Econometrica Supplementary Material

SUPPLEMENT TO "ANALYSIS OF TESTING-BASED FORWARD MODEL SELECTION"

(*Econometrica*, Vol. 88, No. 5, September 2020, 2147–2173)

DAMIAN KOZBUR
Department of Economics, University of Zürich

THIS SUPPLEMENT PROVES Theorems 2 and 3, supporting lemmas for Theorems 1 and 4, and Theorem 5.

S.1. PROOF OF THEOREMS 2 AND 3

Theorem 2 follows by applying Theorem 1 in the following way. If $\widehat{s}$ grows faster than s_0, then there is $m < \widehat{s}$ such that $s_0 < m < K_n$ and m/s_0 exceeds $c'_{\mathrm{F}}(K_n) = O(1)$, giving a contradiction. The first statement of the theorem follows from applying the bound on $\widehat{s}$. Theorem 3 follows by $\|\theta_0 - \widehat{\theta}\|_1 \leq \sqrt{\widehat{s} + s_0}\|\theta_0 - \widehat{\theta}\|_2 \leq \sqrt{\widehat{s} + s_0}\varphi_{\min}(\widehat{s} + s_0)(G)^{-1}\mathbb{E}_n[(x_i'\theta_0 - x_i'\widehat{\theta})^2]^{1/2}$.

S.2. PROOF OF LEMMAS 3 AND 4

S.2.1. *Proof of Lemma 3*

It was already shown that $\ell(\widehat{\theta}) \leq \ell(\theta_0) + s_0 t \varphi_{\min}(\widehat{s} + s_0)(G)^{-1}$. Expanding the above two quadratics in $\ell(\cdot)$ gives

$$\begin{aligned}\mathbb{E}_n\big[\big(x_i'\theta_0 - x_i'\widehat{\theta}\big)^2\big] &\leq \big|2\mathbb{E}_n\big[\varepsilon_i x_i'(\widehat{\theta} - \theta_0)\big]\big| + s_0 t \varphi_{\min}(\widehat{s} + s_0)(G)^{-1} \\ &\leq 2\big\|\mathbb{E}_n[\varepsilon_i x_i]\big\|_\infty \|\theta_0 - \widehat{\theta}\|_1 + s_0 t \varphi_{\min}(\widehat{s} + s_0)(G)^{-1}.\end{aligned}$$

To bound $\|\theta_0 - \widehat{\theta}\|_1$:

$$\begin{aligned}\|\theta_0 - \widehat{\theta}\|_1 &\leq \sqrt{\widehat{s} + s_0}\|\theta_0 - \widehat{\theta}\|_2 \\ &\leq \sqrt{\widehat{s} + s_0}\varphi_{\min}(\widehat{s} + s_0)(G)^{-1/2}\mathbb{E}_n\big[\big(x_i'\theta_0 - x_i'\widehat{\theta}\big)^2\big]^{1/2}.\end{aligned}$$

If $\mathbb{E}_n[(x_i'\theta_0 - x_i'\widehat{\theta})^2]^{1/2} = 0$, then the first conclusion of Theorem 1 holds. Otherwise, combining the above bounds and dividing by $\mathbb{E}_n[(x_i'\theta_0 - x_i'\widehat{\theta})^2]^{1/2}$ gives

$$\begin{aligned}\mathbb{E}_n\big[\big(x_i'\theta - x_i'\widehat{\theta}\big)^2\big]^{1/2} &\leq 2\big\|\mathbb{E}_n[\varepsilon_i x_i]\big\|_\infty \sqrt{\widehat{s} + s_0}\varphi_{\min}(\widehat{s} + s_0)(G)^{-1/2} \\ &\quad + \frac{s_0 t \varphi_{\min}(\widehat{s} + s_0)(G)^{-1}}{\mathbb{E}_n\big[\big(x_i'\theta_0 - x_i'\widehat{\theta}\big)^2\big]^{1/2}}.\end{aligned}$$

Damian Kozbur: damian.kozbur@econ.uzh.ch

 https://doi.org/10.3982/ECTA16273

Finally, either $\mathbb{E}_n[(x_i'\theta_0 - x_i'\widehat{\theta})^2]^{1/2} \leq \sqrt{s_0 t \varphi_{\min}(\widehat{s}+s_0)(G)^{-1/2}}$, in which case Lemma 3 holds, or alternatively $\mathbb{E}_n[(x_i'\theta_0 - x_i'\widehat{\theta})^2]^{1/2} > \sqrt{s_0 t \varphi_{\min}(\widehat{s}+s_0)(G)^{-1/2}}$, in which case

$$\mathbb{E}_n\bigl[\bigl(x_i'\theta - x_i'\widehat{\theta}\bigr)^2\bigr]^{1/2} \leq 2\bigl\|\mathbb{E}_n[\varepsilon_i x_i]\bigr\|_\infty \sqrt{\widehat{s}+s_0}\varphi_{\min}(\widehat{s}+s_0)(G)^{-1/2} + \sqrt{s_0 t \varphi_{\min}(\widehat{s}+s_0)(G)^{-1}}.$$

S.2.2. *Proof of Lemma 4*

For any S, define θ_S^* to be the minimizer of $\mathcal{E}(S)$. For any S, define also $d_S = \theta_S^* - \theta_{S_0\cup S}^*$. Finally, let $\delta_{0,S} = \theta_0 - \theta_{S_0\cup S}^*$. Note that $\mathcal{E}(S) - \mathcal{E}(S_0 \cup S) = d_S'\mathrm{E}[G]d_S$. By arguments in the earlier sections, $d_{\widehat{S}}'\mathrm{E}[G]d_{\widehat{S}} \leq s_0 c_{\text{test}}\varphi_{\min}(K_{\text{test}})(\mathrm{E}[G])^{-1}$. But $d_{\widehat{S}}'\mathrm{E}[G]d_{\widehat{S}} \geq \varphi_{\min}(K_{\text{test}})(\mathrm{E}[G])\|d_{\widehat{S}}\|_2^2$. So $\|d_{\widehat{S}}\|_2 \leq \sqrt{s_0 c_{\text{test}}}\varphi_{\min}(K_{\text{test}})(\mathrm{E}[G])^{-1}$. In addition, $\delta_{0,S}$ is bounded by

$$\begin{aligned}\|\delta_{0,S}\|_2 &= \bigl\|\mathrm{E}\bigl[\mathbb{E}_n\bigl[x_{iS_0\cup S}'\varepsilon_i\bigr]\bigr]\bigr\|_2 \\ &\leq \bigl(|S|+s_0\bigr)^{1/2}\max_j\bigl|\mathrm{E}\bigl[\mathbb{E}_n\bigl[x_{ij}\varepsilon_i^{\mathrm{a}}\bigr]\bigr]\bigr| \leq \frac{1}{2}\sqrt{\bigl(|S|+s_0\bigr)c_{\text{test}}}\varphi_{\min}(K_{\text{test}})\bigl(\mathrm{E}[G]\bigr)^{-1},\end{aligned}$$

where the last bound comes from Cauchy–Schwarz (passing to $\mathrm{E}[\mathbb{E}_n[x_{ij}^2]]^{1/2}\mathrm{E}[\mathbb{E}_n[\varepsilon_i^{\mathrm{a}2}]]^{1/2}$) along with the assumed condition on $\varepsilon_i^{\mathrm{a}}$ and the fact that $c_{\text{test}}' \leq c_{\text{test}}$. Next,

$$\begin{aligned}\widehat{\theta} &= G_{\widehat{S}}^{-1}\mathbb{E}_n\bigl[x_{i\widehat{S}}\bigl(x_{i\widehat{S}}'\theta_{\widehat{S}}^* + \varepsilon_i - x_{i\widehat{S}\cup S_0}'d_{\widehat{S}} + x_{i\widehat{S}\cup S_0}'\delta_{0,\widehat{S}}\bigr)\bigr] \\ &= \theta_{\widehat{S}}^* + G_{\widehat{S}}^{-1}\mathbb{E}_n[x_{i\widehat{S}}\varepsilon_i] + G_{\widehat{S}}^{-1}\mathbb{E}_n\bigl[x_{i\widehat{S}}x_{i\widehat{S}\cup S_0}'(-d_{\widehat{S}} + \delta_{0,\widehat{S}})\bigr] \\ \Rightarrow \quad & \bigl\|\widehat{\theta} - \theta_{\widehat{S}}^*\bigr\|_2 \leq \varphi_{\min}(\widehat{s})(G)^{-1/2}\bigl\|\mathbb{E}_n[x_{i\widehat{S}}\varepsilon_i]\bigr\|_2 + \bigl\|G_{\widehat{S}}^{-1}\mathbb{E}_n\bigl[x_{i\widehat{S}}x_{i\widehat{S}\cup S_0}(-d_{\widehat{S}} + \delta_{0,\widehat{S}})\bigr]\bigr\|_2 \\ &\qquad \leq \varphi_{\min}(\widehat{s})(G)^{-1/2}\widehat{s}^{1/2}\bigl\|\mathbb{E}_n[x_i\varepsilon_i]\bigr\|_\infty \\ &\qquad\quad + \varphi_{\min}(\widehat{s})(G)^{-1/2}\varphi_{\max}(\widehat{s}+s_0)(G)^{1/2}\bigl(\|d_{\widehat{S}}\|_2 + \|\delta_{0,\widehat{S}}\|_2\bigr).\end{aligned}$$

Finally,

$$\begin{aligned}&\bigl(\mathbb{E}_n\bigl[\bigl(x_i'\widehat{\theta} - x_i'\theta_0\bigr)^2\bigr]\bigr)^{1/2} \\ &\quad\leq \varphi_{\max}(s_0+\widehat{s})(G)^{1/2}\|\widehat{\theta} - \theta_0\|_2 \\ &\quad\leq \varphi_{\max}(s_0+\widehat{s})(G)^{1/2}\bigl(\bigl\|\widehat{\theta} - \theta_{\widehat{S}}^*\bigr\|_2 + \|\delta_0\|_2 + \|d_{\widehat{S}}\|_2\bigr) \\ &\quad\leq \varphi_{\max}(s_0+\widehat{s})(G)^{1/2}\varphi_{\min}(s_0+\widehat{s})(G)^{-1/2}\widehat{s}^{1/2}\bigl\|\mathbb{E}_n[x_i\varepsilon_i]\bigr\|_\infty \\ &\qquad + \varphi_{\max}(s_0+\widehat{s})(G)^{1/2}\biggl(\frac{3}{2} + \frac{3}{2}\varphi_{\max}(s_0+\widehat{s})(G)^{1/2}\varphi_{\min}(\widehat{s}+s_0)(G)^{-1/2}\biggr) \\ &\qquad \times \sqrt{(\widehat{s}+s_0)c_{\text{test}}}\varphi_{\min}(K_{\text{test}})\bigl(\mathrm{E}[G]\bigr)^{-1} \\ &\quad\leq \varphi_{\max}(s_0+\widehat{s})(G)^{1/2}\varphi_{\min}(s_0+\widehat{s})(G)^{-1/2}\widehat{s}^{1/2}\bigl\|\mathbb{E}_n[x_i\varepsilon_i]\bigr\|_\infty \\ &\qquad + 3\varphi_{\max}(s_0+\widehat{s})(G)\varphi_{\min}(\widehat{s}+s_0)(G)^{-1/2}\sqrt{(\widehat{s}+s_0)c_{\text{test}}}\varphi_{\min}(K_{\text{test}})\bigl(\mathrm{E}[G]\bigr)^{-1}.\end{aligned}$$

S.3. PROOF OF SUPPORTING LEMMAS FOR SPARSITY BOUNDS FOR THEOREMS 1 AND 4

S.3.1. *Additional Notation*

Additional notation is used for the proof of the lemmas which follow. The inner product from H is hereafter denoted simply with $\langle\cdot,\cdot\rangle_{\mathsf{H}} = \langle\cdot,\cdot\rangle$. The symbol $'$ is kept for use for transposition of finite-dimensional real matrices and vectors derived from certain elements of H defined below. For $a, b \in L^2(\Omega, \mathbb{R}^n)$, $a'b$ is defined pointwise (over Ω) and thus defines a random variable $\Omega \to \mathbb{R}$ and $\langle a, b\rangle = \mathrm{E}[a'b]$. In the case of Theorem 1, $a'b = \langle a, b\rangle$.

Let $V = [v_1, \dots, v_{s_0}]$ with the understanding that V and similar quantities are formally defined as linear mappings $\mathbb{R}^{s_0} \to \mathsf{H}$. Then $y = V\theta_0 + \varepsilon$ is well defined for both Theorems 1 and 4.

Let $\mathcal{M}_k$ denote projection in H onto the space orthogonal to $\operatorname{span}(\{v_1, \dots, v_k\})$. Then note that $\tilde{v}_k = \frac{\mathcal{M}_{k-1} v_k}{\langle v_k, \mathcal{M}_{k-1} v_k\rangle^{1/2}}$ for $k = 1, \dots, s_0$. In addition, $\tilde{\varepsilon} = \frac{\mathcal{M}_{s_0}\varepsilon}{\langle \varepsilon, \mathcal{M}_{s_0}\varepsilon\rangle^{1/2}}$. For more general sets S, let $\mathcal{Q}_S$ be projection onto the space orthogonal to $\operatorname{span}(\{x_j, j \in S\})$. For each selected covariate, w_j, set $S_{\text{pre-}w_j}$ to be the set of (both true and false) covariates selected prior to w_j.

S.3.2. *Proof of Lemma 5*

It is needed to calculate C_1, C_2 such that $\tilde{\gamma}_j'\tilde{\theta} \ge \tilde{\theta}_k C_1$ for $j \in A_{1k}$ and $\tilde{\theta}_k \ge \tilde{\theta}_l C_2$ for $l > k$. Define

$$\Delta_j \ell^{\mathsf{H}}(S) = \begin{cases} \Delta_j \ell(S) \text{ in the case of Theorem 1,} \\ \Delta_j \mathcal{E}(S) \text{ in the case of Theorem 4.} \end{cases}$$

Also recall that $t_{\mathsf{H}} = t$ in the case of Theorem 1 and $t^{\mathsf{H}} = c'_{\text{test}}$ in the case of Theorem 4. Note that $c_{\text{test}''}$ is not defined in the context of Theorem 1. In the case of Theorem 1, during the proof of this lemma, c''_{test} is taken to be equal to 1.

A simple derivation can be made to show that

$$-\Delta_j \ell^{\mathsf{H}}(S_{\text{pre-}w_j}) = \frac{1}{n}\langle y, \tilde{w}_j\rangle\big(\langle \tilde{w}_j, \tilde{w}_j\rangle\big)^{-1}\langle \tilde{w}_j, y\rangle = \frac{1}{n}\frac{1}{\|\tilde{w}_j\|_{\mathsf{H}}^2}\big(\tilde{\theta}'\tilde{\gamma}_j + \tilde{\theta}_{\tilde{\varepsilon}}\tilde{\gamma}_{j\tilde{\varepsilon}}\big)^2.$$

Note the slight abuse of notation in $-\Delta_j \ell^{\mathsf{H}}(S_{\text{pre-}w_j})$ signifying change in loss under inclusion of w_j rather than x_j. Next,

$$\big(\tilde{\theta}'\tilde{\gamma}_j + \tilde{\theta}_{\tilde{\varepsilon}}\tilde{\gamma}_{j\tilde{\varepsilon}}\big)^2 \le 2\big(\tilde{\theta}'\tilde{\gamma}_j\big)^2 + 2(\tilde{\theta}_{\tilde{\varepsilon}}\tilde{\gamma}_{j\tilde{\varepsilon}})^2.$$

Since $\tilde{\theta}_{\tilde{\varepsilon}} = \langle \tilde{\varepsilon}, y\rangle = \langle \varepsilon, \mathcal{M}_{s_0} y\rangle / \langle \varepsilon, \mathcal{M}_{s_0}\varepsilon\rangle^{1/2} = \langle \varepsilon, \mathcal{M}_{s_0}\varepsilon\rangle^{1/2}$, $\|\tilde{w}_j\|_{\mathsf{H}}^2 \ge 1$, and $j \in A_1$, it follows that

$$\frac{1}{n}\frac{1}{\|\tilde{w}_j\|_{\mathsf{H}}^2}(\tilde{\theta}_{\tilde{\varepsilon}}\tilde{\gamma}_{j\tilde{\varepsilon}})^2 \le \frac{1}{n}\frac{1}{\|\tilde{w}_j\|_{\mathsf{H}}^2}\tilde{\theta}_{\tilde{\varepsilon}}^2\left(\frac{t_{\mathsf{H}}^{1/2} n^{1/2}}{\big(3\langle \varepsilon, \mathcal{M}_{s_0}\varepsilon\rangle\big)^{1/2}}\right)^2 \le \frac{t_{\mathsf{H}}}{3}.$$

This implies

$$\frac{1}{2}\big(-\Delta_j \ell^{\mathsf{H}}(S_{\text{pre-}w_j})\big) \le \frac{1}{n}\frac{1}{\|\tilde{w}_j\|_{\mathsf{H}}^2}\big(\tilde{\theta}'\tilde{\gamma}_j\big)^2 + \frac{t_{\mathsf{H}}}{3}.$$

By the condition that the false j is selected, it holds that $-\Delta_j \ell^{\mathsf{H}}(S_{\text{pre-}w_j}) > t_{\mathsf{H}}$ and so $\frac{1}{3}(-\Delta_j \ell^{\mathsf{H}}(S_{\text{pre-}w_j})) > \frac{t_{\mathsf{H}}}{3}$, which implies that $-\frac{t_{\mathsf{H}}}{3} > \frac{1}{3}\Delta_j \ell^{\mathsf{H}}(S_{\text{pre-}w_j})$ and

$$\frac{1}{2}\big(-\Delta_j \ell^{\mathsf{H}}(S_{\text{pre-}w_j})\big) - \frac{t_{\mathsf{H}}}{3} \geq \frac{1}{6}\big(-\Delta_j \ell^{\mathsf{H}}(S_{\text{pre-}w_j})\big).$$

Finally, this yields that

$$\frac{1}{n\|\tilde{w}_j\|_{\mathsf{H}}^2}\big(\tilde{\gamma}_j'\tilde{\theta}\big)^2 \geq \frac{1}{6}\big(-\Delta_j \ell^{\mathsf{H}}(S_{\text{pre-}w_j})\big).$$

By the fact that w_j was selected ahead of v_k, it holds that

$$-\Delta_j \ell^{\mathsf{H}}(S_{\text{pre-}w_j}) \geq -\Delta_k \ell^{\mathsf{H}}(S_{\text{pre-}w_j}) c''_{\text{test}}.$$

Next, to lower bound $-\Delta_k \ell^{\mathsf{H}}(S_{\text{pre-}w_j})$, define a perturbed version of ℓ^{H}. Let $\xi \in \mathsf{H}$. Let $\ell^{\mathsf{H}}_{y+\xi}$ be defined analogously to ℓ^{H} except with the role of y in ℓ played by $y+\xi$ in $\ell^{\mathsf{H}}_{y+\xi}$. Choose ξ such that $\langle \xi, w_j\rangle = 0$ for $j = 1, \dots, m$, $\langle \xi, v_k\rangle = 0$ for $v_k = 1, \dots, s_0$, and $\langle \xi, \varepsilon\rangle = 0$. In the case of Theorem 1, $\xi \neq 0$ exists provided $m + s_0 + 1 < n$. If not, then H can be enlarged appropriately to allow ξ to exist, for example, with the inclusion $\iota : H \to H \oplus \mathbb{R}$, $x \mapsto (x, 0)$, $\xi = (0, 1)$. Then, due to the orthogonality of ξ to w_j and v_k and ε, it follows that

$$-\Delta_k \ell^{\mathsf{H}}(S_{\text{pre-}j}) = -\Delta_k \ell^{\mathsf{H}}_{y+\xi}(S_{\text{pre-}j}),$$

with the right-hand side possibly defined on an enlarged H as described above.

Next, the following reduction holds:

$$\begin{aligned}-\Delta_k \ell^{\mathsf{H}}_{y+\xi}(S_{\text{pre-}w_j}) &\geq -\Delta_k \ell^{\mathsf{H}}_{y+\xi}\big(S_{\text{pre-}w_j} \cup \{\tilde{v}_{k+1}\tilde{\theta}_{k+1} + \cdots + \tilde{v}_{s_0}\tilde{\theta}_{s_0} + \tilde{\varepsilon} + \xi\}\big)\\ &= -\Delta_{\tilde{v}_k} \ell^{\mathsf{H}}_{y+\xi}\big(S_{\text{pre-}w_j} \cup \{\tilde{v}_{k+1}\tilde{\theta}_{k+1} + \cdots + \tilde{v}_{s_0}\tilde{\theta}_{s_0} + \tilde{\varepsilon} + \xi\}\big).\end{aligned}$$

Let $\overset{\leftrightarrow\xi}{\mathcal{M}}_k$ be projection on the corresponding orthogonal space to the span of the vectors listed in $S_{\text{pre-}w_j} \cup \{\tilde{v}_{k+1}\tilde{\theta}_{k+1} + \cdots + \tilde{v}_{s_0}\tilde{\theta}_{s_0} + \tilde{\varepsilon} + \xi\}$. (The accent $\overset{\leftrightarrow}{\cdot}$ is meant to emphasize that covariates selected before and after v_k (or not at all) are considered.) Then the above term is further reduced by

$$= \frac{1}{n}\frac{\langle (y+\xi), \overset{\leftrightarrow\xi}{\mathcal{M}}_k \tilde{v}_k\rangle^2}{\langle \tilde{v}_k, \overset{\leftrightarrow\xi}{\mathcal{M}}_k \tilde{v}_k\rangle} = \frac{1}{n}\frac{\langle \tilde{\theta}_k\tilde{v}_k, \overset{\leftrightarrow\xi}{\mathcal{M}}_k \tilde{v}_k\rangle^2}{\langle \tilde{v}_k, \overset{\leftrightarrow\xi}{\mathcal{M}}_k \tilde{v}_k\rangle} = \frac{1}{n}\tilde{\theta}_k^2 \langle \tilde{v}_k, \overset{\leftrightarrow\xi}{\mathcal{M}}_k \tilde{v}_k\rangle.$$

Then seek a lower bound on $\frac{1}{n}\langle \tilde{v}_k, \overset{\leftrightarrow\xi}{\mathcal{M}}_k \tilde{v}_k\rangle$. Note that for some vector η_k, it holds that $\tilde{v}_k = \langle v_k, \mathcal{M}_{k-1} v_k\rangle^{-1/2} v_k - [v_1, \dots, v_{k-1}]\eta_k$. Then $\langle \tilde{v}_k, \overset{\leftrightarrow\xi}{\mathcal{M}}_k \tilde{v}_k\rangle = \langle v_k, \mathcal{M}_{k-1} v_k\rangle^{-1}\langle v_k, \overset{\leftrightarrow\xi}{\mathcal{M}}_k v_k\rangle$. Let $H = [V \;\; W]$. Let $\tilde{y}_k = \tilde{v}_{k+1}\tilde{\theta}_{k+1} + \cdots + \tilde{v}_{s_0}\tilde{\theta}_{s_0} + \tilde{\varepsilon}$. A lower bound on the term $\langle v_k, \overset{\leftrightarrow\xi}{\mathcal{M}}_k v_k\rangle$ follows from a lower bound on the eigenvalues of the below matrix for any $c > 0$:

$$\langle v_k, \overset{\leftrightarrow\xi}{\mathcal{M}}_k v_k\rangle \geq \lambda_{\min}\big(\big\langle \big[H(\tilde{y}_k + \xi)c\big], \big[H(\tilde{y}_k + \xi)c\big]\big\rangle\big).$$

That is, it is enough to bound the spectrum of $nG_{c,\xi}$ defined by

$$G_{c,\xi} = \frac{1}{n}\begin{bmatrix} \langle H, H\rangle & c\langle \tilde{y}_k + \xi, H\rangle \\ \langle H, \tilde{y}_k + \xi\rangle c & c^2\langle \tilde{y}_k + \xi, \tilde{y}_k + \xi\rangle \end{bmatrix}.$$

Using the fact that ξ is orthogonal to H and ε, $G_{c,\xi}$ reduces to

$$G_{c,\xi} = \frac{1}{n}\begin{bmatrix} \langle H, H\rangle & c\langle \tilde{y}_k, H\rangle \\ \langle H, \tilde{y}_k\rangle c & c^2\langle \tilde{y}_k, \tilde{y}_k\rangle + c^2\langle \xi, \xi\rangle \end{bmatrix}.$$

As a result of the above reductions, for each c, ξ,

$$-\Delta_k \ell^{\mathsf{H}}(S_{\text{pre-}w_j}) \geq \frac{1}{n}\langle v_k, \mathcal{M}_{k-1} v_k\rangle^{-1} n \lambda_{\min}(G_{c,\xi})\tilde{\theta}_k^2.$$

And therefore,

$$-\Delta_k \ell^{\mathsf{H}}(S_{\text{pre-}w_j}) \geq \frac{1}{n}\langle v_k, \mathcal{M}_{k-1} v_k\rangle^{-1} n \tilde{\theta}_k^2 \lim_{\substack{c\to 0 \\ \{\frac{1}{n}\langle \xi,\xi\rangle = c^{-2}\}}} \lambda_{\min}(G_{c,\xi}).$$

By continuity of eigenvalues for symmetric matrices, passing to the limit gives

$$\begin{aligned} -\Delta_k \ell^{\mathsf{H}}(S_{\text{pre-}w_j}) &\geq \frac{1}{n}\langle v_k, \mathcal{M}_{k-1} v_k\rangle^{-1} n \tilde{\theta}_k^2 \lambda_{\min}\left(\frac{1}{n}\begin{bmatrix} \langle H, H\rangle & 0 \\ 0 & 1\end{bmatrix}\right) \\ &\geq \frac{1}{n}\langle v_k, \mathcal{M}_{k-1} v_k/n\rangle^{-1}\tilde{\theta}_k^2 \varphi_{\min}(m+s_0)(G_{\mathsf{H}}) \geq \frac{1}{n}\cdot 1 \cdot \tilde{\theta}_k^2 \varphi_{\min}(m+s_0)(G_{\mathsf{H}}). \end{aligned}$$

This gives

$$\frac{1}{n\|\tilde{w}_j\|_{\mathsf{H}}^2}\big(\tilde{\gamma}_j' \tilde{\theta}\big)^2 \geq c''_{\text{test}} \frac{1}{6}\frac{1}{n}\varphi_{\min}(m+s_0)(G_{\mathsf{H}})\tilde{\theta}_k^2.$$

Using the fact that $\|\tilde{w}_j\|_{\mathsf{H}} \geq 1$ implies that

$$\big(\tilde{\gamma}_j' \tilde{\theta}\big)^2 \geq \tilde{\theta}_k^2 c''_{\text{test}} \frac{1}{6}\varphi_{\min}(m+s_0)(G_{\mathsf{H}}).$$

Now suppose no true variables remain when j is selected. Then $\langle \tilde{w}_j, \tilde{w}_j\rangle = \langle \tilde{u}_j, \tilde{u}_j\rangle = 1$ and

$$-\Delta_j \ell^{\mathsf{H}}(S_{\text{pre-}w_j}) = \frac{1}{n}\tilde{\gamma}_{j\tilde{\varepsilon}}^2 \tilde{\theta}_{\tilde{\varepsilon}}^2 \geq t_{\mathsf{H}}.$$

Note that $\tilde{\theta}_{\tilde{\varepsilon}}$ is given by $\tilde{\theta}_{\tilde{\varepsilon}} = \langle \varepsilon, \mathcal{M}_{s_0}\varepsilon\rangle^{1/2}$. Therefore, $\tilde{\gamma}_{j\tilde{\varepsilon}}^2 \geq t_{\mathsf{H}}\frac{n}{\langle \varepsilon, \mathcal{M}_{s_0}\varepsilon\rangle}$. This implies that $j \in A_2$. Therefore, set

$$C_1^2 = c''_{\text{test}}\frac{1}{6}\varphi_{\min}(m+s_0)(G_{\mathsf{H}}).$$

Next, construct C_2. For each selected true covariate, v_k, set $S_{\text{pre-}v_k}$ to be the set of (both true and false) covariates selected prior to v_k. Note that

$$\frac{1}{n}\tilde{\theta}_k^2 = -\Delta_k \ell^{\mathsf{H}}\big(\{v_1, \ldots, v_{k-1}\}\big) \geq -\Delta_k \ell^{\mathsf{H}}(S_{\text{pre-}v_k})$$

since $\{v_1, \ldots, v_{k-1}\} \subseteq S_{\text{pre}-v_k}$. In addition, if v_k is selected before v_l (or v_l is not selected), then

$$-\Delta_k \ell^{\mathsf{H}}(S_{\text{pre-}v_k}) \geq c''_{\text{test}}\big(-\Delta_l \ell^{\mathsf{H}}(S_{\text{pre-}v_k})\big) \geq c''_{\text{test}} \varphi_{\min}(m+s_0)(G_{\mathsf{H}})\frac{1}{n}\tilde{\theta}_l^2.$$

Therefore, taking

$$C_2^2 = c''_{\text{test}} \varphi_{\min}(m+s_0)(G_{\mathsf{H}})$$

implies that $\tilde{\theta}_k \geq \tilde{\theta}_l C_2$ for any $l > k$.

As a final remark, consider the case that $\tilde{\theta}_k = 0$. Then $\tilde{\theta}_l = 0$ for $l > k$. Then if $j \in A_{1k}$, it follows that $\tilde{\gamma}'_j \tilde{\theta} = 0$. Therefore, using reasoning as above, $-\Delta_j \ell^{\mathsf{H}}(S_{\text{pre-}j}) = \frac{1}{n}\frac{1}{\|\tilde{w}_j\|_{\mathsf{H}}^2}(\tilde{\theta}_{\tilde{\varepsilon}}\tilde{\gamma}_{j\tilde{\varepsilon}})^2 \leq \frac{t_{\mathsf{H}}}{3}$. But this is impossible, because being selected into the model requires $-\Delta_j \ell^{\mathsf{H}}(S_{\text{pre-}j}) > t_{\mathsf{H}}$. Therefore, A_{1k} is empty if $\tilde{\theta}_k = 0$.

S.3.3. *Proof of Lemma 6*

The desired element $\bar{Z}$ of $\mathcal{G}_{s_0}$ is constructed as the covariance matrix of certain real, mean-zero, random vectors

$$\mathsf{X} = (\mathsf{X}_k)_{k=1}^{s_0}, \qquad \mathsf{Y} = (\mathsf{Y}_l)_{l=1}^{s_0}.$$

The random variables X_k, Y_l constituting X, Y are defined as follows. Let $\beta_k = \tilde{\theta}_k / \tilde{\theta}_{k-1}$ for $k = 2, \ldots, s_0$. Then note that the components of B can be expressed $B_{kl} = \prod_{q=k+1}^{l} \beta_q$ for $k < l$, and extended symmetrically for components $l < k$.

Decompose the elements of the sequence β_k into

$$\beta_k = \beta_k^a \beta_k^b$$

in such a way that for all $l \geq k \geq 2$,

$$C_2 \leq \prod_{q=k}^{l} \beta_q^a \leq C_2^{-1},$$

and for all $k \geq 2$,

$$0 \leq \beta_k^b \leq 1.$$

Induction establishes the existence of such a decomposition with the additional property that: $\beta_k^a > \beta_k$ only if there is $q \leq k$ such that $\beta_q^a \cdot \ldots \cdot \beta_k^a = C_2$. The case $s_0 = 2$ follows by taking $\beta_2^a = \max\{C_2, \beta_2\}$ and noting that $\beta_2 = \tilde{\theta}_2/\tilde{\theta}_1 \leq C_2^{-1}$. Assume the complete induction hypothesis that the decomposition exists for sequences with $s_0 = 2, \ldots, s$ for some s. Consider a sequence $\beta_2, \ldots, \beta_{s+1}$. Apply the decomposition to obtain $\beta_k = \beta_k^a \beta_k^b$ for $k \leq s$. The existence of the decomposition fails at $k = s+1$ only if $\beta_{s+1} > 1$ and there is an index j such that $\beta_j^a \cdot \ldots \cdot \beta_s^a \cdot \beta_{s+1} > C_2^{-1}$. Then there must be an index $o \geq j$ such that $\beta_o^a > \beta_o$ as otherwise this contradicts $\tilde{\theta}_{s+1}/\tilde{\theta}_{j-1} \leq C_2^{-1}$. If there are multiple such indices o, then consider the largest one. There must then also be an index q such that $\beta_q^a \cdot \ldots \cdot \beta_o^a = C_2$. There are two cases to consider: $q < j$ and $q \geq j$. Consider the first case

$q < j$. In this case, the above conclusions can be visualized by:

$$\underbrace{\underbrace{\beta_q^a \cdot \ldots \cdot \beta_{j-1}^a \beta_j^a \cdot \ldots \cdot \beta_o^a}_{=C_2} \underbrace{\beta_{o+1} \cdot \ldots \cdot \beta_{s+1}}_{\leq C_2^{-1}}}_{\leq 1} \text{ with } \overbrace{\beta_j^a \cdot \ldots \cdot \beta_o^a \beta_{o+1} \cdot \ldots \cdot \beta_{s+1}}^{>C_2^{-1}}.$$

These imply that $\beta_q^a \cdot \ldots \cdot \beta_{j-1}^a < C_2$ which contradicts the inductive hypothesis. The case $q \geq j$ is similar. This completes the inductive argument and therefore establishes the decomposition $\beta_k = \beta_k^a \beta_k^b$, $k = 2, \ldots, s_0$, for all s_0.

Using the fact that $\beta_k^b \leq 1$ for all k allows the definition of the following autoregressive process. Let $\mathsf{U}_1 \sim \mathrm{N}(0, 1)$ and let $\mathsf{W}_1 = \mathsf{U}_1$. Define $\mathsf{U}_k \sim \mathrm{N}(0, 1)$ independently of previous random variables. Define W_k inductively as

$$\mathsf{W}_k = \beta_k^b \cdot \mathsf{W}_{k-1} + \sqrt{1 - \left(\beta_k^b\right)^2} \cdot \mathsf{U}_k.$$

Note that $\mathrm{E}[\mathsf{W}_k^2] = 1$ and $\mathrm{E}[\mathsf{W}_k \mathsf{W}_l] = \prod_{q=k+1}^{l} \beta_q^b$ if $k < l$. Then set X_k, Y_l as follows:

$$\mathsf{X}_k = C_2 \mathsf{W}_k \left(\prod_{q=2}^{k} \beta_q^a \right)^{-1/2} \left(\prod_{q=k+1}^{s_0} \beta_q^a \right)^{1/2},$$

$$\mathsf{Y}_l = C_2 \mathsf{W}_l \left(\prod_{q=l+1}^{s_0} \beta_q^a \right)^{-1/2} \left(\prod_{q=2}^{l} \beta_q^a \right)^{1/2}.$$

By construction,

$$\mathrm{E}[\mathsf{X}_k \mathsf{Y}_l] = C_2^2 B_{kl} \quad \text{for } k \leq l.$$

Next, note that $\mathrm{E}[\mathsf{X}_k^2] \leq 1$ and $\mathrm{E}[\mathsf{Y}_l^2] \leq 1$. This then implies (taking H_1 to be the span of $\mathsf{U}_1, \ldots, \mathsf{U}_{s_0}$ within the set of square integrable random variables) that both

$$\mathrm{E}\big[\mathsf{X}\mathsf{Y}'\big] \in \mathcal{G}_{s_0} \quad \text{and} \quad \mathrm{E}\big[\mathsf{X}\mathsf{Y}'\big]' \in \mathcal{G}_{s_0}.$$

Take $\bar{Z} = \mathrm{E}[\mathsf{X}\mathsf{Y}']'$. Let $C_3 = C_2^{-2}$. Note Γ is upper triangular due to the way $\tilde{\gamma}_j$ are defined. Because Γ is upper triangular, only lower triangular components of $\mathrm{E}[\mathsf{X}\mathsf{Y}']'$ matter for computing the product $\Gamma C_3 \bar{Z}$. Using this fact and the above calculations gives the desired factorization

$$\Gamma B = \Gamma C_3 \bar{Z} = \Gamma C_3 \mathrm{E}\big[\mathsf{X}\mathsf{Y}'\big]'.$$

S.3.4. *Proof of Lemma 8*

Collect the m_1 false selections into $\tilde{W} = [\tilde{w}_{j_1}, \ldots, \tilde{w}_{j_{m_1}}]$. Set $\tilde{R} = [\tilde{r}_{j_1}, \ldots, \tilde{r}_{j_{m_1}}]$, $\tilde{U} = [\tilde{u}_{j_1}, \ldots, \tilde{u}_{j_{m_1}}]$. Decompose $\tilde{W} = \tilde{R} + \tilde{U}$. Then $\langle \tilde{W}, \tilde{W} \rangle = \langle \tilde{R}, \tilde{R} \rangle + \langle \tilde{U}, \tilde{U} \rangle$. Here, the objects $\langle \tilde{W}, \tilde{W} \rangle$, $\langle \tilde{R}, \tilde{R} \rangle$, and $\langle \tilde{U}, \tilde{U} \rangle$, etc. are formally defined as $m_1 \times m_1$ real matrices with k, l entry given by $\langle \tilde{w}_k, \tilde{w}_l \rangle$, $\langle \tilde{r}_k, \tilde{r}_l \rangle$, $\langle \tilde{u}_k, \tilde{u}_l \rangle$, etc. (which, note, are genuine inner products on H).

Next, by the above normalization, $diag(\langle\tilde{U},\tilde{U}\rangle)=I$ if $\langle\tilde{u}_j,\tilde{u}_j\rangle=1$ for all $j\in A_1$. Recall that this normalization is possible provided $\varphi_{\min}(m+s_0)(G_{\mathsf{H}})>0$. Since $diag(\langle\tilde{U},\tilde{U}\rangle)=I$, it follows that the average inner product between the $\tilde{u}_j$, given by

$$\bar{\rho}=\frac{1}{m_1(m_1-1)}\sum_{j\neq l\in A_1}\langle\tilde{u}_j,\tilde{u}_l\rangle,$$

must be bounded below by

$$\bar{\rho}\geq-\frac{1}{m_1-1}$$

due to the positive definiteness of $\langle\tilde{U},\tilde{U}\rangle$. (This can be checked as a direct consequence of the fact that $1'_{m_1\times 1}\langle\tilde{U},\tilde{U}\rangle 1_{m_1\times 1}\geq 0$.) This implies an upper bound on the average off-diagonal term in $\langle\tilde{R},\tilde{R}\rangle$ since $\langle\tilde{W},\tilde{W}\rangle$ is a diagonal matrix. Since $\tilde{v}_k$ are orthonormal, the sum of all the elements of $\langle\tilde{R},\tilde{R}\rangle$ is given by $\|\sum_{j\in A_1}\tilde{\gamma}_j\|_2^2$. Since $\|\sum_{j\in A_1}\tilde{\gamma}_j\|_2^2=\sum_{j\in A_1}\|\tilde{\gamma}_j\|_2^2+\sum_{j\neq l\in A_1}\tilde{\gamma}_j'\tilde{\gamma}_l$ and since $\langle\tilde{W},\tilde{W}\rangle$ is a diagonal matrix, it must be the case that

$$\frac{1}{m_1(m_1-1)}\sum_{j\neq l\in A_1}\tilde{\gamma}_j'\tilde{\gamma}_l=-\bar{\rho}.$$

Therefore,

$$-\bar{\rho}=\frac{1}{m_1(m_1-1)}\left(\left\|\sum_{j\in A_1}\tilde{\gamma}_j\right\|_2^2-\sum_{j\in A_1}\|\tilde{\gamma}_j\|_2^2\right)\leq\frac{1}{m_1-1}.$$

This implies that

$$\left\|\sum_{j\in A_1}\tilde{\gamma}_j\right\|_2^2\leq m_1+\sum_{j\in A_1}\|\tilde{\gamma}_j\|_2^2.$$

Next, bound $\max_{j\in A_1}\|\tilde{\gamma}_j\|_2^2$.

Note $\|\tilde{\gamma}_j\|_2^2=\|\tilde{r}_j\|_{\mathsf{H}}^2$ since $\tilde{V}$ is orthonormal. Note that $\|\tilde{w}_j\|_{\mathsf{H}}^2$ is upper bounded by $\varphi_{\min}(m+s_0)(G)^{-1}$. To see this, note that $\|\tilde{w}_j\|_{\mathsf{H}}^2=\|c_j\mathcal{Q}_{\text{pre-}j}w_j\|_{\mathsf{H}}^2\leq c_j^2\|w_j\|_{\mathsf{H}}^2=c_j^2 n$, where c_j is the normalizing constant such that $\tilde{w}_j=c_j\mathcal{Q}_{\text{pre-}j}$. At the same time, c_j^2 satisfies $\|\mathcal{M}_{s_0}\mathcal{Q}_{\text{pre-}j}w_j\|_{\mathsf{H}}^2=c_j^{-2}$ whenever $w_j\notin\text{span}(\tilde{V})$. Note also that $\|\mathcal{M}_{s_0}\mathcal{Q}_{\text{pre-}j}w_j\|_{\mathsf{H}}^2\geq\|\mathcal{Q}_{S_0\cup\text{pre-}j}w_j\|_{\mathsf{H}}^2$, where the notation $\mathcal{Q}_{S_0\cup\text{pre-}j}$ denotes projection onto the space orthogonal to covariates indexed in S_0 or selected before w_j. To see this, consider an arbitrary Hilbert space $\check{\mathsf{H}}$, projections onto closed subspaces $1,2,12=\text{span}(1\cup 2)$, $\mathcal{P}_1$, $\mathcal{P}_2$, $\mathcal{P}_{12}$, projections onto the respective orthogonal complements $\mathcal{Q}_1$, $\mathcal{Q}_2$, $\mathcal{Q}_{12}$, and any vector w. Then $w=\mathcal{Q}_{12}w+\mathcal{P}_{12}w$. Then $\mathcal{Q}_2\mathcal{Q}_1w=\mathcal{Q}_2\mathcal{Q}_1\mathcal{Q}_{12}w+\mathcal{Q}_2\mathcal{Q}_1\mathcal{P}_{12}w=\mathcal{Q}_{12}w+\mathcal{Q}_2\mathcal{Q}_1\mathcal{P}_{12}w$. Note that the inner product between the above two terms vanishes: $\langle\mathcal{Q}_{12}w,\mathcal{Q}_2\mathcal{Q}_1\mathcal{P}_{12}w\rangle_{\check{\mathsf{H}}}=\langle w,\mathcal{Q}_{12}\mathcal{P}_{12}w\rangle_{\check{\mathsf{H}}}=\langle w,0w\rangle_{\check{\mathsf{H}}}=0$. Then by the Pythagorean theorem, $\|\mathcal{Q}_2\mathcal{Q}_1w\|_{\check{\mathsf{H}}}^2=\|\mathcal{Q}_{12}w\|_{\check{\mathsf{H}}}^2+\|\mathcal{Q}_2\mathcal{Q}_1\mathcal{P}_{12}w\|_{\check{\mathsf{H}}}^2\geq\|\mathcal{Q}_{12}w\|_{\check{\mathsf{H}}}^2$. So $\|\mathcal{Q}_{12}w\|_{\check{\mathsf{H}}}\leq\|\mathcal{Q}_2\mathcal{Q}_1w\|_{\check{\mathsf{H}}}$. Therefore, the quantity $\|\mathcal{Q}_{S_0\cup\text{pre-}j}w_j\|_{\mathsf{H}}^2$ is lower bounded by $n\varphi_{\min}(m+s_0)(G_{\mathsf{H}})$. As a result, $c_j^2\leq\varphi_{\min}(m+s_0)(G_{\mathsf{H}})^{-1}$, giving the desired bound on $\|\tilde{w}_j\|_{\mathsf{H}}^2$. Therefore, $\|\tilde{r}_j\|_{\mathsf{H}}^2=\|\tilde{w}_j\|_{\mathsf{H}}^2-1\leq\varphi_{\min}(m+s_0)(G_{\mathsf{H}})^{-1}-1$. It follows that

$$\max_{j\in A_1}\|\tilde{\gamma}_j\|_2^2\leq\varphi_{\min}(m+s_0)(G_{\mathsf{H}})^{-1}-1.$$

This then implies that

$$\Big\|\sum_{j\in A_1}\tilde{\gamma}_j\Big\|_2^2 \le m_1\varphi_{\min}(m+s_0)(G_{\mathsf{H}})^{-1}.$$

The same argument as above also shows that for any choice $e_j \in \{-1, 1\}$ of signs, it holds that

$$\Big\|\sum_{j\in A_1}e_j\tilde{\gamma}_j\Big\|_2^2 \le m_1\varphi_{\min}(m+s_0)(G_{\mathsf{H}})^{-1}.$$

(In more detail, take $\tilde{W}_e = [\tilde{w}_{j_1}e_{j_1}, \dots, \tilde{w}_{j_{m_1}}e_{j_{m_1}}]$, etc. and rerun the same argument.)

S.3.5. *Proof of Lemma 10*

In this proof, the number of elements of A_2 is bounded. Recall that the criterion for $j \in A_2$ is that $|\tilde{\gamma}_{j\tilde{\varepsilon}}| > \frac{t_{\mathsf{H}}^{1/2}n^{1/2}}{(3\langle\varepsilon,\mathcal{M}_{s_0}\varepsilon\rangle)^{1/2}}$. Note also that $\tilde{\gamma}_{j\tilde{\varepsilon}}$ is found by the coefficient in the expression

$$\tilde{\gamma}_{j\tilde{\varepsilon}} = \langle\tilde{\varepsilon}, \tilde{w}_j\rangle = \Big\langle\varepsilon, \frac{1}{\langle\varepsilon, \mathcal{M}_{s_0}\varepsilon\rangle^{1/2}}\mathcal{M}_{s_0}\tilde{w}_j\Big\rangle.$$

Next, let H be $H = [v_1, \dots, v_{s_0}, w_1, \dots, w_m]$. Note that

$$\frac{1}{\langle\varepsilon, \mathcal{M}_{s_0}\varepsilon\rangle^{1/2}}\mathcal{M}_{s_0}\tilde{w}_j \in \operatorname{span}(H).$$

Therefore,

$$\tilde{\gamma}_{j\tilde{\varepsilon}} = \langle\varepsilon, H\rangle\langle H, H\rangle^{-1}\Big\langle H, \frac{1}{\big(\langle\varepsilon, \mathcal{M}_{s_0}\varepsilon\rangle\big)^{1/2}}\mathcal{M}_{s_0}\tilde{w}_j\Big\rangle.$$

Let μ_j be the sign $+1$ for each $j \in A_2$ such that $\tilde{\gamma}_{j\tilde{\varepsilon}} > 0$ and -1 for each $j \in A_2$ such that $\tilde{\gamma}_{j\tilde{\varepsilon}} < 0$. By the fact that $j \in A_2$, $\tilde{\gamma}_{j\tilde{\varepsilon}}\mu_j > \frac{t_{\mathsf{H}}^{1/2}n^{1/2}}{(3\langle\varepsilon,\mathcal{M}_{s_0}\varepsilon\rangle)^{1/2}}$, summing over $j \in A_2$ gives

$$\sum_{j\in A_2}\langle\varepsilon, H\rangle\langle H, H\rangle^{-1}\Big\langle H, \frac{1}{\big(\langle\varepsilon, \mathcal{M}_{s_0}\varepsilon\rangle\big)^{1/2}}\mathcal{M}_{s_0}\tilde{w}_j\mu_j\Big\rangle > m_2\frac{t_{\mathsf{H}}^{1/2}n^{1/2}}{\big(3\langle\varepsilon, \mathcal{M}_{s_0}\varepsilon\rangle\big)^{1/2}}.$$

This implies that

$$\Big\|\langle H, H\rangle^{-1}\Big\langle H, \frac{1}{\big(\langle\varepsilon, \mathcal{M}_{s_0}\varepsilon\rangle\big)^{1/2}}\sum_{j\in A_2}\mathcal{M}_{s_0}\tilde{w}_j\mu_j\Big\rangle\Big\|_1\big\|\langle\varepsilon, H\rangle\big\|_\infty > m_2\frac{t_{\mathsf{H}}^{1/2}n^{1/2}}{\big(3\langle\varepsilon, \mathcal{M}_{s_0}\varepsilon\rangle\big)^{1/2}},$$

which further implies that

$$\sqrt{m+s_0}\Big\|\langle H, H\rangle^{-1}\Big\langle H, \frac{1}{\big(\langle\varepsilon, \mathcal{M}_{s_0}\varepsilon\rangle\big)^{1/2}}\sum_{j\in A_2}\mathcal{M}_{s_0}\tilde{w}_j\mu_j\Big\rangle\Big\|_2\big\|\langle\varepsilon, H\rangle\big\|_\infty > m_2\frac{t_{\mathsf{H}}^{1/2}n^{1/2}}{\big(3\langle\varepsilon, \mathcal{M}_{s_0}\varepsilon\rangle\big)^{1/2}}.$$

Next, further upper bound the $\|\cdot\|_2$ term on the left side above by

$$\left\| \langle H, H\rangle^{-1} \left\langle H, \frac{1}{\left(\langle \varepsilon, \mathcal{M}_{s_0}\varepsilon\rangle\right)^{1/2}}\right\rangle \sum_{j\in A_2} \mathcal{M}_{s_0}\tilde{w}_j\mu_j\rangle \right\|_2$$

$$\leq \frac{n^{-1/2}}{\left(\langle \varepsilon, \mathcal{M}_{s_0}\varepsilon\rangle\right)^{1/2}} \varphi_{\min}(s_0+m)(G_{\mathsf{H}})^{-1/2} \left\| \mathcal{M}_{s_0} \sum_{j\in A_2} \tilde{w}_j\mu_j \right\|_{\mathsf{H}}.$$

Next, by the fact that $\mathcal{M}_{s_0}$ is a projection (hence non-expansive) and $\tilde{w}_j$ are mutually orthogonal,

$$\leq \frac{n^{-1/2}}{\left(\langle \varepsilon, \mathcal{M}_{s_0}\varepsilon\rangle\right)^{1/2}} \varphi_{\min}(s_0+m)(G_{\mathsf{H}})^{-1/2} \sqrt{\sum_{j\in A_2} \|\tilde{w}_j\mu_j\|_{\mathsf{H}}^2}.$$

Earlier, it was shown that $\max_j \|\tilde{w}_j\|_{\mathsf{H}}^2 \leq \varphi_{\min}(s_0+m)(G_{\mathsf{H}})^{-1}$. Therefore, putting the above inequalities together,

$$\frac{n^{-1/2}}{\left(\langle \varepsilon, \mathcal{M}_{s_0}\varepsilon\rangle\right)^{1/2}} \sqrt{m+s_0}\varphi_{\min}(m+s_0)(G_{\mathsf{H}})^{-1}\sqrt{m_2}\left\|\langle \varepsilon, H\rangle\right\|_\infty > m_2 \frac{t_{\mathsf{H}}^{1/2} n^{1/2}}{\left(3\langle \varepsilon, \mathcal{M}_{s_0}\varepsilon\rangle\right)^{1/2}}.$$

This implies that

$$m_2 < \frac{1}{n^2}\frac{3}{t_{\mathsf{H}}}\left(\langle \varepsilon, \mathcal{M}_{s_0}\varepsilon\rangle\right)(m+s_0)\frac{\left\|\langle \varepsilon, H\rangle\right\|_\infty^2}{\varepsilon'\mathcal{M}_{s_0}\varepsilon}\varphi_{\min}(m+s_0)(G_{\mathsf{H}})^{-2}.$$

In the case of Theorem 1, this is further bounded by

$$\leq 3(m+s_0)\frac{\left\|\mathbb{E}_n[x_i\varepsilon_i]\right\|_\infty^2}{t}\varphi_{\min}(m+s_0)(G)^{-2}.$$

Under the assumed condition that $t^{1/2} \geq 2\|\mathbb{E}_n[x_i\varepsilon_i]\|_\infty \varphi_{\min}(m+s_0)(G)^{-1}$, it follows that

$$m_2 \leq \frac{3}{4}(m+s_0).$$

Similarly, the condition of Theorem 4 that $\mathrm{E}[\mathbb{E}_n[\varepsilon_i^{a2}]] \leq \frac{1}{2}\varphi_{\min}(\mathrm{E}[G])^{-1}c'_{\text{test}}$ yields $m_2 \leq \frac{3}{4}(m+s_0)$ in the same way. Finally, substituting $m = m_1 + m_2$ gives

$$m_2 \leq 3m_1 + 3s_0.$$

S.3.6. *Proof of Lemma 11*

Combining $m_1 \leq \varphi_{\min}(m+s_0)(G_{\mathsf{H}})^{-1}C_1^{-2}{C_3}^2(K_G^{\mathbb{R}})^2 s_0$ and $m_2 \leq 3(m_1+s_0)$ gives

$$m \leq \left[4\varphi_{\min}(m+s_0)(G_{\mathsf{H}})^{-1}C_1^{-2}{C_3}^2\left(K_G^{\mathbb{R}}\right)^2 + 3\right]s_0.$$

In addition, in the case of Theorem 1, $C_1^2 = \frac{1}{6}\varphi_{\min}(m+s_0)(G_{\mathsf{H}})$, $C_2^2 = \varphi_{\min}(m+s_0)(G_{\mathsf{H}})$, $C_3^2 = (C_2^{-2})^2 = \varphi_{\min}(m+s_0)(G_{\mathsf{H}})^{-2}$, $C_1^{-2}C_3^2 = 6\varphi_{\min}(m+s_0)(G_{\mathsf{H}})^{-3}$, and $K_G^{\mathbb{R}} < 1.783$.

Therefore, $m \le (3 + 24 \times 1.783^2 \times \varphi_{\min}(m+s_0)(G_{\mathsf{H}})^{-4})s_0$. Because $\varphi_{\min}(m+s_0)(G_{\mathsf{H}})^{-1} \ge 1$ and $3 + 24 \times 1.783^2 = 79.2981 < 80$, it holds that

$$m \le 80 \times \varphi_{\min}(m+s_0)(G_{\mathsf{H}})^{-4} s_0.$$

This bound holds for each positive integer m of wrong selections, provided $t^{1/2} \ge 2\varphi_{\min}(m+s_0)(G)^{-1}\|\mathbb{E}_n[x_i\varepsilon_i]\|_\infty$. This concludes the proof of the sparsity bound for Theorem 1. Using similar reasoning in the case of Theorem 4, on the event $\mathcal{T}$, it follows that $m \le 80 \times \varphi_{\min}(m+s_0)(G_{\mathsf{H}})^{-4} c_{\text{test}}^{\prime -3} s_0$ provided $\mathrm{E}[\mathbb{E}_n[\varepsilon_i^{\mathrm{a}2}]] \le \frac{1}{2}\varphi_{\min}(m+s_0)(\mathrm{E}[G])^{-1}c'_{\text{test}}$. Setting $m = K_{\text{test}} - s_0$ contradicts Condition 2 by $K_{\text{test}} \le 80 \times \varphi_{\min}(K_{\text{test}})(\mathrm{E}[G])^{-4} c''^{-3}_{\text{test}} + s_0 < K_{\text{test}}$. Therefore, $m < K_{\text{test}} - s_0$ and thus

$$\widehat{s} \le \big(80 \times \varphi_{\min}(K_{\text{test}})(G_{\mathsf{H}})^{-4} c_{\text{test}}^{\prime -3} + 1\big)s_0,$$

completing the proof of the sparsity bound for Theorem 4.

S.4. PROOF OF THEOREM 5

The strategy is to apply Theorem 4 using the conditional distribution P_x for $\mathcal{D}_n$, conditional on x. The unconditional result is then shown to follow. Let $\mathcal{E}_x(S) = \mathrm{E}[\ell(S)|x]$. In addition, for $j \notin S$, let $\theta_{jS}^{*|x} = (x'_{jS}x_{jS})^{-1}x'_{jS}\mathrm{E}[x'_{jS}(x\theta_0 + \varepsilon^{\mathrm{a}})|x]$ so that $[\theta_{jS}^{*|x}]_j = (x'_j\mathcal{Q}_S x_j)^{-1}\mathrm{E}[x'_j\mathcal{Q}_S(x\theta_0 + \varepsilon^{\mathrm{a}})|x]$. Throughout the proof of Theorem 5, use an abuse of notation by writing $\widehat{V}_{jS} = [\widehat{V}_{jS}]_{jj}$. Let

$$\widehat{Z}_{jS} = \widehat{V}_{jS}^{-1/2}\big([\widehat{\theta}_{jS}]_j - \big[\theta_{jS}^{*|x}\big]_j\big).$$

Let $t_\alpha = \Phi^{-1}(1-\alpha/p)$. Let $\mathcal{A}$ be the event given by

$$\mathcal{A} = \left\{|\widehat{Z}_{jS}| \le \left(\frac{1+c_\tau}{2}\right)\widehat{\tau}_{jS}t_\alpha \text{ for all } j, |S| < K_n\right\}.$$

Note that $-\Delta_j\mathcal{E}_x(S) = [\theta_{jS}^{*|x}]_j^2 A_{jS}$ for A_{jS} defined by $A_{jS} = [G_{jS}^{-1}]_{jj}$.

The next lemma states size, power, and continuity properties of the tests of Definition 1.

LEMMA 12: *The following implications are valid on $\mathcal{A}$ for all j, $|S| < K_n$:*

1. $T_{jS\alpha} = 1$ *if* $-\Delta_j\mathcal{E}_x(S) \ge A_{jS}\widehat{V}_{jS}(2c_\tau)^2\widehat{\tau}_{jS}^2 t_\alpha^2$.
2. $-\Delta_j\mathcal{E}_x(S) \ge A_{jS}\widehat{V}_{jS}(\frac{1-c_\tau}{2})^2\widehat{\tau}_{jS}^2 t_\alpha^2$ *if* $T_{jS\alpha} = 1$.
3. $-\Delta_k\mathcal{E}_x(S) \le \frac{\widehat{V}_{kS}A_{kS}}{\widehat{V}_{jS}A_{jS}}(1 + \frac{1+c_\tau}{c_\tau-1}(1 + \frac{\widehat{\tau}_{kS}}{\widehat{\tau}_{jS}}))^2(-\Delta_j\mathcal{E}_x(S))$ *if* $T_{jS\alpha} = 1$, $W_{jS} \ge W_{kS}$.

Next, define a sequence of sets $\mathcal{X} = \mathcal{X}_n$ which will be shown to have the property that both $\mathrm{P}(x \in \mathcal{X}) \to 1$ and

$$\mathrm{P}^{\mathcal{X}}(\mathcal{A}) = \operatorname*{ess\,inf}_{x\in\mathcal{X}} \mathrm{P}(\mathcal{A}|x) \to 1.$$

In addition, there will be constants $\tilde{c}_{\text{test}}, \tilde{c}'_{\text{test}}, c''_{\text{test}} > 0$ which are independent of n and the realization of x, such that for $c_{\text{test}} = \frac{1}{n}\tilde{c}_{\text{test}}$, $c'_{\text{test}} = \frac{1}{n}\tilde{c}'_{\text{test}}$ and for the set $\mathcal{B}$ defined by

$$\mathcal{B} = \begin{cases} 1. \quad A_{jS}\widehat{V}_{jS}(2c_\tau)^2\widehat{\tau}^2_{jS}t^2_\alpha \leq c_{\text{test}}, \\ 2. \quad A_{jS}\widehat{V}_{jS}\left(\frac{1-c_\tau}{2}\right)^2\widehat{\tau}^2_{jS}t^2_\alpha \geq c'_{\text{test}}, \\ 3. \quad \frac{A_{kS}\widehat{V}_{kS}}{A_{jS}\widehat{V}_{jS}}\left(1+\frac{1+c_\tau}{1-c_\tau}\left(1+\frac{\widehat{\tau}_{kS}}{\widehat{\tau}_{jS}}\right)\right)^2 \geq c''_{\text{test}}, \end{cases} \qquad |S| < K_n$$

it holds that $\mathrm{P}^{\mathcal{X}}(\mathcal{B}) \to 1$.

Define sets $\mathcal{X} = \mathcal{X}_n$ as follows. Set $\mathcal{X} = \mathcal{X}_1 \cap \mathcal{X}_2 \cap \mathcal{X}_3 \cap \mathcal{X}_4$ with

$\mathcal{X}_1 = \{x : \max_{j\leq p} \mathbb{E}_n[x_{ij}^{12}] = O(1)\}$,

$\mathcal{X}_2 = \{x : \varphi_{\min}(K_n)(G)^{-1} = O(1)\}$,

$\mathcal{X}_3 = \{x : \max_{j,|S|<K_n} \|\eta_{jS}\|_1 = O(1)\}$,

$\mathcal{X}_4 = \{x : \mathrm{P}(\varphi_{\min}(K_n)(\mathbb{E}_n[\varepsilon_i^2 x_i x_i'])^{-1} = O(1)|x) = 1 - o(1)\}$.

Note that $\mathrm{P}(\mathcal{X}_1), \mathrm{P}(\mathcal{X}_2), \mathrm{P}(\mathcal{X}_3) \to 1$ by assumption. In addition, failure of $\mathrm{P}(\mathcal{X}_4) \to 1$ would contradict the unconditional statement in Condition 4 that

$$\mathrm{P}\big(\varphi_{\min}(K_n)\big(\mathbb{E}_n\big[\varepsilon_i^2 x_i x_i'\big]\big)^{-1} = O(1)\big) = 1 - o(1).$$

Therefore, $\mathrm{P}(\mathcal{X}) \to 1$.

The next two sections prove the following two lemmas.

LEMMA 13: $\mathrm{P}^{\mathcal{X}}(\mathcal{A}) \to 1$.

LEMMA 14: $\mathrm{P}^{\mathcal{X}}(\mathcal{B}) \to 1$ *for some* c_{test}, c'_{test}, c''_{test} *as described in the definition of* $\mathcal{B}$ *above*.

The previous results show that for each n, Theorem 4 can be applied conditionally on x with c_{test}, c'_{test}, c''_{test} defined above, with $K_{\text{test}} = K_n - 1$, and with $1 - \alpha - \delta_{\text{test}} = \mathrm{P}^{\mathcal{X}}(\mathcal{A} \cap \mathcal{B})$. Note that renormalizing the covariates to satisfy $\mathbb{E}_n[x_{ij}^2] = 1$ does not affect $\mathcal{E}_x(S)$ and therefore does not affect the conclusions above. Moreover, on $\mathcal{X}$, renormalizing does not affect boundedness of sparse eigenvalues of G. The unconditional result is shown as follows. By Theorem 4,

$$\mathrm{P}^{\mathcal{X}}\big(\mathbb{E}_n\big[\big(x_i'\theta^{*|x}_{S_0} - x_i\widehat{\theta}\big)^2\big]^{1/2} \leq O(\sqrt{s_0 \log p/n})\big) \to 1.$$

Note that $\theta^{*|x}_{S_0} - \theta_0 = (x'_{S_0}x_{S_0})^{-1}x'_{S_0}\mathrm{E}[\varepsilon^{\mathrm{a}}|x]$. As a result,

$$\begin{aligned} \big\|\theta_0 - \theta^{*|x}_{S_0}\big\|_2 &\leq \varphi_{\min}(s_0)(G)^{-1/2}\big\|\mathbb{E}_n\big[x_{is_0}\mathrm{E}\big[\varepsilon_i^{\mathrm{a}}|x\big]\big]\big\|_2 \\ &\leq \varphi_{\min}(s_0)(G)^{-1/2}\sqrt{s_0}\big\|\mathbb{E}_n\big[x_{ij}\mathrm{E}\big[\varepsilon_i^{\mathrm{a}}|x\big]\big]\big\|_\infty. \end{aligned}$$

By the assumed rate conditions, sparse eigenvalue conditions, and by $\max_i \mathrm{E}[\varepsilon_i^{\mathrm{a}}] = O(n^{-1/2})$, the bound on $\|\theta_0 - \theta^{*|x}_{S_0}\|_2$ implies further that $\mathrm{P}^{\mathcal{X}}(\mathbb{E}_n[(x_i'\theta^{*|x}_{S_0} - x_i\theta_0)^2]^{1/2} \leq O(\sqrt{s_0 \log p/n})) \to 1$. Theorem 5 follows by triangle inequality.

S.5. PROOF OF SUPPORTING LEMMAS FOR THEOREM 5

S.5.1. *Proof of Lemma 12*

For this proof, work on $\mathcal{A}$ and suppose $|S| < K_n$. To prove the first statement, suppose that $-\Delta_j \mathcal{E}_x(S) \geq A_{jS}\widehat{V}_{jS}(2c_\tau)^2\widehat{\tau}_{jS}^2 t_\alpha^2$. Then

$$\begin{aligned}
\big[\theta_{jS}^{*|x}\big]_j^2 A_{jS} &\geq A_{jS}\widehat{V}_{jS}(2c_\tau)^2\widehat{\tau}_{jS}^2 t_\alpha^2, \\
\big|\big[\theta_{jS}^{*|x}\big]_j\big| &\geq \widehat{V}_{jS}^{1/2}(2c_\tau)\widehat{\tau}_{jS}t_\alpha, \\
\big|[\widehat{\theta}_{jS}]_j\big| &\geq \widehat{V}_{jS}^{1/2}(2c_\tau)\widehat{\tau}_{jS}t_\alpha - \big|\big[\theta_{jS}^{*|x}\big]_j - [\widehat{\theta}_{jS}]_j\big|, \\
\big|[\widehat{\theta}_{jS}]_j\big| &\geq \widehat{V}_{jS}^{1/2}(2c_\tau)\widehat{\tau}_{jS}t_\alpha - \widehat{V}_{jS}^{1/2}\left(\frac{1+c_\tau}{2}\right)\widehat{\tau}_{jS}t_\alpha, \\
\big|[\widehat{\theta}_{jS}]_j\big| &\geq \widehat{V}_{jS}^{1/2}c_\tau\widehat{\tau}_{jS}t_\alpha,
\end{aligned}$$

which implies $T_{jS\alpha} = 1$.

Next, prove the second statement. By construction, if $T_{jS\alpha} = 1$, then $|\widehat{V}_{jS}^{-1/2}[\widehat{\theta}_{jS}]_j| \geq c_\tau\widehat{\tau}_{jS}t_\alpha$, which is equivalent to

$$\big|[\widehat{\theta}_{jS}]_j\big| \geq c_\tau\widehat{\tau}_{jS}t_\alpha\widehat{V}_{jS}^{1/2}.$$

Note that $|[\widehat{\theta}_{jS}]_j - [\theta_{jS}^{*|x}]_j| \leq \widehat{V}_{jS}^{1/2}(\frac{1+c_\tau}{2})\widehat{\tau}_{jS}t_\alpha$. Then $T_{jS\alpha} = 1 \Rightarrow$

$$\big|\big[\theta_{jS}^{*|x}\big]_j\big| \geq c_\tau\widehat{\tau}_{jS}t_\alpha\widehat{V}_{jS}^{1/2} - \widehat{V}_{jS}^{1/2}\left(\frac{1+c_\tau}{2}\right)\widehat{\tau}_{jS}t_\alpha = \widehat{V}_{jS}^{1/2}\widehat{\tau}_{jS}t_\alpha\left(\frac{c_\tau - 1}{2}\right).$$

Therefore, $-\Delta_j\mathcal{E}_x(S) \geq A_{jS}\widehat{V}_{jS}\widehat{\tau}_{jS}^2 t_\alpha^2(\frac{c_\tau-1}{2})^2$.

Finally, prove the third statement. Note that $W_{kS} \leq W_{jS}$ implies $\widehat{V}_{kS}^{-1/2}|[\widehat{\theta}_{kS}]_k| \leq \widehat{V}_{jS}^{-1/2}|[\widehat{\theta}_{jS}]_j|$. Then

$$\begin{aligned}
&\widehat{V}_{kS}^{-1/2}\big|\big[\theta_{kS}^{*|x}\big]_k\big| - \left(\frac{1+c_\tau}{2}\right)\widehat{\tau}_{kS}t_\alpha \leq \widehat{V}_{jS}^{-1/2}\big|\big[\theta_{jS}^{*|x}\big]_k\big| + \left(\frac{1+c_\tau}{2}\right)\widehat{\tau}_{jS}t_\alpha \\
\Rightarrow\quad &\widehat{V}_{kS}^{-1/2}\big|\big[\theta_{kS}^{*|x}\big]_k\big| \leq \widehat{V}_{jS}^{-1/2}\big|\big[\theta_{jS}^{*|x}\big]_j\big| + \left(\frac{1+c_\tau}{2}\right)(\widehat{\tau}_{kS} + \widehat{\tau}_{jS})t_\alpha \\
\Rightarrow\quad &\widehat{V}_{kS}^{-1/2}A_{kS}^{-1/2}\big(-\Delta_k\mathcal{E}_x(S)\big)^{1/2} \\
&\quad\leq \widehat{V}_{jS}^{-1/2}A_{jS}^{-1/2}\big(-\Delta_j\mathcal{E}_x(S)\big)^{1/2} + \left(\frac{1+c_\tau}{2}\right)(\widehat{\tau}_{kS} + \widehat{\tau}_{jS})t_\alpha \\
&\quad= \widehat{V}_{jS}^{-1/2}A_{jS}^{-1/2}\big(-\Delta_j\mathcal{E}_x(S)\big)^{1/2} \\
&\qquad+ \left(\frac{1+c_\tau}{2}\right)(\widehat{\tau}_{kS} + \widehat{\tau}_{jS})t_\alpha\left(\frac{A_{jS}\widehat{V}_{jS}\left(\frac{1-c_\tau}{2}\right)^2\widehat{\tau}_{jS}^2 t_\alpha^2}{A_{jS}\widehat{V}_{jS}\left(\frac{1-c_\tau}{2}\right)^2\widehat{\tau}_{jS}^2 t_\alpha^2}\right)^{1/2}.
\end{aligned}$$

Using the fact that $-\Delta_j \mathcal{E}_x(S) \geq A_{jS}\widehat{V}_{jS}(\frac{1-c_\tau}{2})^2 \widehat{\tau}_{jS}^2 t_\alpha^2$ (because $T_{jS\alpha} = 1$), gives that the previous expression is bounded by

$$\leq \widehat{V}_{jS}^{-1/2} A_{jS}^{-1/2} \big(-\Delta_j \mathcal{E}_x(S)\big)^{1/2} + \frac{\left(\frac{1+c_\tau}{2}\right)(\widehat{\tau}_{kS} + \widehat{\tau}_{jS}) t_\alpha}{\left(A_{jS}\widehat{V}_{jS}\left(\frac{1-c_\tau}{2}\right)^2 \widehat{\tau}_{jS}^2 t_\alpha^2\right)^{1/2}} \big(-\Delta_j \mathcal{E}_x(S)\big)^{1/2}$$

$$= \widehat{V}_{jS}^{-1/2} A_{jS}^{-1/2} \left(1 + \frac{1+c_\tau}{c_\tau - 1} \frac{\widehat{\tau}_{kS} + \widehat{\tau}_{jS}}{\widehat{\tau}_{jS}}\right) \big(-\Delta_j \mathcal{E}_x(S)\big)^{1/2}.$$

This gives $-\Delta_k \mathcal{E}_x(S) \leq \frac{\widehat{V}_{kS} A_{kS}}{\widehat{V}_{jS} A_{jS}} (1 + \frac{1+c_\tau}{c_\tau - 1}(1 + \frac{\widehat{\tau}_{kS}}{\widehat{\tau}_{jS}}))^2 (-\Delta_j \mathcal{E}_x(S))$.

S.5.2. *Proof of Lemma 13*

Note that

$$\begin{aligned}
\widehat{Z}_{jS} &= \widehat{V}_{jS}^{-1/2}\big([\widehat{\theta}_{jS}]_j - \big[\theta_{jS}^{*|x}\big]_j\big) \\
&= \widehat{V}_{jS}^{-1/2}\big(x_j' \mathcal{Q}_S x_j\big)^{-1} x_j' \mathcal{Q}_S \big(\varepsilon - \mathrm{E}[\varepsilon|x]\big) \\
&= \Big(\big(x_j' \mathcal{Q}_S x_j\big)^{-1} \mathbb{E}_n\big[\widehat{\varepsilon}_{ijS}^2 [\mathcal{Q}_S x_{jS}]_i^2\big]\big(x_j' \mathcal{Q}_S x_j\big)^{-1}\Big)^{-1/2} \big(x_j' \mathcal{Q}_S x_j\big)^{-1} x_j' \mathcal{Q}_S \big(\varepsilon - \mathrm{E}[\varepsilon|x]\big) \\
&= \mathbb{E}_n\big[\widehat{\varepsilon}_{ijS}^2 [\mathcal{Q}_S x_{jS}]_i^2\big]^{-1/2} x_j' \mathcal{Q}_S \big(\varepsilon - \mathrm{E}[\varepsilon|x]\big) \\
&= \mathbb{E}_n\big[\widehat{\varepsilon}_{ijS}^2 \big(\eta_{jS}' x_{ijS}\big)^2\big]^{-1/2} \eta_{jS}' x_{jS} \big(\varepsilon - \mathrm{E}[\varepsilon|x]\big). \\
&= \mathbb{E}_n\big[\widehat{\varepsilon}_{ijS}^2 \big(\eta_{jS}' x_{ijS}\big)^2\big]^{-1/2} \eta_{jS}' x_{jS} \big(\varepsilon^{\mathrm{o}} + \varepsilon^{\mathrm{a}} - \mathrm{E}\big[\varepsilon^{\mathrm{a}}|x\big]\big).
\end{aligned}$$

Let $\ddot{\varepsilon} = \varepsilon^{\mathrm{o}} + \varepsilon^{\mathrm{a}} - \mathrm{E}[\varepsilon^{\mathrm{a}}|x]$. Define the *Regularization Event* by

$$\mathcal{R} = \left\{ \frac{\left|\sum_{i=1}^n x_{ik}\ddot{\varepsilon}_i\right|}{\sqrt{\sum_{i=1}^n x_{ik}^2 \ddot{\varepsilon}_i^2}} \leq t_\alpha \text{ for every } k \leq p \right\}.$$

In addition, define the *Variability Domination Event* by

$$\mathcal{V} = \left\{ \sum_{i=1}^n x_{ik}^2 \ddot{\varepsilon}_i^2 \leq \left(\frac{1+c_\tau}{2}\right)^2 \sum_{i=1}^n x_{ik}^2 \widehat{\varepsilon}_{ijS}^2 \text{ for every } k \in jS, \text{for every } |S| < K_n \right\}.$$

The definitions of the Regularization Event and the Variability Domination Event are useful because

$$\mathcal{R} \cap \mathcal{V} \Rightarrow \mathcal{A}.$$

To see this, note that on $\mathcal{R}$, the following inequality holds for any conformable vector ν:

$$\left(\sum_{i=1}^{n}\sum_{k\in jS}\nu_k x_{ik}\ddot{\varepsilon}_i\right)^2 \leq \left(t_\alpha \sum_{k\in jS}|\nu_k|\sqrt{\sum_{i=1}^{n}x_{ik}^2\ddot{\varepsilon}_i^2}\right)^2.$$

Furthermore, on $\mathcal{V}$, the previous expression can be further bounded by

$$\begin{aligned}
&\leq \left(\frac{1+c_\tau}{2}\right)^2\left(t_\alpha\sum_{k\in jS}|\nu_k|\sqrt{\sum_{i=1}^{n}x_{ik}^2\widehat{\varepsilon}_{ijS}^2}\right)^2 \\
&= \left(\frac{1+c_\tau}{2}\right)^2 \frac{\left(t_\alpha\sum_{k\in jS}|\nu_k|\sqrt{\sum_{i=1}^{n}x_{ik}^2\widehat{\varepsilon}_{ijS}^2}\right)^2}{\sum_{i=1}^{n}\left(\sum_{k\in jS}\nu_k x_{ik}\right)^2\widehat{\varepsilon}_{ijS}^2}\sum_{i=1}^{n}\left(\sum_{k\in jS}\nu_k x_{ik}\right)^2\widehat{\varepsilon}_{ijS}^2 \\
&= \left(\frac{1+c_\tau}{2}\right)^2 t_\alpha^2 \frac{\left\|\nu' \operatorname{Diag}\left(\Psi_{jS}^{\widehat{\varepsilon}}\right)^{1/2}\right\|_1^2}{\nu'\Psi_{jS}^{\widehat{\varepsilon}}\nu}\sum_{i=1}^{n}\left(\sum_{k\in jS}\nu_k x_{ik}\right)^2\widehat{\varepsilon}_{ijS}^2.
\end{aligned}$$

Specializing to the case that $\nu = \eta_{jS}$ and using $\widehat{\tau}_{jS} = \frac{\|\nu' \operatorname{Diag}(\Psi_{jS}^{\widehat{\varepsilon}})^{1/2}\|_1}{\sqrt{\nu'\Psi_{jS}^{\widehat{\varepsilon}}\nu}}$ gives that

$$|\widehat{Z}_{jS}| \leq \frac{1+c_\tau}{2}\widehat{\tau}_{jS}t_\alpha \quad \text{on } \mathcal{R}\cap\mathcal{V}.$$

It is therefore sufficient to prove that $\mathcal{R}$ and $\mathcal{V}$ have probability $\to 1$ under $\mathrm{P}^{\mathcal{X}}$. $\mathrm{P}^{\mathcal{X}}(\mathcal{R}) \to 1$ follows immediately from the moderate deviation bounds for self-normalized sums given in Jing, Shao, and Wang (2003). For details on the application of this result, see Belloni, Chen, Chernozhukov, and Hansen (2012).

Therefore, it is only left to show that $\mathrm{P}^{\mathcal{X}}(\mathcal{V}) \to 1$. Define $\varepsilon_{ijS} = y_i - x_{ijS}'\theta_{jS}^{*|x}$. Furthermore, define ξ_{ijS} through the decomposition $\varepsilon_{ijS} = \ddot{\varepsilon}_i + \xi_{ijS}$. Let ε_{jS} and ξ_{jS} be the respective stacked versions. Let $\tilde{c}_\tau = ((1+c_\tau)/2)^2$. Then

$$\begin{aligned}
\tilde{c}_\tau\sum_{i=1}^{n}x_{ik}^2\widehat{\varepsilon}_{ijS}^2 &= \tilde{c}_\tau\left[\sum_{i=1}^{n}x_{ik}^2\left(\widehat{\varepsilon}_{ijS}^2 - \varepsilon_{ijS}^2\right) + \sum_{i=1}^{n}x_{ik}^2\ddot{\varepsilon}_i^2 + 2\sum_{i=1}^{n}x_{ik}^2\ddot{\varepsilon}_i\xi_{ijS} + \sum_{i=1}^{n}x_{ik}^2\xi_{ijS}^2\right] \\
&\geq \tilde{c}_\tau\left[\sum_{i=1}^{n}x_{ik}^2\left(\widehat{\varepsilon}_{ijS}^2 - \varepsilon_{ijS}^2\right) + \sum_{i=1}^{n}x_{ik}^2\ddot{\varepsilon}_i^2 + 2\sum_{i=1}^{n}x_{ik}^2\ddot{\varepsilon}_i\xi_{ijS}\right] \\
&= \sum_{i=1}^{n}x_{ik}^2\ddot{\varepsilon}_i^2 + \tilde{c}_\tau\sum_{i=1}^{n}x_{ik}^2\left(\widehat{\varepsilon}_{ijS}^2 - \varepsilon_{ijS}^2\right) + \frac{(\tilde{c}_\tau - 1)}{2}\sum_{i=1}^{n}x_{ik}^2\ddot{\varepsilon}_i^2 \\
&\quad + 2\tilde{c}_\tau\sum_{i=1}^{n}x_{ik}^2\ddot{\varepsilon}_i\xi_{ijS} + \frac{(\tilde{c}_\tau - 1)}{2}\sum_{i=1}^{n}x_{ik}^2\ddot{\varepsilon}_i^2.
\end{aligned}$$

Define the two events

$$\mathcal{V}' = \left\{ \tilde{c}_\tau \mathbb{E}_n\left[x_{ik}^2\left(\widehat{\varepsilon}_{ijS}^2 - \varepsilon_{ijS}^2\right)\right] + \frac{(\tilde{c}_\tau - 1)}{2}\mathbb{E}_n\left[x_{ik}^2 \ddot{\varepsilon}_i^2\right] \geq 0 \text{ for all } j, k \leq p, |S| < K_n \right\},$$

$$\mathcal{V}'' = \left\{ 2\tilde{c}_\tau \mathbb{E}_n\left[x_{ik}^2 \ddot{\varepsilon}_i \xi_{ijS}\right] + \frac{(\tilde{c}_\tau - 1)}{2}\mathbb{E}_n\left[x_{ik}^2 \ddot{\varepsilon}_i^2\right] \geq 0 \text{ for all } j, k \leq p, |S| < K_n \right\}.$$

Therefore, $\mathcal{V}' \cap \mathcal{V}'' \Rightarrow \mathcal{V}$.

Note that $\mathbb{E}_n[x_{ik}^2 \ddot{\varepsilon}_i^2] \geq \frac{1}{2}\mathbb{E}_n[x_{ik}^2 \varepsilon_i^2] - \mathbb{E}_n[x_{ik}^2 \mathrm{E}[\varepsilon_i^{\mathrm{a}}|x]] \geq \frac{1}{2}\mathbb{E}_n[x_{ik}^2 \varepsilon_i^2] - \max_{i \leq n} \mathrm{E}[\varepsilon_i^{\mathrm{a}2}|x]^{1/2} \times \mathbb{E}_n[x_{ik}^4]^{1/2}$. This is bounded below with $\mathrm{P}^{\mathcal{X}} \to 1$ by a positive constant independent of n. Therefore, to show that $\mathrm{P}^{\mathcal{X}}(\mathcal{V}') \to 1$, $\mathrm{P}^{\mathcal{X}}(\mathcal{V}'') \to 1$, it suffices to show $\mathbb{E}_n[x_{ik}^2(\widehat{\varepsilon}_{ijS}^2 - \varepsilon_{ijS}^2)]$ and $\mathbb{E}_n[x_{ik}^2 \ddot{\varepsilon}_i \xi_{ijS}]$, respectively, are suitably smaller order.

First consider $\mathbb{E}_n[x_{ik}^2(\widehat{\varepsilon}_{ijS}^2 - \varepsilon_{ijS}^2)]$. It is convenient to bound the slightly more general sum $\mathbb{E}_n[x_{ik}x_{il}(\widehat{\varepsilon}_{ijS}^2 - \varepsilon_{ijS}^2)]$, because this will show up again:

$$\begin{aligned}
&\mathbb{E}_n\left[x_{ik}x_{il}\left(\widehat{\varepsilon}_{ijS}^2 - \varepsilon_{ijS}^2\right)\right] \\
&\quad = 2\mathbb{E}_n\left[x_{ik}x_{il}\varepsilon_{ijS}x_{ijS}'\left(\theta_{jS}^{*|x} - \widehat{\theta}_{jS}\right)\right] + \mathbb{E}_n\left[x_{ik}x_{il}\left(x_{ijS}'\left(\theta_{jS}^{*|x} - \widehat{\theta}_{jS}\right)\right)^2\right] \\
&\quad \leq 2\left\|\mathbb{E}_n\left[x_{ik}x_{il}\varepsilon_{ijS}x_{ijS}'\right]\right\|_2 \left\|\theta_{jS}^{*|x} - \widehat{\theta}_{jS}\right\|_2 + \lambda_{\max}\mathbb{E}_n\left[x_{ik}x_{il}x_{ijS}x_{ijS}'\right]\left\|\theta_{jS}^{*|x} - \widehat{\theta}_{jS}\right\|_2^2.
\end{aligned}$$

Standard reasoning gives that $\|\theta_{jS}^{*|x} - \widehat{\theta}_{jS}\|_2 \leq \varphi_{\min}(K_n)(G)^{-1/2}\sqrt{K_n}\|\mathbb{E}_n x_{ijS}\varepsilon_{ijS}\|_\infty$. Therefore, the bound continues:

$$\begin{aligned}
&\leq 2\left\|\mathbb{E}_n\left[x_{ik}x_{il}\varepsilon_{ijS}x_{ijS}'\right]\right\|_2 \varphi_{\min}(K_n)(G)^{-1/2}\sqrt{K_n}\|\mathbb{E}_n x_{ijS}\varepsilon_{ijS}\|_\infty \\
&\quad + \lambda_{\max}\mathbb{E}_n\left[x_{ik}x_{il}x_{ijS}x_{ijS}'\right]\varphi_{\min}(K_n)(G)^{-1}K_n\|\mathbb{E}_n x_{ijS}\varepsilon_{ijS}\|_\infty^2.
\end{aligned}$$

Note that $\lambda_{\max}\mathbb{E}_n[x_{ik}x_{il}x_{ijS}x_{ijS}'] \leq K_n \max_{j \leq p}\mathbb{E}_n[x_{ij}^4]$:

$$\begin{aligned}
&\leq 2\left\|\mathbb{E}_n\left[x_{ik}x_{il}\varepsilon_{ijS}x_{ijS}'\right]\right\|_2 \varphi_{\min}(K_n)(G)^{-1/2}\sqrt{K_n}\|\mathbb{E}_n x_{ijS}\varepsilon_{ijS}\|_\infty \\
&\quad + K_n^2 \max_{j \leq p}\mathbb{E}_n\left[x_{ij}^4\right]\varphi_{\min}(K_n)(G)^{-1}\|\mathbb{E}_n x_{ijS}\varepsilon_{ijS}\|_\infty^2.
\end{aligned}$$

An application of Cauchy–Schwarz to the top line gives

$$\begin{aligned}
&\leq 2\sqrt{K_n}\max_j \mathbb{E}_n\left[x_{ik}^4\right]^{1/2}\max_{j,S}\mathbb{E}_n\left[\varepsilon_{ijS}^2 x_{ij}^2\right]^{1/2}\varphi_{\min}(K_n)(G)^{-1/2}\sqrt{K_n}\|\mathbb{E}_n x_{ijS}\varepsilon_{ijS}\|_\infty \\
&\quad + K_n^2 \max_{j \leq p}\mathbb{E}_n\left[x_{ij}^4\right]\varphi_{\min}(K_n)(G)^{-1}\|\mathbb{E}_n x_{ijS}\varepsilon_{ijS}\|_\infty^2.
\end{aligned}$$

Next, $\|\mathbb{E}_n x_{ijS}\varepsilon_{ijS}\|_\infty$ and $\mathbb{E}_n[\varepsilon_{ijS}^2 x_{ij}^2]^{1/2}$ are bounded using $\varepsilon_{ijS} = \varepsilon_i - \mathrm{E}[\varepsilon_i|x] + \xi_{ijS}$. Note that by construction, $\|\mathbb{E}_n[x_{ijS}\xi_{ijS}]\|_\infty = 0$. Then

$$\begin{aligned}
\left\|\mathbb{E}_n[x_{ijS}\varepsilon_{ijS}]\right\|_\infty &\leq \left\|\mathbb{E}_n[x_i\varepsilon_i]\right\|_\infty + \left\|\mathbb{E}_n\left[x_i \mathrm{E}\left[\varepsilon_i^{\mathrm{a}}|x\right]\right]\right\|_\infty \\
&\leq \left\|\mathbb{E}_n[x_i\varepsilon_i]\right\|_\infty + \max_{j \leq p}\mathbb{E}_n\left[x_{ij}^2\right]^{1/2}\mathbb{E}_n\left[\mathrm{E}\left[\varepsilon_i^{\mathrm{a}}|x\right]^2\right]^{1/2} = O(\sqrt{\log p/n})
\end{aligned}$$

with $\mathrm{P}^{\mathcal{X}} \to 1$. Next,

$$\begin{aligned}\mathbb{E}_n\big[\varepsilon_{ijS}^2 x_{ij}^2\big] &\le 3\mathbb{E}_n\big[\varepsilon_i^2 x_{ij}^2\big] + 3\mathbb{E}_n\big[\mathrm{E}\big[\varepsilon_i^{\mathrm{a}2}|x\big]x_{ij}^2\big] + 3\mathbb{E}_n\big[\xi_{ijS}^2 x_{ij}^2\big] \\ &\le 3\mathbb{E}_n\big[\varepsilon_i^2 x_{ij}^2\big] + 3\mathbb{E}_n\big[x_{ij}^2\big]\max_{i\le n}\mathrm{E}\big[\varepsilon_i^{\mathrm{a}2}|x\big] + 3\mathbb{E}_n\big[\xi_{ijS}^4\big]^{1/2}\mathbb{E}_n\big[x_{ij}^4\big]^{1/2}.\end{aligned}$$

Next, $(\mathbb{E}_n[\xi_{ijS}^4])^{1/2} \le O(1)s_0^2$ on $\mathcal{X}_1 \cap \mathcal{X}_3$. To see this, note $\xi_{jS} = \mathcal{Q}_{jS}x\theta_0 = \sum_{l=1}^{s_0}\mathcal{Q}_{jS}x_l\theta_{0,l} = \sum_{l=1}^{s_0}\eta_{l,(jS)}x_{ljS} = \tilde{\eta}_{jS}x_{S_0\cup jS}$ for some new linear combination $\tilde{\eta}_{jS}$. Note that $\|\tilde{\eta}_{jS}\|_1 \le s_0 O(1)$. Then $(\mathbb{E}_n[\xi_{ijS}^4])^{1/4} \le \|\tilde{\eta}_{jS}\|_1 \max_{k\le p}\mathbb{E}_n[x_{ik}^4]^{1/4}$ from which the bound follows.

Next consider $\mathbb{E}_n[x_{ik}^2\ddot{\varepsilon}_i\xi_{ijS}]$. Consider two cases. In Case 1,

$$\mathbb{E}_n\big[x_{ik}^4\xi_{ijS}^2\big]^{1/2} \le \frac{1}{2\tilde{c}_\tau}\frac{(\tilde{c}_\tau - 1)}{2}\frac{\mathbb{E}_n\big[x_{ik}^2\ddot{\varepsilon}_i^2\big]}{\mathbb{E}_n\big[\ddot{\varepsilon}_i^2\big]^{1/2}}.$$

In this case, $2\tilde{c}_\tau\mathbb{E}_n[x_{ik}^2\ddot{\varepsilon}_i\xi_{ijS}] \le \mathbb{E}_n[x_{ik}^4\xi_{ijS}^2]^{1/2}\mathbb{E}_n[\ddot{\varepsilon}_i^2]^{1/2} \le \frac{\tilde{c}_\tau - 1}{2}$, and the requirement of $\mathcal{V}''$ for k, j, S holds.

For Case 2, suppose the alternative that $\mathbb{E}_n[x_{ik}^4 x_{ijS}^2] > \frac{1}{2\tilde{c}_\tau}\frac{(\tilde{c}_\tau-1)}{2}\frac{\mathbb{E}_n[x_{ik}^2\ddot{\varepsilon}_i^2]}{\mathbb{E}_n[\ddot{\varepsilon}_i^2]^{1/2}}$ holds. Then $\mathrm{E}[\mathbb{E}_n[x_{ik}^4\xi_{ijS}^2\ddot{\varepsilon}_i^2]|x]$ is bounded away from zero by conditions on $\mathrm{E}[\varepsilon_i^2|x]$ and $\max_i|\varepsilon_i^{\mathrm{a}}|$. In addition, $\mathrm{E}[\mathbb{E}_n[|x_{ik}|^6|\xi_{ijS}|^3|\ddot{\varepsilon}_i|^3]|x] \le \max_i \mathrm{E}[|\ddot{\varepsilon}_i|^3|x]\mathbb{E}_n[|x_{ik}|^6|\xi_{ijS}|^3] \le O(1)\mathbb{E}_n[|x_{ik}|^6|\xi_{ijS}|^3]$. This term is further bounded by $O(1)\mathbb{E}_n[x_{ik}^{12}]^{1/2}\mathbb{E}_n[|\xi_{ijS}|^6]^{1/2}$. Using the same reasoning as bounding $\mathbb{E}_n[\xi_{ijS}^4]$ earlier, it follows that $\mathbb{E}_n[|\xi_{ijS}|^6]^{1/2} = O(1)s_0^3$. In addition, $\mathbb{E}_n[x_{ik}^{12}] = O(1)$. As a result, for those k, j, S which fall in Case 2, the self-normalized sum

$$= \max_{j,k,S\in\text{Case 2}}\frac{\sqrt{n}\big|\mathbb{E}_n\big[x_{ik}^2\xi_{ijS}\ddot{\varepsilon}_i\big]\big|}{\sqrt{\mathbb{E}_n\big[x_{ik}^4\xi_{ijS}^2\ddot{\varepsilon}_i^2\big]}}$$

is $O(\log(p^{K_n}))$ with probability $1 - o(1)$ provided $\sqrt{\log(p^{K_n})} = o(n^{1/6}/(s_0^3)^{1/3})$. This holds under the assumed rate conditions. Then $\max_{j,k,S}|\mathbb{E}_n[x_{ik}^2\xi_{ijS}\ddot{\varepsilon}_i]|$ is bounded by $\frac{1}{\sqrt{n}}O(\log(p^{K_n}))\max_{j,k,S}\sqrt{\mathbb{E}_n[x_{ik}^4\xi_{ijS}^2\ddot{\varepsilon}_i^2]}$. Furthermore, $\mathbb{E}_n[x_{ik}^4\xi_{ijS}^2\ddot{\varepsilon}_i^2] \le \mathbb{E}_n[x_{ik}^8\xi_{ijS}^4]^{1/2}\mathbb{E}_n[\ddot{\varepsilon}_i^4]^{1/2} \le (\mathbb{E}_n[x_{ik}^{12}]^{2/3}\mathbb{E}_n[\xi_{ijS}^{12}]^{1/3})^{1/2}\mathbb{E}_n[\ddot{\varepsilon}_i^4]^{1/2} \le O(1)s_0^2\mathbb{E}_n[\ddot{\varepsilon}_i^4]^{1/2}$. Note that $\mathbb{E}_n[\ddot{\varepsilon}_i^4]^{1/2} \le O(1)$ with $\mathrm{P}^{\mathcal{X}} \to 1$. Together, these give that $\max_{j,k,S}\mathbb{E}_n[x_{ik}^2\ddot{\varepsilon}_i\xi_{ijS}] = o(1)$ with $\mathrm{P}^{\mathcal{X}} \to 1$. Finally, $\mathrm{P}^{\mathcal{X}}(\mathcal{V}) \to 1$.

S.5.3. *Proof of Lemma 14*

First, A_{jS} depend only on x and are bounded above and below by constants which do not depend on n on $\mathcal{X}$ from the assumption on the sparse eigenvalues of G. For bounding $\widehat{\tau}_{jS}$ above and away from zero, since $1 \le \|\eta_{jS}\|_1, \|\eta_{jS}\|_2 \le O(1)$ on $\mathcal{X}$, it is sufficient to show that the eigenvalues of $\Psi_{jS}^{\widehat{\varepsilon}} = \mathbb{E}_n[x_{ijS}x_{ijS}'\widehat{\varepsilon}_{ijS}^2]$ remain bounded above and away from zero and that the diagonal terms of $\Psi_{jS}^{\widehat{\varepsilon}}$ remain bounded above and away from zero. Note that by arguments in the last section, it was shown that $\mathbb{E}_n[x_{ik}x_{il}(\widehat{\varepsilon}_{ijS} - \varepsilon_{ijS})] = O(\sqrt{\log p/n})$ with $\mathrm{P}^{\mathcal{X}} \to 1$. Therefore, $\|\mathbb{E}_n[x_{ijS}x_{ijS}'\widehat{\varepsilon}_{ijS}^2] - \mathbb{E}_n[x_{ijS}x_{ijS}'\varepsilon_{ijS}^2]\|_{\mathcal{F}} = O(K_n\sqrt{\log p/n})$ with $\mathrm{P}^{\mathcal{X}} \to 1$. Here, $\mathcal{F}$ is the Frobenius norm. By the assumed rate condition, the above quantity therefore vanishes with $\mathrm{P}^{\mathcal{X}} \to 1$.

Next,

$$\mathbb{E}_n\big[x_{ijS}x'_{ijS}\varepsilon^2_{ijS}\big] = \mathbb{E}_n\big[x_{ijS}x'_{ijS}\varepsilon^2_i\big] + 2\mathbb{E}_n\big[x_{ijS}x'_{ijS}\varepsilon_i\big(\xi_{ijS} + \mathrm{E}\big[\varepsilon^{\mathrm{a}}_i|x\big]\big)\big]$$
$$+ \mathbb{E}_n\big[x_{ijS}x'_{ijS}\big(\xi_{ijS} + \mathrm{E}\big[\varepsilon^{\mathrm{a}}_i|x\big]\big)^2\big].$$

The first term above, $\mathbb{E}_n[x_{ijS}x'_{ijS}\varepsilon^2_i]$, has eigenvalues bounded away from zero for all j, S with $\mathrm{P}^{\mathcal{X}} \to 1$. The third term above, $\mathbb{E}_n[x_{ijS}x'_{ijS}(\xi_{ijS} + \mathrm{E}[\varepsilon^{\mathrm{a}}_i|x])^2]$, is positive semidefinite by construction. The second term above has Frobenius norm tending to zero for all j, S with $\mathrm{P}^{\mathcal{X}} \to 1$. This, in conjunction with the fact that the eigenvalues of $\mathbb{E}_n[x_{ijS}x'_{ijS}\widehat{\varepsilon}_{ijS}]$ are bounded above and away from zero with $\mathrm{P}^{\mathcal{X}} \to 1$, shows that the eigenvalues of $\Psi^{\widehat{\varepsilon}}_{jS} = \mathbb{E}_n[x_{ijS}x'_{ijS}\widehat{\varepsilon}^2_{ijS}]$ are bounded above and away from zero with $\mathrm{P}^{\mathcal{X}} \to 1$. Finally, for bounding $\widehat{V}_{jS}$, it is sufficient to show that $\max_{k\le p}\mathbb{E}_n[\varepsilon^2_i(\eta'_{jS}x_{ijS})^2]$ be bounded above. This follows immediately from $\mathrm{E}[\varepsilon^4_i|x]$ being uniformly bounded and $\max_{j,S}\|\eta_{jS}\|_1 = O(1)$ and $\max_{k\le p}\mathbb{E}_n[x^4_{ik}] = O(1)$. These imply that $\mathrm{P}^{\mathcal{X}}(\mathcal{B}) \to 1$.

REFERENCES

BELLONI, A., D. CHEN, V. CHERNOZHUKOV, AND C. HANSEN (2012): "Sparse Models and Methods for Optimal Instruments With an Application to Eminent Domain," *Econometrica*, 80, 2369–2429. [15]

JING, B.-Y., Q.-M. SHAO, AND Q. WANG (2003): "Self-Normalized Cramèr-Type Large Deviations for Independent Random Variables," *Annals of Probability*, 31 (4), 2167–2215. [15]

Co-editor Ulrich K. Müller handled this manuscript.

Manuscript received 18 April, 2018; final version accepted 16 March, 2020; available online 9 April, 2020.