(p)}\|\beta^V\|_0$. To see this, note that $\|B_1\beta^V\| \leq \|B_1\beta^V\|_1 \leq \|B_1\|_1\|\beta^V\|_1$, where $\|B_1\|_1 \leq \|B\|_1 \leq C$, with $C > 0$ a constant independent of $|\mathcal{I}|$. At the same time, $\|\beta^V\|_1 \leq a\tau_p\|\beta^V\|_0$. So the argument holds for $C_4 = 2ra^2C^2$. Additionally, $\|\beta^V\|_0$ has a multinomial distribution, where the number of trials is $|V| \leq L_p$ and the success probability is $\epsilon_p = p^{-\vartheta}$. Combining these, we have

$$I \leq P\big(\|\beta^V\|_0 \geq \sqrt{(C_1/C_4)|\mathcal{I}|}\big) \leq L_p p^{-\vartheta\lfloor\sqrt{(C_1/C_4)|\mathcal{I}|}\rfloor}, \tag{B.67}$$

where $\lfloor x \rfloor$ denotes the the largest integer k such that $k \leq x$.

Next, consider II. Note that $\|H_1\| \leq \|H\| \leq C_5$, where $C_5 > 0$ is a constant independent of $|\mathcal{I}|$. It follows that $\|\tilde{z}\|^2 \leq C_5^{-1}\|H_1^{-1/2}\tilde{z}\|^2$, where

$\|H_1^{-1/2}\tilde{z}\|^2$ has a χ^2-distribution with degree of freedom $|\mathcal{I}^{ps}| \leq C|\mathcal{I}|$. Using the property of χ^2-distributions,

$$II \leq P\big(\|H_1^{-1/2}\tilde{z}\|^2 \geq C_1C_5|\mathcal{I}|\log(p)\big) \leq L_p p^{-(C_1C_5/2)|\mathcal{I}|}. \tag{B.68}$$

Inserting (B.67) and (B.68) into (B.66), (B.64) follows by taking $C_2 > \vartheta\sqrt{C_1/C_4}$ and $C_3 > C_1C_5/2$. □

B.8.1. *Proof of Lemma B.10.* For preparation, we need some notations. First, for a constant $\delta_0 > 0$ to be determinded, define the $p \times p$ matrices $\tilde{B}$ and $\tilde{H}$ by

$$\tilde{B}(i,j) = B(i,j)1\{|B(i,j)| > \delta_0\}, \qquad \tilde{H}(i,j) = H(i,j)1\{|H(i,j)| > \delta_0\}.$$

Second, view $\mathcal{U}_p^*$ as a subgraph of $\mathcal{G}^+$. Note that in the PS-step, each $\mathcal{G}_t$ is a connected subgraph of $\mathcal{G}^+$. Hence, any $\mathcal{G}_t$ that passed the test must be contained as a whole in one component of $\mathcal{U}_p^*$. It follows that for any $\mathcal{I} \lhd \mathcal{U}_p^*$, there exists a (random) set $\mathcal{T} \subset \{1, \cdots, T\}$ such that $\mathcal{I} = \cup_{t\in\mathcal{T}}\mathcal{G}_t$. Therefore, we write

$$\mathcal{I} = \cup_{i=1}^{\hat{s}_0} V_i,$$

where each $V_i = \mathcal{G}_t$ for some $t \in \mathcal{T}$, and these V_i's are listed in the order they were tested. Denote $\hat{N}_i = \mathcal{U}^{(t-1)} \cap \mathcal{G}_t$ and $\hat{F}_i = \mathcal{G}_t \backslash \hat{N}_i$. Let $W_{(i)}$ and $Q_{(i)}$ be the vector W and matrix Q in (2.17). From basic algebra, the test statistic can be rewritten as

$$T(d, \hat{F}_i, \hat{N}_i) = \|u_{(i)}\|^2, \qquad u_{(i)} \equiv \Sigma_{(i)}^{-1/2}[W_{(i)}^{\hat{F}_i} - Q_{(i)}^{\hat{F}_i,\hat{N}_i}(Q_{(i)}^{\hat{N}_i,\hat{N}_i})^{-1}W_{(i)}^{\hat{N}_i}], \tag{B.69}$$

where $\Sigma_{(i)} = Q_{(i)}^{\hat{F}_i,\hat{F}_i} - Q_{(i)}^{\hat{F}_i,\hat{N}_i}[Q_{(i)}^{\hat{N}_i,\hat{N}_i}]^{-1}Q_{(i)}^{\hat{N}_i,\hat{F}_i}$.

Third, define

$$W_{(i)}^* = (\tilde{B}^{V_i^{ps},V_i})'(\tilde{H}^{V_i^{ps},V_i^{ps}})^{-1}d^{V_i^{ps}},$$

and $u_{(i)}^*$ as in (B.69) with $W_{(i)}$ replaced by $W_{(i)}^*$. Let u be the $|\mathcal{I}| \times 1$ vector by putting $\{u_{(i)}, 1 \leq i \leq \hat{s}_0\}$ together, and define u^* similarly.

With these notations, to show the claim, it suffices to show there exist positive constants C_6, C_7 such that with probability at least $1 - o(1/p)$, for any $\mathcal{I} \lhd \mathcal{U}_p^*$,

$$\|u^*\|^2 \geq C_6|\mathcal{I}|\log(p), \tag{B.70}$$

and

$$\|u^*\|^2 \leq C_7\|d^{\mathcal{I}^{ps}}\|^2. \tag{B.71}$$

Consider (B.70) first. Since each V_i passed the test, $\|u_{(i)}\|^2 \geq t(\hat{F}_i, \hat{N}_i)$. If $t(\hat{F}_i, \hat{N}_i)$ is chosen from (2.33), $t(\hat{F}_i, \hat{N}_i) \geq 2q_0 \log(p) \geq 2(q_0/m)|\hat{F}_i| \log(p)$; otherwise it is chosen from (2.38), then $t(\hat{F}_i, \hat{N}_i) \geq 2\tilde{q}|\hat{F}_i| \log(p)$. In both cases, there is a constant $q > 0$ such that

$$\|u_{(i)}\|^2 \geq 2q|\hat{F}_i| \log(p), \qquad 1 \leq i \leq \hat{s}_0.$$

In addition, it is easy to see that $\cup_i \hat{F}_i$ is a partition of $\mathcal{I}$. It follows that

$$\|u\|^2 = \sum_{i=1}^{\hat{s}_0} \|u_{(i)}\|^2 \geq 2q|\mathcal{I}| \log(p). \tag{B.72}$$

At the same time, let A_p be the event $\{\|d\|_\infty \leq C_0\sqrt{\log(p)}\}$, where we argue that when C_0 is sufficiently large, $P(A_p^c) = o(1/p)$. To see this, recall that $d = B\beta + H^{1/2}\tilde{z}$, where $\tilde{z} \sim N(0, I_p)$. By the assumptions, $\|B\|_\infty \leq C$, $\|\beta\|_\infty \leq C\sqrt{\log(p)}$ and $\|H\|_\infty \leq C$. Therefore, $\|d\|_\infty \leq C(\sqrt{\log(p)} + \|\tilde{z}\|_\infty)$. It is well-known that $P(\|\tilde{z}\|_\infty > \sqrt{2a\log(p)}) = L_p p^{-a}$ for any $a > 0$. Hence, when C_0 is sufficiently large, $P(A_p^c) = o(1/p)$.

We shall show that over the event A_p, by choosing δ_0 a sufficiently small constant,

$$\|u - u^*\|^2 \leq q|\mathcal{I}| \log(p)/2. \tag{B.73}$$

Once this is proved, combining (B.72) and (B.73), and applying the inequality $(a + b)^2 \leq 2(a^2 + b^2)$ for any $a, b \in \mathbb{R}$, we have

$$2q|\mathcal{I}| \log(p) \leq \|u\|^2 \leq 2(\|u^*\|^2 + \|u - u^*\|^2) \leq 2\|u^*\|^2 + q|\mathcal{I}| \log(p).$$

Hence, (B.70) holds with $C_6 = q/2$.

What remains is to prove (B.73). It follows from $G \in \mathcal{M}_p^*(\gamma, g, c_0, A_1)$ and $|V_i| \leq m \leq g$ that $\|(G^{V_i,V_i})^{-1}\| \leq c_0^{-1}$. As a result, $\|Q_{(i)}^{-1}\| \leq C$. Also, $\Sigma_{(i)}^{-1}$ is a submatrix of $Q_{(i)}^{-1}$; and hence $\|\Sigma_{(i)}^{-1}\| \leq C$. This implies

$$\|u_{(i)} - u_{(i)}^*\| \leq C\|W_{(i)} - W_{(i)}^*\|, \qquad 1 \leq i \leq \hat{s}_0. \tag{B.74}$$

Since B enjoys a polynomial off-diagonal decay with rate α, $\|(B-\tilde{B})^{V_i^{ps},V_i}\|_\infty \leq C\delta_0^{1-1/\alpha}$. Noting that $|V_i^{ps}| \leq C$, this implies $\|(B - \tilde{B})^{V_i^{ps},V_i}\| \leq C\delta_0^{1-1/\alpha}$. Similarly, we can derive $\|(H - \tilde{H})^{V_i^{ps},V_i^{ps}}\| \leq C\delta_0^{1-1/\alpha}$. These together imply

$$\|W_{(i)} - W_{(i)}^*\| \leq C\delta_0^{1-1/\alpha}\|d^{V_i^{ps}}\| \leq C\delta_0^{1-1/\alpha}\|d\|_\infty, \quad 1 \leq i \leq \hat{s}_0, \tag{B.75}$$

where in the last inequality we use the facts that $|V_i^{ps}| \leq C$ and $\|d^{V_i^{ps}}\|_\infty \leq \|d\|_\infty$. Combining (B.74) and (B.75), over the event A_p,

$$\|u_{(i)} - u^*_{(i)}\|^2 \leq C\delta_0^{2(1-1/\alpha)} \log(p), \qquad 1 \leq i \leq \hat{s}_0.$$

Noting that $\alpha > 1$, we can choose a sufficiently small δ_0 such that $C\delta_0^{2(1-1/\alpha)} \leq q/2$, and (B.73) follows by noting $|\hat{s}_0| \leq |\mathcal{I}|$.

Next, consider (B.71). We write

$$u^* = \Xi\Gamma\Theta \; d^{\mathcal{I}^{ps}},$$

where the matrices Ξ, Γ and Θ are defined as follows: Ξ is a block-wise diagonal matrix with the i-th block equals to $\Sigma_{(i)}^{-1}$. Γ is a $|\mathcal{I}| \times (\sum_{i=1}^{\hat{s}_0} |V_i|)$ matrix, with the $(\hat{F}_i, V_i)$-block is given by

$$\Gamma^{\hat{F}_i, V_i} = \big[I, \; -Q_{(i)}^{\hat{F}_i, \hat{N}_i} (Q_{(i)}^{\hat{N}_i, \hat{N}_i})^{-1}\big].$$

and 0 elsewhere. Θ is a $(\sum_{i=1}^{\hat{s}_0} |V_i|) \times |\mathcal{I}^{ps}|$ matrix, with the (V_i, V_i^{ps})-block

$$\Theta^{V_i, V_i^{ps}} = (\tilde{B}^{V_i^{ps}, V_i})' (\tilde{H}^{V_i^{ps}, V_i^{ps}})^{-1},$$

and 0 elsewhere.

Note that these matrices are random (they depend on $\mathcal{U}_p^*$ and $\mathcal{I}$). Below, we show that for any realization of $\mathcal{U}_p^*$ and any component $\mathcal{I} \lhd \mathcal{U}_p^*$,

$$\|\Xi\Gamma\Theta\| \leq C. \tag{B.76}$$

Once (B.76) is proved, (B.71) follows by letting $C_7 = C^2$.

We now show (B.76). Since $\|\Xi\Gamma\Theta\| \leq \|\Xi\| \|\Gamma\| \|\Theta\|$, it suffices to show

$$\|\Xi\|, \|\Gamma\|, \|\Theta\| \leq C.$$

First, $\|\Xi\| \leq \max_i \|Q_{(i)}^{-1}\| \leq C$. Second, the entries in Γ and Θ have a uniform upper bound in magnitude, and each row and column of Γ has $\leq m$ non-zero entries. So $\|\Gamma\| \leq C$. Finally, each row of Θ has no more than $2m\ell^{ps}$ entries; as a result, to show $\|\Theta\| \leq C$, we only need to prove that each column of Θ also has a bounded number of non-zero entries.

Towards this end, write for short $\tilde{B}_{(i)} = \tilde{B}^{V_i^{ps}, V_i}$ and $\tilde{H}_{(i)} = \tilde{H}^{V_i^{ps}, V_i^{ps}}$ for each $1 \leq i \leq \hat{s}_0$. By definition,

$$\Theta(k, j) = \sum_{j' \in V_i^{ps}} \tilde{B}_{(i)}(j', k) \tilde{H}_{(i)}^{-1}(j', j), \qquad k \in V_i, \; j \in V_i^{ps}.$$

First, given the chosen δ_0, each row or column of $\tilde{B}$ and $\tilde{H}$ has $\leq L_0$ non-zero entries, where L_0 is a constant integer. Therefore, for each j', the number of k such that $\tilde{B}(j',k) \neq 0$ is upper bounded by L_0. Second, we define a graph $\mathcal{G} = \mathcal{G}(\delta_0)$ where there is an edge between nodes j and j' if and only if $\tilde{H}(j,j') \neq 0$. For each $1 \leq i \leq \hat{s}_0$, let $\mathcal{G}_i$ be the restriction of $\mathcal{G}$ to the nodes in V_i^{ps}. We see that $\tilde{H}_{(i)}$ is block-diagonal with each block corresponding to a component of $\mathcal{G}_i$, and so is $(\tilde{H}_{(i)})^{-1}$. This means $(\tilde{H}_{(i)})^{-1}(j',j)$ can be non-zero only when j and j' belong to the same component of $\mathcal{G}_i$. Since $|V_i^{ps}| \leq 2m\ell^{ps}$ for all i, necessarily, there exits a path in $\mathcal{G}$ of length $\leq 2m\ell^{ps}$ that connects j and j'. Third, since $\mathcal{G}$ is L_0-sparse, for each j, the number of j' that is connected to j with a path of length $\leq 2m\ell^{ps}$ is upper bounded by $L_0^{2m\ell^{ps}}$. In summary, for each fixed j, there are no more than $L_0 \cdot L_0^{2m\ell^{ps}}$ nodes k such that $\Theta(k,j) \neq 0$, i.e., each column of Θ has $\leq L_0^{2m\ell^{ps}+1}$ nonzero entries and the claim follows. □

B.9. Proof of Lemma A.3. Fix $\mathcal{I}$ and recall that $\mathcal{J} = \{j : D(i,j) \neq 0, \text{ for some } i \in \mathcal{I}\}$. In this lemma, $\mathcal{I}^{pe}$ is as in Definition 2.6, but $\mathcal{J}^{pe}$ is redefined as $\mathcal{J}^{pe} = \{j : D(i,j) \neq 0, \text{ for some } i \in \mathcal{I}^{pe}\}$. Denote $M = |\mathcal{J}^{pe}| - |\mathcal{I}^{pe}|$ and write $G^{\mathcal{J}^{pe},\mathcal{J}^{pe}} = G_1$ for short. Let $\mathcal{F}$ be the mapping from $\mathcal{J}^{pe}$ to $\{1, \cdots, |\mathcal{J}^{pe}|\}$ that maps each $j \in \mathcal{J}^{pe}$ to its order in $\mathcal{J}^{pe}$. Denote $\mathcal{I}_1 = \mathcal{F}(\mathcal{I})$. By these notations, the claim reduces to: for any $|\mathcal{J}^{pe}| \times M$ matrix U whose columns contain an orthonormal basis of $Null(\mathcal{I}^{pe}, \mathcal{J}^{pe})$,

$$\left\|[U(U'G_1^{-1}U)^{-1}U']^{\mathcal{I}_1,\mathcal{I}_1}\right\| = o(1).$$

It suffices to show

$$\left\|[U(U'G_1^{-1}U)^{-1}U']^{\mathcal{I}_1,\mathcal{I}_1}\right\| \leq C(\ell^{pe})^{-\gamma}, \tag{B.77}$$

where $\gamma > 0$ is the same as in $\mathcal{M}_p^*(\gamma, g, c_0, A_1)$. In fact, once this is proved, the claim follows by noting that $\ell^{pe} = (\log(p))^{\nu} \to \infty$.

We now show (B.77). By elementary algebra,

$$\left\|[U(U'G_1^{-1}U)^{-1}U']^{\mathcal{I}_1,\mathcal{I}_1}\right\| \leq \left\|(U'G_1^{-1}U)^{-1}\right\| \left\|(UU')^{\mathcal{I}_1,\mathcal{I}_1}\right\|. \tag{B.78}$$

Consider $\left\|(U'G_1^{-1}U)^{-1}\right\|$ first. Since $U'U$ is an identity matrix, we have $\|(U'G_1^{-1}U)^{-1}\| = [\lambda_{\min}(U'G_1^{-1}U)]^{-1} \leq [\lambda_{\min}(G_1^{-1})]^{-1} = \|G_1\|$. Additionally, the assumption $G \in \mathcal{M}_p^*(\gamma, g, c_0, A_1)$ implies that $\|G_1\| \leq A_1 \sum_{j=1}^{|\mathcal{J}^{pe}|} j^{-\gamma} \leq C|\mathcal{J}^{pe}|^{1-\gamma}$. Last, when $|\mathcal{I}| \leq l_0$, $2\ell^{pe} + 1 \leq |\mathcal{J}^{pe}| \leq (2\ell^{pe}+1)l_0$. Combining the above yields

$$\left\|(U'G_1^{-1}U)^{-1}\right\| \leq C(\ell^{pe})^{1-\gamma}. \tag{B.79}$$

Next, consider $\|(UU')^{\mathcal{I}_1,\mathcal{I}_1}\|$. Note that $\|(UU')^{\mathcal{I}_1,\mathcal{I}_1}\| \leq |\mathcal{I}_1| \cdot \max_{i,i'\in\mathcal{I}_1} |(UU')(i,i')|$, where $\max_{i,i'\in\mathcal{I}_1} |U'U(i,i')| \leq M \cdot \max_{i\in\mathcal{I}_1, 1\leq j\leq M} |U(i,j)|^2$. Here $|\mathcal{I}_1| = |\mathcal{I}| \leq l_0$ and $M \leq h|\mathcal{I}| \leq hl_0$. It follows that

$$\|(UU')^{\mathcal{I}_1,\mathcal{I}_1}\| \leq C \max_{i\in\mathcal{I}_1, 1\leq j\leq M} |U(i,j)|^2. \tag{B.80}$$

The following lemma is proved in Appendix C.

Lemma B.11. *Under the conditioins of Lemma A.3, for any $\mathcal{I} \trianglelefteq \mathcal{G}^+$ such that $|\mathcal{I}| \leq l_0$, and any matrix U whose columns form an orthonormal basis of $Null(\mathcal{I}^{pe}, \mathcal{J}^{pe})$,*

$$\max_{i\in\mathcal{F}(\mathcal{I}), 1\leq j\leq |\mathcal{J}^{pe}|-|\mathcal{I}^{pe}|} |U(i,j)|^2 \leq C(\ell^{pe})^{-1}.$$

Using Lemma B.11, it follows from (B.80) that

$$\|(UU')^{\mathcal{I}_1,\mathcal{I}_1}\| \leq C(\ell^{pe})^{-1}. \tag{B.81}$$

Inserting (B.79) and (B.81) into (B.78), we obtain (B.77). □

B.10. Proof of Lemma A.4. Write for short

$$M_1 = \sum_{j=1}^{p} \sum_{\mathcal{I}: j\in\mathcal{I}\trianglelefteq\mathcal{G}^+, |\mathcal{I}|\leq l_0} P\big(\mathcal{I} \lhd \mathcal{U}_p^*,\ A_p \cap E_{p,\mathcal{I}}^c\big), \quad M_2 = \sum_{k=1}^{p} P\big(\beta_k \neq 0,\ k \notin \mathcal{U}_p^*\big).$$

With these notations, the claim reduces to $M_1 \leq L_p \cdot M_2$.

The key is to prove

(a) for each $\mathcal{I} \trianglelefteq \mathcal{G}^+$, over the event $\{\mathcal{I} \lhd \mathcal{U}_p^*, A_p \cap E_{p,\mathcal{I}}^c\}$, it always holds that $(S(\beta) \cap \mathcal{E}(\mathcal{I}^{pe})) \setminus \mathcal{U}_p^* \neq \emptyset$;

(b) for each k, there are no more than L_p different $\mathcal{I}$ such that $\mathcal{I} \trianglelefteq \mathcal{G}^+$, $|\mathcal{I}| \leq l_0$ and $k \in \mathcal{E}(\mathcal{I}^{pe})$.

Once (a) and (b) are proved, the claim follows easily. To see the point, we note that

$$P\left((S(\beta) \cap \mathcal{E}(\mathcal{I}^{pe}))\setminus\mathcal{U}_p^* \neq \emptyset\right) \leq \sum_{k\in\mathcal{E}(\mathcal{I}^{pe})} P\big(\beta_k \neq 0,\ k \notin \mathcal{U}_p^*\big).$$

Combining this with (a), we have

$$M_1 \leq \sum_{j=1}^{p} \sum_{\mathcal{I}: j\in\mathcal{I}\trianglelefteq\mathcal{G}^+, |\mathcal{I}|\leq l_0} \sum_{k\in\mathcal{E}(\mathcal{I}^{pe})} P\big(\beta_k \neq 0,\ k \notin \mathcal{U}_p^*\big).$$

By re-organizing the summation, the right hand side is equal to

$$\sum_{k=1}^{p} \sum_{\mathcal{I}:\mathcal{I}\trianglelefteq\mathcal{G}^{+},|\mathcal{I}|\le l_0,k\in\mathcal{E}(\mathcal{I}^{pe})} |\mathcal{I}|\cdot P\big(\beta_k\neq 0,\ k\notin\mathcal{U}_p^*\big),$$

which $\le L_p\cdot M_2$ by (b), and the claim follows.

We now show (a) and (b). Consider (a) first. Fix $\mathcal{I}\trianglelefteq\mathcal{G}^{+}$. Suppose (a) does not hold, i.e., the following event

$$\big\{\mathcal{I}\lhd\mathcal{U}_p^*,\ (S(\beta)\cap\mathcal{E}(\mathcal{I}^{pe}))\backslash\mathcal{U}_p^*=\emptyset,\ A_p\cap E_{p,\mathcal{I}}^c\big\}$$

is non-empty. View $\mathcal{U}_p^*$ as a subgraph of $\mathcal{G}^{+}$. Applying Lemma B.2 to $V=\mathcal{U}_p^*$, we find that $\mathcal{I}\lhd\mathcal{U}_p^*$ implies $(\mathcal{U}_p^*\backslash\mathcal{I})\cap\mathcal{E}(\mathcal{I}^{pe})=\emptyset$. Therefore, the following event

$$\text{(B.82)}\qquad \big\{(\mathcal{U}_p^*\backslash\mathcal{I})\cap\mathcal{E}(\mathcal{I}^{pe})=\emptyset,\ (S(\beta)\cap\mathcal{E}(\mathcal{I}^{pe}))\backslash\mathcal{U}_p^*=\emptyset,\ A_p\cap E_{p,\mathcal{I}}^c\big\}$$

is non-empty. Note that $\mathcal{I}\subset\mathcal{E}(\mathcal{I}^{pe})$. From basic set operations, $(\mathcal{U}_p^*\backslash\mathcal{I})\cap\mathcal{E}(\mathcal{I}^{pe})=\emptyset$ and $(S(\beta)\cap\mathcal{E}(\mathcal{I}^{pe}))\backslash\mathcal{U}_p^*=\emptyset$ together imply

$$\big(S(\beta)\cap\mathcal{E}(\mathcal{I}^{pe})\big)\subset\mathcal{I}.$$

By definition, this belongs to the event $E_{p,\mathcal{I}}$. Hence, the event in (B.82) is empty, which is a contradiction.

Consider (b) next. Fix k and denote $\mathcal{K}$ the collection of $\mathcal{I}$ satisfying the conditions in (b). Let $V=\big\{1\le i\le p:\ \ k\in\mathcal{E}(\{i\}^{pe})\big\}$. Since $\mathcal{E}(\mathcal{I}^{pe})=\cup_{i\in\mathcal{I}}\mathcal{E}(\{i\}^{pe})$, we observe that

$$\mathcal{K}=\cup_{i\in V}\mathcal{K}_i,\qquad \text{where}\quad \mathcal{K}_i\equiv\big\{\mathcal{I}:\ \mathcal{I}\trianglelefteq\mathcal{G}^{+},\ |\mathcal{I}|\le l_0,\ i\in\mathcal{I}\big\}.$$

Note that by Lemma B.1 and B.2, $\mathcal{G}^*$ is K_p-sparse and $\mathcal{G}^{+}$ is K_p^{+}-sparse, where both K_p and K_p^{+} are L_p terms. First, we bound $|V|$: By definition, $k\in\mathcal{E}(\{i\}^{pe})$ if and only if there exits a node $k'\in\{i\}^{pe}$ such that k' and k are connected by a length-1 path in $\mathcal{G}^*$. Since $\mathcal{G}^*$ is K_p-sparse, given k, the number of such k' is bounded by K_p. In addition, for each k', there are no more than $(2\ell^{pe}+1)$ nodes i such that $k'\in\{i\}^{pe}$. Hence, $|V|\le(2\ell^{pe}+1)K_p$. Second, we bound $\max_{i\in V}|\mathcal{K}_i|$: For each node $i\in V$, there are no more than $C(eK_p^{+})^{l_0}$ connected subgraph of $\mathcal{G}^{+}$ that contain i and have a size $\le l_0$ (Frieze and Molloy, 1999), i.e., $|\mathcal{K}_i|\le C(eK_p^{+})^{l_0}$. Combining the two parts, $|\mathcal{K}|\le K_p(2\ell^{pe}+1)\cdot C(eK_p^{+})^{l_0}$, which is an L_p term. □

B.11. Proof of Lemma A.5. Let $V_1 = S(\beta) \cap \mathcal{E}(\mathcal{I}^{pe})$ and $V_2 = S(\beta) \backslash \mathcal{E}(\mathcal{I}^{pe})$. We have $(B\beta)^{\mathcal{I}^{pe}} = B^{\mathcal{I}^{pe}, V_1} \beta^{V_1} + \zeta$, where $\zeta = B^{\mathcal{I}^{pe}, V_2} \beta^{V_2}$. Note that over the event $E_{p,\mathcal{I}}$, $V_1 \subset \mathcal{I}$. It follows that $B^{\mathcal{I}^{pe}, V_1} \beta^{V_1} = B^{\mathcal{I}^{pe}, \mathcal{I}} \beta^{\mathcal{I}}$. Combining these, to show the claim, it is sufficient to show

$$\|\zeta\| \leq C(\ell^{pe})^{1/2} [\log(p)]^{-(1-1/\alpha)} \tau_p. \tag{B.83}$$

Recall the matrix B^{**} defined in Section B.1. Since $B^{**}(i,j) = 0$ for $j \in V_2$, we have $\|B^{\mathcal{I}^{pe}, V_2}\|_\infty \leq \|B - B^{**}\|_\infty$, where by Lemma B.1, $\|B - B^{**}\|_\infty \leq C[\log(p)]^{-(1-1/\alpha)}$. Moreover, $\|\beta\|_\infty \leq a\tau_p$. Consequently,

$$\|\zeta\|_\infty \leq \|B - B^{**}\|_\infty \|\beta^{V_2}\|_\infty \leq C[\log(p)]^{-(1-1/\alpha)} \tau_p. \tag{B.84}$$

At the same time, note that $|\mathcal{I}^{pe}| \leq l_0(2\ell^{pe} + 1) \leq C\ell^{pe}$. It follows from the Cauchy-Schwartz inequality that $\|\zeta\| \leq \sqrt{|\mathcal{I}^{pe}|} \|\zeta\|_\infty \leq C(\ell^{pe})^{1/2} \|\zeta\|_\infty$. Combining this with (B.84) gives the claim. □

B.12. Proof of Lemma A.6. Fix $(j, V_0, V_1, \mathcal{I})$ and write for short $\rho_j(V_0, V_1) = \rho_j(V_0, V_1; \mathcal{I})$ and $\rho_j^* = \rho_j^*(\vartheta, r, G)$. The goal is to show $\rho_j(V_0, V_1) \geq \rho_j^* + o(1)$. We show this for the case $V_0 \neq V_1$ and the case $V_0 = V_1$ separately.

Consider the first case. By definition, $\rho_j^* \leq \rho(V_0, V_1)$, where $\rho(V_0, V_1)$ is as in (2.28). Therefore, it suffices to show

$$\rho_j(V_0, V_1) = \rho(V_0, V_1) + o(1). \tag{B.85}$$

Introduce the function

$$f(x) = \max\{|V_0|, |V_1|\}\vartheta + \frac{1}{4}\big[\big(\sqrt{x} - \big||V_0| - |V_1|\big|\vartheta/\sqrt{x}\big)_+\big]^2, \qquad x > 0.$$

Then $\rho_j(V_0, V_1) = f(\varpi_j r)$ and $\rho(V_0, V_1) = f(\varpi^* r)$, where $\varpi_j = \varpi_j(V_0, V_1; \mathcal{I})$ and $\varpi^* = \varpi^*(V_0, V_1)$, defined in (A.15) and (2.27) respectively. Since $f(x)$ is an increasing function and $|f(x) - f(y)| \leq |x - y|/4$ for all $x, y > 0$, to show (B.85), it suffices to show

$$\varpi_j \geq \varpi^* + o(1). \tag{B.86}$$

Now, we show (B.86). Introduce the quantity $\varpi = \min_{j \in (V_0 \cup V_1)} \varpi_j$. Write $B_1 = B^{\mathcal{I}^{ps}, \mathcal{I}}$, $H_1 = H^{\mathcal{I}^{ps}, \mathcal{I}^{ps}}$ and $Q_1 = B_1' H_1^{-1} B_1$. Given any $C > 0$, define $\Theta(C)$ as the collection of vectors $\xi \in \mathbb{R}^{|\mathcal{I}|}$ such that for all i, either $\xi_i^{(k)} = 0$ or $|\xi_i^{(k)}| \geq 1$, and that $\mathrm{Supp}(\xi^{(k)}) = V_k$, $\|\xi^{(k)}\|_\infty \leq C$, for $k = 0, 1$. Denote $\Theta = \Theta(\infty)$. By these notations and the definitions of ϖ_j and ϖ^*, we have

$$\varpi = \min_{(\xi^{(0)}, \xi^{(1)}):\ \xi^{(k)} \in \Theta, k=0,1; \mathrm{sgn}(\xi^{(0)}) \neq \mathrm{sgn}(\xi^{(1)})} (\xi^{(1)} - \xi^{(0)})' Q_1 (\xi^{(1)} - \xi^{(0)}),$$

$$\varpi^* = \min_{(\xi^{(0)}, \xi^{(1)}):\ \xi^{(k)} \in \Theta(a), k=0,1; \mathrm{sgn}(\xi^{(0)}) \neq \mathrm{sgn}(\xi^{(1)})} (\xi^{(1)} - \xi^{(0)})' G^{\mathcal{I}, \mathcal{I}} (\xi^{(1)} - \xi^{(0)}).$$

First, since $a > a_g^*(G)$, in the expression of ϖ^*, $\Theta(a)$ can be replaced by $\Theta(C)$ for any $C \geq a$. Second, since $\lambda_{\min}(Q_1) \geq C$, from basic properties of the quadratic programming, there exists a constant $a_0 > 0$ such that for any $(\xi_*^{(0)}, \xi_*^{(1)})$, a minimizer in the expression of ϖ, $\max\{\|\xi_*^{(0)}\|_\infty, \|\xi_*^{(1)}\|_\infty\} \leq a_0$. Therefore, in the expression of ϖ, Θ can be replaced by $\Theta(C)$ for any $C \geq a_0$. Now, let $a_1 = \max\{a_0, a\}$ and we can unify the constraints in two expressions to that $\xi^{(k)} \in \Theta(a_1)$, for $k = 0, 1$, and $\mathrm{sgn}(\xi^{(0)}) \neq \mathrm{sgn}(\xi^{(1)})$. It follows that

$$|\varpi - \varpi^*| \leq \max_{\xi \in \mathbb{R}^{|\mathcal{I}|}: \|\xi\|_\infty \leq 2a_1} \left|\xi'(G^{\mathcal{I},\mathcal{I}} - Q_1)\xi\right| \leq C\|G^{\mathcal{I},\mathcal{I}} - Q_1\|. \tag{B.87}$$

Note that Q_1 is the Fisher Information Matrix associated with model $d_1 \sim N(B_1\beta^{\mathcal{I}}, H_1)$, by Lemma 2.2 and Lemma A.3, $\|G^{\mathcal{I},\mathcal{I}} - Q_1\| = o(1)$. Plugging this into (B.87) gives $|\varpi - \varpi^*| = o(1)$. Hence, $\varpi_j \geq \varpi \geq \varpi^* + o(1)$ and (B.86) follows.

Next, consider the case $V_0 = V_1$. Pick an arbitrary minimizer in the definition of ϖ_j, denoted as $(\xi_*^{(0)}, \xi_*^{(1)})$, and define $F = \{k : \mathrm{sgn}(\xi_{*k}^{(0)}) \neq \mathrm{sgn}(\xi_{*k}^{(1)})\}$ and $N = V_0 \backslash F$. It is seen that $j \in F$. By Lemma B.3, $\rho_j^* \leq \psi(F, N)$, where $\psi(F, N)$ is defined in (2.35). Hence, it suffices to show

$$\rho_j(V_0, V_1) \geq \psi(F, N) + o(1). \tag{B.88}$$

On one hand, when $|V_0| = |V_1|$, the function f introduced above is equal to $|V_0|\vartheta + x/4$ and hence

$$\rho_j(V_0, V_1) = f(\varpi_j r) = |V_0|\vartheta + \varpi_j r/4.$$

On the other hand, using the expression of $\psi(F, N)$ in (2.35) and noting that $|F| \geq 1$,

$$\psi(F, N) \leq (|F| + |N|)\vartheta + \omega r/4 = |V_0|\vartheta + \omega r/4,$$

where $\omega = \omega(F, N)$ is defined in (2.36). Therefore, to show (B.88), it suffices to show

$$\varpi_j \geq \omega + o(1). \tag{B.89}$$

Now, we show (B.89). From the definition (2.36) and basic algebra, we can write

$$\omega = \min_{\xi \in \mathbb{R}^{|\mathcal{I}|}:\ \xi_i = 0, i \notin V_0; |\xi_i| \geq 1, i \in F} \xi' G^{\mathcal{I},\mathcal{I}} \xi.$$

Denote $\xi_* = \xi_*^{(1)} - \xi_*^{(0)}$. By our construction, $\varpi_j = \xi_*' Q_1 \xi_*$, $\xi_{*i} = 0$ for $i \notin V_0$, and $|\xi_{*i}| \geq 2$ for $i \in F$. As a result,

$$\xi_*' G^{\mathcal{I},\mathcal{I}} \xi_* \geq \omega. \tag{B.90}$$

At the same time, we have seen in the derivation of (B.87) that there exists a constant $a_0 > 0$ such that $\|\xi_*^{(0)}\|_\infty, \|\xi_*^{(1)}\|_\infty \leq a_0$ and $\|G^{\mathcal{I},\mathcal{I}} - Q_1\| = o(1)$. Therefore, $\|\xi_*\|^2 \leq 2a_0|\mathcal{I}| \leq C$ and

$$|\varpi_j - \xi_*' G^{\mathcal{I},\mathcal{I}} \xi_*| \equiv |\xi_*' Q_1 \xi_* - \xi_*' G^{\mathcal{I},\mathcal{I}} \xi_*| \leq C \|G^{\mathcal{I},\mathcal{I}} - Q_1\| = o(1). \tag{B.91}$$

Combining (B.90) and (B.91) gives (B.89). □

APPENDIX C: SUPPLEMENTARY PROOFS

In this section, we prove Lemma B.5, B.7, B.8 and B.11.

C.1. Proof of Lemma B.5. Write $\overline{\kappa}_p = \max_{\log(p) \leq j \leq p - \log(p)} G^{-1}(j,j)$ and $a_0 = \frac{1}{2\pi}\int_{-\pi}^{\pi} f^{-1}(\omega) d\omega$. The assertion of Lemma B.5 is

$$\lim_{p\to\infty} \overline{\kappa}_p = a_0.$$

To show this, denote $\underline{\kappa}_p = \min_{\log(p) \leq j \leq p - \log(p)} G^{-1}(j,j)$, and $\kappa_p = \mathrm{trace}(G^{-1})/p$. Since $\log(p) \ll p$ and all diagonals of G^{-1} are bounded from above, it follows from definitions that

$$\underline{\kappa}_p + o(1) \leq \kappa_p \leq \overline{\kappa}_p + o(1). \tag{C.1}$$

At the same time, the conditions of Lemma 2.5 ensure that $f^*(\omega)$ is continuously differentiable on $[-\pi, \pi]$. By Rambour and Seghier (2005),

$$\lim_{p\to\infty} \kappa_p = a_0.$$

Therefore, $\mathrm{liminf}_{p\to\infty} \overline{\kappa}_p \geq \lim_{p\to\infty} \kappa_p = a_0$, and all we need to show is $\mathrm{limsup}_{p\to\infty} \overline{\kappa}_p \leq a_0$.

Towards this end, write $G = G_p$ to emphasize on its dependence of p. For any positive definite $p \times p$ matrix A and a subset $V \subset \{1, \cdots, p\}$, if we let B_1 be the inverse of $A^{V,V}$ and B_2 the (V,V)-block of A^{-1}, then by elementary algebra, $B_2 - B_1$ is positive semi-definite. Now, for any (i,j) such that $\log(p) < j < p - \log(p)$ and $1 \leq i \leq \lfloor \log(p) \rfloor$, let $V = \{j - i + 1, \cdots, j - i + \lfloor \log(p) \rfloor\}$ ($\lfloor x \rfloor$ denotes the largest integer k such that $k \leq x$). Applying the above argument to the set V and matrix $A = G_p$, we have $[(G_p)^{V,V}]^{-1}(i,i) \leq G_p^{-1}(j,j)$. At the same time, the Toeplitz structure yields $(G_p)^{V,V} = G_{\lfloor \log(p) \rfloor}$. As a result, $G_{\lfloor \log(p) \rfloor}^{-1}(i,i) \leq G_p^{-1}(j,j)$. Since this holds for all i and j, we have

$$\overline{\kappa}_{\lfloor \log(p) \rfloor} \leq \underline{\kappa}_p.$$

Combining this with the first inequality of (C.1), $\overline{\kappa}_{\lfloor \log(p) \rfloor} \leq \kappa_p + o(1)$. It follows that $\mathrm{limsup}_{p\to\infty} \overline{\kappa}_p \leq \lim_{p\to\infty} \kappa_p$ and the claim follows.

We remark that additionally $\lim_{p\to\infty} \underline{\kappa}_p = a_0$, whose proof is similar so we omit. □

C.2. Proof of Lemma B.7. Fix $\log(p) \leq j \leq p - \log(p)$. Denote the collection of pairs of sets

$$\mathcal{C}_j = \big\{(F, N) : \min(F \cup N) = j,\ F \cap N = \emptyset,\ F \neq \emptyset\big\},$$

and its sub-collection

$$\mathcal{C}_j^* = \big\{(F, N) \in \mathcal{C}_j : F \in \mathcal{R}_p,\ (F \cup N) \in \mathcal{R}_p,\ |F| \leq 3 \text{ and } |N| \leq 2\big\},$$

where we recall that $\mathcal{R}_p$ is the collection of sets that are formed by consecutive nodes. The claim now reduces to

$$\min_{(F,N)\in\mathcal{C}_j^*} \psi(F, N) = \min_{(F,N)\in\mathcal{C}_j} \psi(F, N) + o(1).$$

Noting that $\mathcal{C}_j^* \subset \mathcal{C}_j$, it suffices to show for any $(F, N) \in \mathcal{C}_j$, there exists (F', N') such that

$$\psi(F', N') \leq \psi(F, N) + o(1) \quad \text{and} \quad (F', N') \in \mathcal{C}_j^*. \tag{C.2}$$

To show (C.2), we introduce the notation $(F', N') \preceq (F, N)$ to indicate

$$\psi(F', N') \leq \psi(F, N), \quad |F'| \leq |F|, \quad \text{and} \quad |N'| \leq |N|.$$

Using these notations, we claim:

(a) For any $(F, N) \in \mathcal{C}_j$, there exists $(F', N') \in \mathcal{C}_j$ such that $\psi(F', N') \leq \psi(F, N) + o(1)$ and $|F'| \leq 3$.
(b) For any $(F, N) \in \mathcal{C}_j$, there exists $(F', N') \in \mathcal{C}_j$ such that $(F', N') \preceq (F, N)$ and $(F' \cup N') \in \mathcal{R}_p$.
(c) For any $(F, N) \in \mathcal{C}_j$ satisfying $(F \cup N) \in \mathcal{R}_p$, there exists $(F', N') \in \mathcal{C}_j$ such that $(F', N') \preceq (F, N)$, $(F' \cup N') \in \mathcal{R}_p$ and $F' \in \mathcal{R}_p$.
(d) For any $(F, N) \in \mathcal{C}_j$ satisfying $(F \cup N) \in \mathcal{R}_p$ and $F \in \mathcal{R}_p$, there exists $(F', N') \in \mathcal{C}_j$ such that $(F', N') \preceq (F, N)$, $(F' \cup N') \in \mathcal{R}_p$, $F' \in \mathcal{R}_p$ and $|N'| \leq 2$.

Now, for any $(F, N) \in \mathcal{C}_j$, we construct (F', N') as follows: First, by (a), there exists (F_1, N_1) such that $\psi(F_1, N_1) \leq \psi(F, N) + o(1)$, and $|F_1| \leq 3$. Second, by (b) and (c), there exists (F_2, N_2) such that $(F_2, N_2) \preceq (F_1, N_1)$, $F_2 \in \mathcal{R}_p$ and $(F_2 \cup N_2) \in \mathcal{R}_p$. Finally, by (d), there exists (F_3, N_3) such that $(F_3, N_3) \preceq (F_2, N_2)$, $(F_3 \cup N_3) \in \mathcal{R}_p$, $F_3 \in \mathcal{R}_p$ and $|N_3| \leq 2$. Let $(F', N') = (F_3, N_3)$.

By the construction, $(F' \cup N') \in \mathcal{R}_p$, $F' \in \mathcal{R}_p$ and

$$\psi(F', N') = \psi(F_3, N_3) \leq \psi(F_2, N_2) \leq \psi(F_1, N_1) \leq \psi(F, N) + o(1).$$

Moreover, $|F'| = |F_3| \leq |F_2| \leq |F_1| \leq 3$, and $|N'| = |N_3| \leq 2$. So (F', N') satisfies (C.2).

All remains is to verify the claims (a)-(d). We need the following results, which follow from basic algebra and we omit the proof: First, recall the definition of $\omega(F, N)$ in (2.36). For any fixed (F, N), let $\mathcal{I} = F \cup N$ and $R = (G^{\mathcal{I},\mathcal{I}})^{-1}$. Then

$$\omega(F, N) = \min_{\xi \in \mathbb{R}^{|F|}: |\xi_i| \geq 1} \xi'(R^{F,F})^{-1}\xi. \tag{C.3}$$

Second, when $(F \cup N) \in \mathcal{R}_p$,

$$R = \frac{1}{j}\eta\eta' + \Sigma_*^{(k)}, \qquad k = |F \cup N|, \tag{C.4}$$

where $\eta = (1, 0, \cdots, 0)'$ and $\Sigma_*^{(k)}$ is as in (B.44).

Now, we show (a). The case $|F| \leq 3$ is trivial, so without loss of generality we assume $|F| \geq 4$. Take

$$F' = \{j + 1, j + 2\}, \qquad N' = \{j\}.$$

We check that (F', N') satisfies the requirement in (a). It is obvious that $(F', N') \in \mathcal{C}_j$ and $|F'| \leq 3$. We only need to check $\psi(F', N') \leq \psi(F, N) + o(1)$. On one hand, direct calculations yield $\omega(F', N') = (j + 1)/(j + 2) = 1 + o(1)$, and

$$\psi(F', N') \leq 2\vartheta + r/4 + o(1).$$

On the other hand, by (C.3), $\omega(F, N) \geq |F| \cdot [\lambda_{\max}(R)]^{-1} \geq |F| \cdot \lambda_{\min}(G)$. Noting that $G^{-1} = H$, we have $\|G^{-1}\| \leq \|H\|_\infty \leq 4$. So $\lambda_{\min}(G) \geq 1/4$. Therefore, $\omega(F, N) \geq 1$. It follows that

$$\psi(F, N) \geq |F|\vartheta/2 + \omega(F, N)r/4 \geq 2\vartheta + r/4.$$

Combining the two parts, we have $\psi(F', N') \leq \psi(F, N) + o(1)$.

Next, we verify (b). We construct (F', N') by constructing a sequence of $(F^{(t)}, N^{(t)})$ recursively: Initially, set $F^{(1)} = F$ and $N^{(1)} = N$. On round t, write $F^{(t)} \cup N^{(t)} = \{j_1, \cdots, j_k\}$, where the nodes are arranged in the acceding order and $k = |F^{(t)} \cup N^{(t)}|$. Let l_0 be the largest index such that $j_l = j_1 + l - 1$ for all $l \leq l_0$. If $l_0 = k$, then the process terminates. Otherwise, let $L = j_{l_0+1} - j_1 - l_0$ and update

$$F^{(t+1)} = \big\{j_l - L \cdot 1\{l > l_0\} : j_l \in F^{(t)}\big\}, \quad N^{(t+1)} = \big\{j_l - L \cdot 1\{l > l_0\} : j_l \in N^{(t)}\big\}.$$

By the construction, it is not hard to see that l_0 strictly increases as t increases, and k remains unchanged. So the process terminates in finite rounds. Let T be the number of rounds when the process terminates, we construct (F', N') by

$$F' = F^{(T)}, \qquad N' = N^{(T)}.$$

Now, we justify that (F', N') satisfies the requirement in (b). First, it is seen that $\min(F^{(t)} \cup N^{(t)}) = j$ on every round t. So $\min(F \cup N) = j$ and $(F, N) \in \mathcal{C}_j$. Second, on round T, $l_0 = k$, which implies $(F' \cup N') \in \mathcal{R}_p$. Third, $|F^{(t)}|$ and $|N^{(t)}|$ keep unchanged as t increases, so $|F'| = |F|$ and $|N'| = |N|$. Finally, it remains to check $\psi(F', N') \leq \psi(F, N)$. It suffices to show

$$\psi(F^{(t+1)}, N^{(t+1)}) \leq \psi(F^{(t)}, N^{(t)}), \quad \text{for } t = 1, \cdots, T-1. \tag{C.5}$$

Let $\mathcal{I} = F^{(t)} \cup N^{(t)}$ and $\mathcal{I}_1 = F^{(t+1)} \cup N^{(t+1)}$. We observe that $G^{\mathcal{I}_1,\mathcal{I}_1} = G^{\mathcal{I},\mathcal{I}} - L\eta\eta'$, where $\eta = (0'_{l_0}, 1'_{k-l_0})'$. So $G^{\mathcal{I},\mathcal{I}} - G^{\mathcal{I}_1,\mathcal{I}_1}$ is positive semi-definite. It follows form (C.3) that $\omega(F^{(t+1)}, N^{(t+1)}) \leq \omega(F^{(t)}, N^{(t)})$, and hence (C.5) holds by recalling that $|F^{(t+1)}| = |F^{(t)}|$ and $|N^{(t+1)}| = |N^{(t)}|$.

Third, we prove (c). By assumptions, $(F \cup N) \in \mathcal{R}_p$, so that we can write $F \cup N = \{j, j+1, \cdots, j+k\}$, where $k + 1 = |F \cup N|$. The case $F \in \mathcal{R}_p$ is trivial. In the case $F \notin \mathcal{R}_p$, we construct (F', N') as follows: Let i_0 be the smallest index such that $i_0 \notin F$ and both $F_1 = F \cap \{i : i < i_0\}$ and $F_2 = F \backslash F_1$ are not empty. We note that such i_0 exists because $F \notin \mathcal{R}_p$. Let

$$F' = F_1 = \{i \in F : i < i_0\}, \qquad N' = \{i \in N : i \leq i_0\}.$$

To check that (F', N') satisfies the requirement in (c), first note that $\min(F' \cup N') = j$ and hence $(F', N') \in \mathcal{C}_j$. Second, it is easy to see that $|F'| \leq |F|$ and $|N'| \leq |N|$. Third, from the definition of i_0, $F' \in \mathcal{R}_p$. Additionally, since $i_0 \in N$, we have $F' \cup N' = \{j, j+1, \cdots, i_0\} \in \mathcal{R}_p$. Last, we check $\psi(F', N') \leq \psi(F, N)$: Since $|F'| \leq |F|$ and $|N'| \leq |N|$, it suffices to show

$$\omega(F', N') \leq \omega(F, N). \tag{C.6}$$

Write $\mathcal{I} = F \cup N$ and denote $R = (G^{\mathcal{I},\mathcal{I}})^{-1}$. From (C.4), R is tri-diagonal. So $R^{F,F}$ is block-diagonal in the partition $F = F_1 \cup F_2$. Using (C.3), it is easy to see

$$\omega(F_1, \mathcal{I} \backslash F_1) \leq \omega(F, \mathcal{I} \backslash F) \equiv \omega(F, N).$$

At the same time, notice that both $\mathcal{I}$ and $\mathcal{I}' = F' \cup N'$ have the form $\{j, j+1, \cdots, m\}$ with $m \geq \max(F_1) + 1$. Applying (C.3) and (C.4), by direct calculations,

$$\omega(F_1, \mathcal{I} \backslash F_1) = \omega(F_1, \mathcal{I}' \backslash F_1) \equiv \omega(F', N').$$

Combining the two parts gives (C.6).

Finally, we justify (d). By assumptions, $(F \cup N) \in \mathcal{R}_p$ and $F \in \mathcal{R}_p$, so that we write $F \cup N = \{j, j+1, \cdots, k\}$, and $F = \{j_0, j_0+1, \cdots, k_0\}$, where $j_0 \geq j$ and $k_0 \leq k$. The case $|N| \leq 2$ is trivial. In the case $|N| > 2$, let $m_0 = |F|$ and we construct (F', N') as follows:

$$\begin{array}{ll} F' = F, \quad N' = \{k_0+1\}, & \text{when } j_0 = j; \\ F' = \{j+1, j+2, \cdots, j+m_0\}, \quad N' = \{j, \ j+m_0+1\}, & \text{when } j_0 > j, k_0 < k; \\ F' = \{j+1, j+2, \cdots, j+m_0\}, \quad N' = \{j\}, & \text{when } j_0 > j, k_0 = k. \end{array}$$

Now, we show that (F', N') satisfies the requirement in (d). First, by the construction, $(F', N') \in \mathcal{C}_j$, $(F' \cup N') \in \mathcal{R}_p$ and $F' \in \mathcal{R}_p$. Second, $|F'| = |F|$, $|N'| \leq 2 < |N|$. Third, we check $\psi(F', N') \leq \psi(F, N)$. Applying (C.3) and (C.4), direct calculations yield $\omega(F', N') = \omega(F, N)$. This, together with $|F'| \leq |F|$ and $|N'| \leq |N|$, proves $\psi(F', N') \leq \psi(F, N)$. □

C.3. Proof of Lemma B.8. Recalling the definition of $\mathcal{C}_j^*$ in the proof of Lemma B.7, the claim reduces to

$$\min_{(F,N)\in\mathcal{C}_j^*} \psi(F,N) = \min_{(F,N)\in\mathcal{C}_1^*} \psi^{(\infty)}(F,N) + o(1), \qquad \log(p) \leq j \leq p - \log(p).$$

We argue that on both sides, the minimum is not attained on (F, N) such that $|N| = 0$ and $|F| = 1$. In this case, on the left hand side, $F = \{j\}$ and $N = \emptyset$. By direct calculations, $\omega(F, N) = j \geq \log(p)$, and hence $\psi(F, N)$ can not be the minimum. Similarly, on the right hand side, $\omega^{(\infty)}(F, N) = \infty$ by definition, and the same conclusion follows. Therefore, the claim is equivalent to

$$\min_{(F,N)\in\mathcal{C}_j^*:|F|+|N|>1} \psi(F,N) = \min_{(F,N)\in\mathcal{C}_1^*:|F|+|N|>1} \psi^{(\infty)}(F,N) + o(1).$$

Fix $\log(p) \leq j \leq p - \log(p)$. Define a one-to-one mapping from $\mathcal{C}_j^*$ to $\mathcal{C}_1^*$, where given any $(F, N) \in \mathcal{C}_j^*$, it is mapped to (F_1, N_1) such that

$$F_1 = \{i - j + 1 : i \in F\}, \qquad N_1 = \{i - j + 1 : i \in N\}.$$

To show the claim, it suffices to show when $|F| + |N| > 1$,

$$\psi(F, N) = \psi^{(\infty)}(F_1, N_1) + o(1).$$

Since $|F_1| = |F|$ and $|N_1| = |N|$, it is sufficient to show

$$\omega(F, N) = \omega^{(\infty)}(F_1, N_1) + o(1). \tag{C.7}$$

Now, we show (C.7). Consider the case $N \neq \emptyset$ first. Suppose $|\mathcal{I}| = k$ and write $\mathcal{I} = F \cup N = \{j, \cdots, j+k-1\}$, where $1 < k \leq 5$. Let $R = (G^{\mathcal{I},\mathcal{I}})^{-1}$ and $R_* = (\Sigma_*^{(k)})^{F_1,F_1}$, where $\Sigma_*^{(k)}$ is defined in (B.44). We note that when $N \neq \emptyset$, R_* is invertible. Using (C.3) and the definition of $\omega^{(\infty)}$,

$$|\omega(F,N) - \omega^{(\infty)}(F_1,N_1)| \leq \max_{\xi \in \mathbb{R}^k : |\xi_i| \leq 2a} \left|\xi'[(R^{F,F})^{-1} - R_*^{-1}]\xi\right|. \tag{C.8}$$

Since $\mathcal{I} \in \mathcal{R}_p$, we apply (C.4) and obtain

$$R^{F,F} = \frac{1}{j}(\eta^{F_1})(\eta^{F_1})' + (\Sigma_*^{(k)})^{F_1,F_1},$$

where $\eta = (1, 0, \cdots, 0)' \in \mathbb{R}^k$. By matrix inverse formula,

$$\xi'[(R^{F,F})^{-1} - R_*^{-1}]\xi = -[j + (\eta^{F_1})' R_*^{-1} \eta^{F_1}]^{-1} (\xi' R_*^{-1} \eta^{F_1})^2. \tag{C.9}$$

Combining (C.8) and (C.9),

$$|\omega(F,N) - \omega^{(\infty)}(F_1,N_1)| \leq j^{-1} \max_{\xi \in \mathbb{R}^k : |\xi_i| \leq 2a} |\xi' R_*^{-1} \eta^{F_1}|^2 \leq j^{-1} \cdot C \|R_*^{-1}\|^2.$$

Since $N_1 \neq \emptyset$ and k is finite, $\lambda_{\min}(R_*) \geq C > 0$ and hence $\|R_*^{-1}\| \leq C$. Noting that $j \geq \log(p)$, (C.7) follows directly.

Next, consider the case $N = \emptyset$. Suppose $|F| = k$ and write $F = \{j, \cdots, j + k - 1\}$, where $1 < k \leq 3$. We observe that $G^{F,F} = j11' + \Omega_*^{(k)}$, where $\Omega_*^{(k)}$ is defined in (B.45). By definition

$$\omega(F,N) = \min_{\xi \in \mathbb{R}^k : |\xi_i| \geq 1} \xi' G^{F,F} \xi = \min_{\xi \in \mathbb{R}^k : |\xi_i| \geq 1} \left[j(1'\xi)^2 + \xi' \Omega_*^{(k)} \xi \right].$$

On one hand, if we let ξ^* be one minimizer in the definition of $\omega^{(\infty)}(F_1, N_1)$, then $1'\xi^* = 0$. As a result,

$$\omega(F,N) \leq j(1'\xi^*)^2 + (\xi^*)' \Omega_*^{(k)} \xi^* = (\xi^*)' \Omega_*^{(k)} \xi^* \equiv \omega^{(\infty)}(F_1, N_1). \tag{C.10}$$

On the other hand, we can show

$$\omega(F,N) \geq \omega^{(\infty)}(F_1,N_1) - 1/(j+1). \tag{C.11}$$

Combing (C.10) and (C.11), and noting that $j \geq \log(p)$, we obtain (C.7).

It remains to show (C.11). When $k = 2$, by direct calculations, $\omega(F,N) = \omega^{(\infty)}(F,N) = 1$. When $k > 2$, write $\xi = (\xi_1, \xi_2, \tilde{\xi}')'$ for any $\xi \in \mathbb{R}^k$, and introduce the function $g(x) = \sum_{i=1}^{k-2}(x_i + x_{i+1} + \cdots + x_{k-2})^2$, for $x \in \mathbb{R}^{k-2}$. We observe that

$$\xi'\Omega_*^{(k)}\xi = (1'\xi - \xi_1)^2 + g(\tilde{\xi}). \tag{C.12}$$

Let $g_{\min} = \min_{x\in\mathbb{R}^{k-2}:|x_i|\geq 1} g(x)$. We claim that there exists $q \in \mathbb{R}^k$ such that

$$1'q = 0, \qquad q'\Omega_*^{(k)}q = 1 + g_{\min}, \qquad \text{and} \qquad |q_i| \geq 1, \quad \text{for } 1 \leq i \leq k.$$

To see this, note that under the constraints $|x_i| \geq 1$, $g(x)$ is obviously minimized at $x^* = (\cdots, -1, 1, -1, 1)$. Observing that $1'(x^*)$ is either 0 or 1, we let $q = (1, -1, (x^*)')'$ when $1'(x^*) = 0$, and let $q = (1, -2, (x^*)')'$ when $1'(x^*) = 1$. Using (C.12), it is easy to check that q satisfies the above requirements. It follows that

$$\omega^{(\infty)}(F_1, N_1) = \min_{\xi\in\mathbb{R}^k:|\xi_i|\geq 1, 1'\xi=0} \xi'\Omega_*^{(k)}\xi \leq q'\Omega_*^{(k)}q = 1 + g_{\min}. \tag{C.13}$$

At the same time, since $G^{F,F} = j11' + \Omega_*^{(k)}$, we can write from (C.12) that

$$\xi'G^{F,F}\xi = j(1'\xi)^2 + (1'\xi - \xi_1)^2 + g(\tilde{\xi}). \tag{C.14}$$

Note that $\min_y\{jy^2 + (y-c)^2\} = c^2\ j/(j+1)$, for any $c \in \mathbb{R}$. So

$$j(1'\xi)^2 + (1'\xi - \xi_1)^2 \geq |\xi_1|^2 j/(j+1).$$

Plugging this into (C.14), we find that

$$\omega(F, N) = \min_{\xi\in\mathbb{R}^k:|\xi_i|\geq 1} \xi'G^{F,F}\xi \geq j/(j+1) + g_{\min}. \tag{C.15}$$

Combining (C.13) and (C.15) gives (C.11). □

C.4. Proof of Lemma B.11. To show the claim, we first introduce a key lemma: Fix a linear filter $D_{h,\eta}$, for any dimension $k > h$, let $\tilde{D}^{(k)}$ be the $(k-h) \times k$ matrix, where for each $1 \leq i \leq k-h$, $\tilde{D}^{(k)}(i,i) = 1$, $\tilde{D}^{(k)}(i,i+1) = \eta_1$, $\cdots$, $\tilde{D}^{(k)}(i,i+h) = \eta_h$, and $\tilde{D}^{(k)}(i,j) = 0$ for other j. Define the null space of $D_{h,\eta}$ in dimension k, $Null_k(\eta)$, as the collection of all vectors $\xi \in \mathbb{R}^k$ that satisfies $\tilde{D}^{(k)}\xi = 0$. The following lemma is proved below.

Lemma C.1. *For a given η, if RCA holds, then for sufficiently large n and any $k \geq n$, there exists an orthonormal basis of $Null_k(\eta)$, denoted as $\xi^{(1)}, \cdots, \xi^{(h)}$, such that*

$$\max_{1\leq i\leq k-n, 1\leq j\leq h} |\xi_i^{(j)}|^2 \leq C_\eta n^{-1},$$

where $C_\eta > 0$ is a constant that only depends on η.

Second, we state some observations. Fix $\mathcal{I} \trianglelefteq \mathcal{G}^+$. Partition $\mathcal{I}^{pe}$ uniquely as $\mathcal{I}^{pe} = \cup_{t=1}^T V_t$, so that $V_t = \{i_t, i_t+1, \cdots, j_t-1, j_t\}$ is formed by consecutive nodes and $j_t < i_{t+1}$ for all t. Denote $M = |\mathcal{J}^{pe}| - |\mathcal{I}^{pe}|$. It is easy to see that $T \leq M$ and $M \leq h|\mathcal{I}| \leq l_0 h$, so both M and T are finite. Let $\tilde{V}_t = \{1 \leq j \leq p : D(i,j) \neq 0 \text{ for some } i \in V_t\}$ and define $Null(V_t, \tilde{V}_t)$ in the same way as $Null(\mathcal{I}^{pe}, \mathcal{J}^{pe})$. Recall that $\mathcal{F}$ is the mapping from nodes in $\mathcal{J}^{pe}$ to their orders in $\mathcal{J}^{pe}$. Similarly, define the mapping $\mathcal{F}_t$ from $\tilde{V}_t$ to $\{1, \cdots, |\tilde{V}_t|\}$ that maps each $j \in \tilde{V}_t$ to its order in $\tilde{V}_t$. Denote $\mathcal{I}_t = \mathcal{F}_t(\mathcal{I} \cap V_t)$. We observe that:

- (O_1) $\tilde{V}_t \cap \tilde{V}_{t'} \neq \emptyset$ only when $|t - t'| \leq 1$; and $|\tilde{V}_t \cap \tilde{V}_{t+1}| \leq h - 1$, for all t.
- (O_2) $Null(V_t, \tilde{V}_t) = Null_{|\tilde{V}_t|}(\eta)$ for all t, where $Null_k(\eta)$ is as in Lemma C.1.
- (O_3) $\mathcal{J}^{pe} = \cup_{t=1}^T \tilde{V}_t$; and $|\tilde{V}_t| \geq |V_t| \geq 2\ell^{pe} + 1$, for all t.
- (O_4) Any node $i \in \mathcal{I}_t$ satisfies that $1 \leq i < |\tilde{V}_t| - \ell^{pe}$, for all t.
- (O_5) For any $\xi \in \mathbb{R}^{|\mathcal{J}^{pe}|}$, $\xi \in Null(\mathcal{I}^{pe}, \mathcal{J}^{pe})$ if and only if $\xi^{\mathcal{F}(\tilde{V}_t)} \in Null(V_t, \tilde{V}_t)$ for all t, where $\xi^{\mathcal{F}(\tilde{V}_t)}$ is the subvector of ξ formed by elements in $\mathcal{F}(\tilde{V}_t)$.

Due to (O_2) and Lemma C.1, for each t, there exists an orthonormal basis $\xi^{(t,1)}, \cdots, \xi^{(t,h)}$ for $Null(V_t, \tilde{V}_t)$ such that

$$\max_{1 \leq i \leq |\tilde{V}_t| - n, 1 \leq j \leq h} |\xi_i^{(t,j)}|^2 \leq C_\eta n^{-1}, \quad \text{for any } 1 \leq n < |\tilde{V}_t|. \tag{C.16}$$

Let U_t be the matrix formed by the last h rows of $[\xi^{(t,1)}, \cdots, \xi^{(t,h)}]$. From the explicit form of the basis in the proof of Lemma C.1, we further observe:

- (O_6) $c' \leq \lambda_{\min}(U_t U_t') \leq \lambda_{\max}(U_t U_t') \leq 1 - c$, where $0 < c, c' < 1$ and $c + c' < 1$.
- (O_7) For each $1 \leq h_0 \leq h$, the submatrix of U_t formed by its last h_0 rows has a rank h_0.

Now, we show the claim by constructing a matrix W, whose columns form an orthonormal baisis for $Null(\mathcal{I}^{pe}, \mathcal{J}^{pe})$, and it satisfies

$$\max_{1 \leq i \leq |\mathcal{J}^{pe}| - n, 1 \leq j \leq M} |W(i,j)|^2 \leq C n^{-1}, \quad \text{for any } 1 \leq n < |\mathcal{J}^{pe}|. \tag{C.17}$$

In fact, once such W is constructed, any U whose columns form an orthonromal baisis for $Null(\mathcal{I}^{pe}, \mathcal{J}^{pe})$ can be written as

$$U = WR,$$

where R has the dimension $M \times M$ and $R'R$ is an identity matrix. By basic algebra, for any $m \times n$ matrix A and $n \times p$ matrix B, $\max_{1 \leq j \leq p} |(AB)(i,j)|^2 \leq$

$n\|B'B\| \cdot \max_{1\leq k\leq n} |A(i,k)|^2$ for each $1 \leq i \leq m$. Applying this to W and R, and noting that $\|R'R\| = 1$ and that M is finite, we obtain

$$\max_{i\in\mathcal{F}(\mathcal{I}),1\leq j\leq M} |U(i,j)|^2 \leq C \max_{i\in\mathcal{F}(\mathcal{I}),1\leq j\leq M} |W(i,j)|^2. \tag{C.18}$$

At the same time, for any $i \in \mathcal{I}$, there exists a unique t such that $i \in \mathcal{I} \cap V_t$. In addition, from (O_4), $\mathcal{F}_t(i) < |\tilde{V}_t| - \ell^{pe}$. By the construction, this implies $\mathcal{F}(i) < |\mathcal{J}^{pe}| - \ell^{pe}$. Combining this to (C.17), we find that

$$\max_{i\in\mathcal{F}(\mathcal{I}),1\leq j\leq M} |W(i,j)|^2 \leq C(\ell^{pe})^{-1}. \tag{C.19}$$

The claim then follows from (C.18) and (C.19).

To construct W, the key is to recursively construct matrices $W_T, W_{T-1}, \cdots, W_1$. Denote $m_t = h - |\tilde{V}_t \cap \tilde{V}_{t+1}|$, with $m_T = h$ by convention; $M_t = \sum_{s=t}^{T} m_s$ and $L_t = |\cup_{s=t}^{T} \tilde{V}_s|$; in particular, $M_1 = |\mathcal{J}^{pe}| - |\mathcal{I}^{pe}| = M$ and $L_1 = |\mathcal{J}^{pe}|$. Initially, construct the $L_T \times M_T$ matrix

$$W_T = \left[\xi^{(T,1)}, \cdots, \xi^{(T,h)}\right],$$

where $\{\xi^{(T,j)} : 1 \leq j \leq h\}$ is the orthonormal basis in (C.16). Given W_{t+1}, construct the $L_t \times M_t$ matrix W_t as follows: Denote $\widetilde{W}_{t+1}$ the submatrix of W_{t+1} formed by its first $|\tilde{V}_t \cap \tilde{V}_{t+1}|$ $(= h - m_t)$ rows and write

$$\left[\xi^{(t,1)}, \cdots, \xi^{(t,h)}\right] = \begin{bmatrix} A_t \\ B_t \end{bmatrix},$$

where A_t has $(|\tilde{V}_t| - h - m_t)$ rows and B_t has $(h - m_t)$ rows. From (O_7), the rank of B_t is $(h - m_t)$. Hence, there exists an $h \times m_t$ matrix Q_t, such that $Q_t'Q_t$ is an identity matrix and $B_tQ_t = 0$. Now, construct

$$W_t = \begin{bmatrix} A_tB_t'(B_tB_t')^{-1}\widetilde{W}_{t+1} & A_tQ_t \\ W_{t+1} & 0 \end{bmatrix}. \tag{C.20}$$

Continue this process until we obtain W_1 and let

$$W = W_1(W_1'W_1)^{-1/2}.$$

Below, we check that W satisfies the requirement. First, we show that the columns of W form an orthonormal basis of $Null(\mathcal{I}^{pe}, \mathcal{J}^{pe})$. Since W has $M = |\mathcal{J}^{pe}| - |\mathcal{I}^{pe}|$ columns and its columns are orthonormal, it suffices to show that all its columns belong to $Null(\mathcal{I}^{pe}, \mathcal{J}^{pe})$. By (O_5), we only need to show that for each $1 \leq t \leq T$, in the submatix of W formed by restricting rows into $\mathcal{F}(\tilde{V}_t)$, all its columns belong to $Null(V_t, \tilde{V}_t)$. By the construction, only the first M_t columns of this submatrix are non-zero and they are equal to

$$\begin{bmatrix} A_t B_t'(B_t B_t')^{-1}\widetilde{W}_{t+1} & A_t Q_t \\ \widetilde{W}_{t+1} & 0 \end{bmatrix} = \begin{bmatrix} A_t \\ B_t \end{bmatrix} \left[B_t'(B_t B_t')^{-1}\widetilde{W}_t, \ Q_t \right],$$

where in the equality we have used the facts that $\widetilde{W}_t = B_t B_t'(B_t B_t')^{-1}\widetilde{W}_t$ and $B_t Q_t = 0$. Combining this to the definition of A_t and B_t, we find that each column of the above matrix is a linear combination of $\{\xi^{(t,1)}, \cdots, \xi^{(t,h)}\}$ and hence belongs to $Null(V_t, \tilde{V}_t)$.

Second, we show that W satisfies (C.17). It suffices to show, for $t = T, \cdots, 1$,

(a) $\max_{1\leq i\leq L_t-n, 1\leq j\leq M_t} |W_t(i,j)|^2 \leq Cn^{-1}$, for any $1 \leq n < L_t$.
(b) $\lambda_{\min}(W_t' W_t) \geq C > 0$.

In fact, once (a) and (b) are proved, by taking $t = 1$ and noticing that $L_1 = |\mathcal{J}^{pe}|$, we have $\max_{1\leq i\leq |\mathcal{J}^{pe}|-n, 1\leq j\leq M} |W_1(i,j)|^2 \leq Cn^{-1}$, for $1 \leq n < |\mathcal{J}^{pe}|$; and $\|(W_1'W_1)^{-1}\| = [\lambda_{\min}(W_1'W_1)]^{-1} \leq C$. Hence, by similar arguments in (C.18), for each $1 \leq i \leq |\mathcal{J}^{pe}| - n$, $\max_{1\leq j\leq M} |W(i,j)|^2 \leq M\|(W_1'W_1)^{-1}\| \cdot \max_{1\leq j\leq M} |W_1(i,j)|^2 \leq Cn^{-1}$. This gives (C.17).

It remains to show (a) and (b). Note that for W_T, by the construction and (C.16), (a) and (b) hold trivially. We aim to show that if (a) and (b) hold for W_{t+1}, then they also hold for W_t. For preparation, we argue that

$$\|A_t B_t'(B_t B_t')^{-1}\widetilde{W}_{t+1}\|^2 \leq C(\ell^{pe})^{-1} = o(1). \tag{C.21}$$

To see this, note that $L_{t+1} \geq 2\ell^{pe} + 1$ from (O_3); in particular, $h - m_t \ll L_{t+1} - \ell^{pe}$. Hence, if (a) holds for W_{t+1}, $\max_{1\leq i\leq h-m_t, 1\leq j\leq M_{t+1}} |W_{t+1}(i,j)|^2 \leq C(\ell^{pe})^{-1}$, i.e., $|\widetilde{W}_{t+1}(i,j)| \leq C(\ell^{pe})^{-1}$, for any (i,j). Since $\widetilde{W}_{t+1}$ has a finite dimension, this yields $\|\widetilde{W}_{t+1}\|^2 \leq C(\ell^{pe})^{-1}$. Furthermore, from (O_6) and that $B_t B_t'$ is a submatrix of $U_t U_t'$, $\lambda_{\min}(B_t B_t') \geq c' > 0$. So $\|(B_t B_t')^{-1}\| \leq C$. In addition, $\|A_t\|, \|B_t\| \leq 1$. Combining the above gives (C.21).

Consider (a) first. By (C.20), (C.21) and the assumption on W_{t+1}, it suffices to show

$$\max_{1\leq i\leq |\tilde{V}_t|-n, 1\leq j\leq m_t} |A_t Q_t(i,j)|^2 \leq Cn^{-1}, \quad \text{for any } 1 \leq n < |\tilde{V}_t|. \tag{C.22}$$

By similar arguments in (C.18) and the fact that $\|Q_t'Q_t\| = 1$, the left hand side is bounded by $C\max_{1\leq i\leq |\tilde{V}_t|-n, 1\leq j\leq m_t} |A_t(i,j)|^2$. Therefore, (C.22) follows from (C.16) and the definition of A_t.

Next, consider (b). Using (C.20) and (C.21), we can write

$$W_t'W_t = \begin{bmatrix} W_{t+1}'W_{t+1} + \Delta_1 & \Delta_2 \\ \Delta_2' & Q_t'A_t'A_tQ_t \end{bmatrix},$$

where $\|\Delta_1\| = o(1)$ and $\|\Delta_2\| = o(1)$. So it suffices to show $\lambda_{\min}(W_{t_1}'W_{t+1}) \geq C$ and $\lambda_{\min}(Q_t'A_t'A_tQ_t) \geq C$. The former follows from the assumption on W_{t+1}. To show the latter, note that $Q_t'Q_t$ is an identity matrix, and so $\lambda_{\min}(Q_t'A_t'A_tQ_t) \geq \lambda_{\min}(A_t'A_t)$. Also, since $A_t'A_t + B_t'B_t$ is an identity matrix, $\lambda_{\min}(A_t'A_t) = 1 - \lambda_{\max}(B_t'B_t)$. Additionally, $\lambda_{\max}(B_t'B_t) = \lambda_{\max}(B_tB_t')$, where B_tB_t' is a submatrix of U_tU_t', and by (O_6), $\lambda_{\max}(U_tU_t') \leq 1 - c$. Combining the above yields $\lambda_{\min}(Q_t'A_t'A_tQ_t) \geq c > 0$. This proves (b). ☐

C.4.1. *Proof of Lemma C.1.* For each $k \geq h$, we construct a $k \times h$ matrix U whose columns form an orthonormal basis of $Null_k(\eta)$ as follows: Recall the characteristic polynomial $\varphi_\eta(z) = 1 + \eta_1 z + \cdots + \eta_h z^h$. Let $z_1, \cdots, z_m$ be m different roots of $\varphi_\eta(z)$, each replicating $h_1, \cdots, h_m$ times respectively $(h_1 + \cdots + h_m = h)$. For $1 \leq j \leq m$ and $1 \leq s \leq h_j$, when z_i is a real root, let

$$\mu^{(j,s)} = \left(k^{s-1}\frac{1}{z_j^{k-1}},\ \cdots,\ 3^{s-1}\frac{1}{z_j^2},\ 2^{s-1}\frac{1}{z_j},\ 1 \right)';$$

and when $z_{j\pm} = |z_j|e^{\pm\sqrt{-1}\theta_j}$, $\theta_j \in (0, \pi/2]$, are a pair of conjugate roots, let

$$\mu^{(j+,s)} = \left(k^{s-1}\frac{\cos(k-1)\theta_j}{|z_j|^{k-1}},\ \cdots,\ 3^{s-1}\frac{\cos 2\theta_j}{|z_j|^2},\ 2^{s-1}\frac{\cos\theta_j}{|z_j|},\ 1 \right)',$$

$$\mu^{(j-,s)} = \left(k^{s-1}\frac{\sin(k-1)\theta_j}{|z_j|^{k-1}},\ \cdots,\ 3^{s-1}\frac{\sin 2\theta_j}{|z_j|^2},\ 2^{s-1}\frac{\sin\theta_j}{|z_j|},\ 1 \right)'.$$

It is seen that $\{\mu^{(j,s)}, 1 \leq j \leq m, 1 \leq s \leq h_j\}$ are h vectors in $\mathbb{R}^k$. Let $\xi^{(j,s)} = \mu^{(j,s)}/\|\mu^{(j,s)}\|$ for each (j,s), and construct the $k \times h$ matrix

$$R = \left[\xi^{(1,1)}, \cdots, \xi^{(1,h_1)}, \cdots, \xi^{(m,1)}, \cdots, \xi^{(m,h_m)}\right].$$

Define

$$U = R(R'R)^{-1/2}.$$

Now, we show that the vectors $\{\mu^{(j,s)}, 1 \leq j \leq m, 1 \leq s \leq h_j\}$ are linearly independent and span $Null_k(\eta)$. Therefore, U is well defined and its columns form an orthonormal basis of $Null_k(\eta)$. To see this, note that for any vector $\eta \in \mathbb{R}^k$, if we write $\eta_1 = f(k)$, $\cdots$, $\eta_k = f(1)$, then $\xi \in Null_k(\eta)$ if and only if $f(i)$'s satisfy the difference equation:

$$f(i) + \eta_1 f(i-1) + \cdots + \eta_h f(i-h) = 0, \quad h+1 \leq i \leq k. \tag{C.23}$$

It is well-known in theories of difference equations that (C.23) has h independent base solutions:

$$f_{j,s}(i) = i^{s-1} z_j^{-i}, \qquad 1 \le j \le m, \quad 1 \le s \le h_j.$$

By the construction, when z_j is a real root, $\mu^{(j,s)} = (f_{j,s}(k), \cdots, f_{j,s}(1))'$; and when $z_{j\pm}$ are a pair of conjugate roots, $\mu^{(j+,s)}$ and $\mu^{(j-,s)}$ are the real and imaginary parts of the vector $(f_{j,s}(k), \cdots, f_{j,s}(1))'$. So the vectors $\{\mu^{(j,s)}\}$ are linearly independent and they span $Null_k(\eta)$.

Next, we check that the columns of U satisfy the requirement in the claim, i.e., there exists a constant C_η such that for any (n,k) satisfying $k \ge n \ge h$,

$$\max_{1 \le i \le k-n, 1 \le j \le h} |U(i,j)|^2 \le C_\eta n^{-1}.$$

Since $\max_{1 \le j \le h} |U(i,j)| \le h \|(R'R)^{-1}\| \cdot \max_{1 \le j \le h} |R(i,j)|^2$, it suffices to show that

$$\text{(C.24)} \qquad \max_{1 \le i \le k-n, 1 \le j \le h} |R(i,j)|^2 \le C n^{-1},$$

and that for all $k \ge h$,

$$\text{(C.25)} \qquad \lambda_{\min}(R'R) \ge C > 0.$$

Consider (C.24) first. It is equivalent to show that

$$\text{(C.26)} \qquad \max_{1 \le i \le k-n} |\mu_i^{(j,s)}| / \|\mu^{(j,s)}\| \le C n^{-1/2}, \qquad 1 \le j \le m, 1 \le s \le h_j.$$

In the case $|z_j| > 1$, $\|\mu^{(j,s)}\| \le C$. In addition, $|z_j|^i \ge C i^{s-1/2}$ for sufficiently large i, and hence $\max_{1 \le i \le k-n} |\mu_i^{(j,s)}| \le \max_{i > n} C(i^{s-1} i^{1/2-s}) \le C n^{-1/2}$. So (C.26) holds. In the case $|z_j| = 1$, it can be shown in analysis that $\|\mu^{(j,s)}\| \ge C k^{s-1/2}$, where $C > 0$ is a constant depending on θ_j but independent of k. Also, $\max_{1 \le i \le k-n} |\mu_i^{(j,s)}| \le \max_{n < i \le k} C i^{s-1} \le C k^{s-1}$. Hence, $\max_{1 \le i \le k-n} |\mu_i^{(j,s)}| / \|\mu^{(j,s)}\| \le C k^{-1/2} \le C n^{-1/2}$ and (C.26) holds.

Next, consider (C.25). $R'R$ is an $h \times h$ matrix. For convenience, we use $\{(j,s) : 1 \le j \le m, 1 \le s \le h_j\}$ to index the entries in $R'R$. By construction, all the diagonals of $R'R$ are equal to 1, and the off-diagonals are equal to

$$\text{(C.27)} \qquad (R'R)_{(j,s),(j',s')} = \frac{\langle \mu^{(j,s)}, \mu^{(j',s')} \rangle}{\|\mu^{(j,s)}\| \|\mu^{(j',s')}\|}, \quad (j,s) \ne (j',s').$$

It is easy to see that as $k \to \infty$, each entry of $R'R$ has a finite limit. Therefore, as $k \to \infty$, $R'R$ approaches a fixed $h \times h$ matrix A element-wise.

In particular, $\lambda_{\min}(R'R) \to \lambda_{\min}(A)$. Hence, to show (C.25), we only need to prove that A is non-singular.

Write $R = (R_1, R_2)$, where R_1 is the submatrix formed by columns corresponding to those roots $|z_j| > 1$, and R_2 the submatrix formed by columns corresponding to those roots $|z_j| = 1$. Note that when $|z_j| = 1$ and $|z_{j'}| > 1$, as $k \to \infty$, $|\langle \mu^{(j,s)}, \mu^{(j',s')}\rangle| \leq C$, $\|\mu^{(j,s)}\| \to \infty$ and $\|\mu^{(j',s')}\| \geq C$; so $(R'R)_{(j,s),(j',s')} \to 0$. This means $R_1'R_2$ approaches the zero matrix as $k \to \infty$. Consequently,

$$A = \mathrm{diag}(A_1, A_2), \qquad \text{where } \; R_1'R_1 \to A_1 \; \text{ and } \; R_2'R_2 \to A_2, \quad \text{as } k \to \infty.$$

Therefore, it suffices to show that both A_1 and A_2 are non-singular.

Consider A_1 first. Denote $h_0 = \sum_j h_j 1\{|z_j| > 1\}$ so that R_1 is a $k \times h_0$ matrix. Let R_1^* be the $k \times h_0$ matrix whose columns are $\{\mu^{(j,s)} : |z_j| > 1\}$, M be the $h_0 \times h_0$ submatrix formed by the last h_0 rows of R^* and $\Lambda = \mathrm{diag}(\|\mu^{(j,s)}\|)$ is the $h_0 \times h_0$ diagonal matrix. Now, suppose A_1 is singular, i.e., there exists a non-zero vector b such that $b'A_1 b = 0$. This implies $\|R_1 b\| \to 0$ as $k \to \infty$. Using the matrices defined above, we can write $R_1 = R_1^*\Lambda$; so $\|R_1^*\Lambda b\| \to 0$. Since $\|M\Lambda b\| \leq \|R_1^*\Lambda b\|$, it further implies $\|M\Lambda b\| \to 0$. First, we observe that M is a fixed matrix independent of k. Second, note that when $|z_j| > 1$, $\|\mu^{(j,s)}\| \to c_{js}$, as $k \to \infty$, for some constant $c_{js} > 0$; as a result, $\Lambda \to \Lambda^*$ as $k \to \infty$, where Λ^* is a positive definite diagonal matrix. Combining the two parts, $\|M\Lambda b\| \to 0$ implies $\|M(\Lambda^* b)\| = 0$, where $\Lambda^* b$ is a fixed non-zero vector. This means M is singular. Therefore, if we can prove M is non-singular, then by contradiction, A_1 is also non-singular.

Now, we show M is non-singular. Let $\widetilde{M}$ be the matrix by re-arranging the rows in M in the inverse order. It is easy to see that M is non-singular if and only if $\widetilde{M}$ is non-singular. For convenience, we use $\{1, \cdots, h_0\} \times \{(j,s) : |z_j| > 1, 1 \leq s \leq h_j\}$ to index the entries in $\widetilde{M}$. It follows by the construction that

$$\widetilde{M}_{i,(j,s)} = i^{s-1} z_j^{-(i-1)}, \qquad z_j \text{ is a real}, \; 1 \leq i \leq h_0$$
$$\left.\begin{array}{l} \widetilde{M}_{i,(j-,s)} = i^{s-1}|z_j|^{-(i-1)}\cos((i-1)\theta_j), \\ \widetilde{M}_{i,(j-,s)} = i^{s-1}|z_j|^{-(i-1)}\sin((i-1)\theta_j), \end{array}\right. \quad z_{j\pm} \text{ are conjugates}, \; 1 \leq i \leq h_0.$$

Define an $h_0 \times h_0$ matrix T by

$$T_{i,(j,s)} = i^{s-1} z_j^{-(i-1)}, \quad 1 \leq i \leq h_0.$$

Let V be the $h_0 \times h_0$ confluent Vandermonde matrices generated by $\{z_j^{-1} : |z_j| > 1\}$:

$$V_{i,(j,s)} = \begin{cases} 0 & 1 \leq i \leq s-1, \\ \frac{(i-1)!}{(i-s)!} z_j^{-(i-s)} & s \leq i \leq h_0. \end{cases}$$

First, it is seen that each column of T is a (complex) linear combination of columns in $\widetilde{M}$. Second, we argue that each column of V is a linear combination of columns in T. To see this, note that $V_{i,(j,s)}$ can be written in the form $V_{i,(j,s)} = g_{s-1}(i)z_j^{-(i-s)}$, where $g_{s-1}(x) = (x-1)(x-2)\cdots(x-s+1)$ is a polynomial of degree $s-1$. Let $c_0, \cdots, c_{s-1}$ be the coefficients of this polynomial. Then, for each $i > s$, $V_{i,(j,s)} = z_j^{-(i-s)}\sum_{l=0}^{s-1} c_l i^l = \sum_{l=1}^{s} \alpha_l T_{i,(j,l)}$, where $\alpha_l \equiv z_j^{s-1} c_{l-1}$. The argument follows. Finally, it is well known that $\det(V) \neq 0$. Combining these, we see that $\det(\widetilde{M}) \neq 0$. Therefore, $\widetilde{M}$ is non-singular.

Next, we show A_2 is non-singular. Note that $\sum_{i=1}^{k} i^s = \frac{k^{s+1}}{s+1}(1+o(1))$, $\sum_{i=1}^{k} i^s \cos^2((i-1)\theta) = \frac{k^{s+1}}{2(s+1)}(1+o(1))$ and $\sum_{i=1}^{k} i^s \sin^2((i-1)\theta) = \frac{k^{s+1}}{2(s+1)}(1+o(1))$, for $\theta \neq -\frac{\pi}{2}, 0, \frac{\pi}{2}$. Also, $\sum_{i=1}^{k} i^s \sin((i-1)\theta) = o(k^{s+1})$ for all θ, and $\sum_{i=1}^{k} i^s \cos((i-1)\theta) = o(k^{s+1})$ for $\theta \neq 0$. Using these arguments and basic equalities in trigonometric functions, we have

$$(R'R)_{(j,s),(j',s')} = o(1) + \begin{cases} \frac{\sqrt{(2s-1)(2s'-1)}}{s+s'-1}, & j = j', \\ 0, & \text{elsewhere.} \end{cases}$$

As a result, A_2 is a block-diagonal matrix, where each block corresponds to one z_j on the unit circle and is equal to the matrix $W(h_j)$, where h_j is the replication number of z_j and $W(h)(s,s') = \sqrt{(2s-1)(2s'-1)}/(s+s'-1)$, for $1 \leq s, s' \leq h$. Since such $W(h)$'s are non-singular, A_2 is non-singular.

□

REFERENCES

Frieze, A. M. and Molloy, M. (1999). Splitting an expander graph. *Journal of Algorithms* **33** 166–172.

Jin, J., Zhang, C.-H. and Zhang, Q. (2012). Optimality of Graphlet Screening in high dimensional variable selection. *arXiv preprint arXiv:1204.6452*.

Ke, Z. T., Jin, J. and Fan, J. (2013). Covariate assisted screening and estimation. *arXiv preprint arXiv:1205.4645*.

Lehmann, E. L. and Casella, G. (1998). *Theory of point estimation* **31**. Springer.

Rambour, P. and Seghier, A. (2005). Szego type trace theorems for the singular case. *BULLETIN DES SCIENCES MATHEMATIQUES* **129** 149–174.

Z. Ke
Department of Statistics
University of Chicago
Chicago, Illinois, 60637
USA
E-mail: zke@galton.uchicago.edu

J. Jin
Department of Statistics
Carnegie Mellon University
Pittsburgh, Pennsylvania, 15213
USA
E-mail: jiashun@stat.cmu.edu

J. Fan
Department of Operations Research
and Financial Engineering
Princeton University
Princeton, New Jersey, 08544
USA
E-mail: jqfan@princeton.edu