The Broken Adaptive Ridge Procedure and Its Applications

Linlin Dai, Kani Chen and Gang Li

Southwestern University of Finance and Economics,

Hong Kong University of Science and Technology and

University of California, Los Angeles

Supplementary Material

S1 Theorem Proofs

We first present some preliminaries. Let $\mathbf{\Sigma}_{n1} = \mathbf{T}_1'\mathbf{\Sigma}_n\mathbf{T}_1$ and $\mathbf{\Sigma}_{n2} = \mathbf{T}_2'\mathbf{\Sigma}_n\mathbf{T}_2$. It follows from (2.1) that $\|\mathbf{g}(\tilde{\boldsymbol{\beta}})\| = O_p(\|\hat{\boldsymbol{\beta}}(\text{OLS})\|)$. Multiplying both sides of equation (2.2) by $(\mathbf{X}'\mathbf{X})^{-1}\{\mathbf{X}'\mathbf{X} + \lambda_n\mathbf{D}(\boldsymbol{\beta})\}$ yields

$$\mathbf{g}(\boldsymbol{\beta}) + \lambda_n(\mathbf{X}'\mathbf{X})^{-1}\mathbf{D}(\boldsymbol{\beta})\mathbf{g}(\boldsymbol{\beta}) = \boldsymbol{\beta}_0 + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}. \tag{S1.1}$$

Then, transform (S1.1) by $\mathbf{T}'$ and we have

$$\mathbf{T}'\{\mathbf{g}(\boldsymbol{\beta}) - \boldsymbol{\beta}_0\} + \frac{\lambda_n}{n}\mathbf{T}'\mathbf{\Sigma}_n^{-1}\mathbf{D}(\boldsymbol{\beta})\mathbf{g}(\boldsymbol{\beta}) = \mathbf{T}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon},$$

which is equivalent to

$$\mathbf{T}_1'\{\mathbf{g}(\boldsymbol{\beta}) - \boldsymbol{\beta}_0\} + \frac{\lambda_n}{n}\mathbf{T}_1'\mathbf{\Sigma}_n^{-1}\mathbf{D}(\boldsymbol{\beta})\mathbf{g}(\boldsymbol{\beta}) = \mathbf{T}_1'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}, \tag{S1.2}$$

$$\mathbf{T}_2'\{\mathbf{g}(\boldsymbol{\beta}) - \boldsymbol{\beta}_0\} + \frac{\lambda_n}{n}\mathbf{T}_2'\boldsymbol{\Sigma}_n^{-1}\mathbf{D}(\boldsymbol{\beta})\mathbf{g}(\boldsymbol{\beta}) = \mathbf{T}_2'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}. \tag{S1.3}$$

Note that $\mathbf{T}_2'\boldsymbol{\beta}_0 = 0$. The equality (S1.3) can be written as

$$\begin{aligned}\mathbf{T}_2'\mathbf{g}(\boldsymbol{\beta}) + \frac{\lambda_n}{n}\mathbf{T}_2'\boldsymbol{\Sigma}_n^{-1}\mathbf{D}_1(\boldsymbol{\beta})\mathbf{g}(\boldsymbol{\beta}) + \frac{\lambda_n}{n}\mathbf{T}_2'\boldsymbol{\Sigma}_n^{-1}\mathbf{D}_2(\boldsymbol{\beta})\mathbf{g}(\boldsymbol{\beta})& \\ = \mathbf{T}_2'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon},&\end{aligned} \tag{S1.4}$$

where $\mathbf{D}_1(\boldsymbol{\beta}) = \sum_{k=1}^{q_n}\mathbf{d}_k\mathbf{d}_k'/c_k^2(\boldsymbol{\beta})$ and $\mathbf{D}_2(\boldsymbol{\beta}) = \sum_{k=q_n+1}^{K_n}\mathbf{d}_k\mathbf{d}_k'/c_k^2(\boldsymbol{\beta})$. Furthermore, let $\boldsymbol{\Sigma}_{n2}^* = \mathbf{T}_2'\boldsymbol{\Sigma}_n^{-1}\mathbf{T}_2$. Since $\mathbf{d}_k'\mathbf{T}_1 = \mathbf{0}$ for $k = q_n+1, \ldots, K_n$, equation (S1.4) equals

$$\begin{aligned}\mathbf{T}_2'\mathbf{g}(\boldsymbol{\beta}) + \frac{\lambda_n}{n}\mathbf{T}_2'\boldsymbol{\Sigma}_n^{-1}\mathbf{D}_1(\boldsymbol{\beta})\mathbf{g}(\boldsymbol{\beta}) + \frac{\lambda_n}{n}\boldsymbol{\Sigma}_{n2}^*\mathbf{T}_2'\mathbf{D}_2(\boldsymbol{\beta})\mathbf{g}(\boldsymbol{\beta})& \\ = \mathbf{T}_2'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}.&\end{aligned} \tag{S1.5}$$

S1.1 Proof of Lemma 1

Proof. It follows from assumption (A1) that

$$\begin{aligned}E(\|\mathbf{T}_2'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}\|^2) &= E[\mathrm{tr}\{\boldsymbol{\varepsilon}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{T}_2\mathbf{T}_2'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}\}] \\ &= \mathrm{tr}\{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{T}_2\mathbf{T}_2'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')\mathbf{X}\} \\ &= \frac{\sigma^2}{n}\mathrm{tr}\{\mathbf{T}_2'\boldsymbol{\Sigma}_n^{-1}\mathbf{T}_2\} \\ &= O\Big(\frac{p_n}{n}\Big).\end{aligned}$$

Recall that $\mathfrak{B} \equiv \{\boldsymbol{\beta} \in \mathbb{R}^{p_n} : \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| \leq \delta_n \sqrt{p_n/n}\}$. According to assumptions (A2)–(A3), we have

$$\begin{aligned}\sup_{\boldsymbol{\beta}\in\mathfrak{B}} \left\| \frac{\lambda_n}{n} \mathbf{T}_2' \boldsymbol{\Sigma}_n^{-1} \mathbf{D}_1(\boldsymbol{\beta}) \mathbf{g}(\boldsymbol{\beta}) \right\| &\leq \frac{\lambda_n}{n} \|\boldsymbol{\Sigma}_n^{-1}\| \sup_{\boldsymbol{\beta}\in\mathfrak{B}} \|\mathbf{D}_1(\boldsymbol{\beta})\mathbf{g}(\boldsymbol{\beta})\| \\ &= O_p\Big(\frac{\lambda_n q_n \sqrt{p_n}}{n b_n^2}\Big) = o_p\Big(\sqrt{\frac{p_n}{n}}\Big).\end{aligned}$$

Therefore, (S1.5) equals

$$\sup_{\boldsymbol{\beta}\in\mathfrak{B}} \left\| \mathbf{T}_2'\mathbf{g}(\boldsymbol{\beta}) + \frac{\lambda_n}{n} \boldsymbol{\Sigma}_{n2}^{*} \mathbf{T}_2' \mathbf{D}_2(\boldsymbol{\beta}) \mathbf{g}(\boldsymbol{\beta}) \right\| = O_p\Big(\sqrt{\frac{p_n}{n}}\Big). \qquad \text{(S1.6)}$$

Since $\mathbf{d}_k'\mathbf{T}_1 = \mathbf{0}$, we have

$$\begin{aligned}\mathbf{D}_2(\boldsymbol{\beta})\mathbf{g}(\boldsymbol{\beta}) &= \mathbf{D}_2(\boldsymbol{\beta})\mathbf{T}\mathbf{T}'\mathbf{g}(\boldsymbol{\beta}) \\ &= \sum_{k=q_n+1}^{K_n} \frac{\mathbf{d}_k \mathbf{d}_k'}{c_k^2(\boldsymbol{\beta})} (\mathbf{T}_1 \vdots \mathbf{T}_2) \begin{pmatrix} \mathbf{T}_1' \\ \mathbf{T}_2' \end{pmatrix} \mathbf{g}(\boldsymbol{\beta}) \\ &= \{\mathbf{0} \vdots \mathbf{D}_2(\boldsymbol{\beta})\mathbf{T}_2\} \begin{pmatrix} \mathbf{T}_1' \\ \mathbf{T}_2' \end{pmatrix} \mathbf{g}(\boldsymbol{\beta}) \\ &= \mathbf{D}_2(\boldsymbol{\beta})\mathbf{T}_2\mathbf{T}_2'\mathbf{g}(\boldsymbol{\beta}).\end{aligned}$$

Set $\boldsymbol{\gamma}^*(\boldsymbol{\beta}) = \mathbf{T}_2'\mathbf{g}(\boldsymbol{\beta})$ and $\tilde{\mathbf{D}}_2(\boldsymbol{\beta}) = \mathbf{T}_2'\mathbf{D}_2(\boldsymbol{\beta})\mathbf{T}_2$. Then, by multiplying both sides of equation (S1.6) with $\boldsymbol{\gamma}^*(\boldsymbol{\beta})'\boldsymbol{\Sigma}_{n2}^{*-1}/\|\boldsymbol{\gamma}^*(\boldsymbol{\beta})\|$, we obtain

$$\sup_{\boldsymbol{\beta}\in\mathfrak{B}} \left\{ \frac{\boldsymbol{\gamma}^*(\boldsymbol{\beta})'\boldsymbol{\Sigma}_{n2}^{*-1}\boldsymbol{\gamma}^*(\boldsymbol{\beta})}{\|\boldsymbol{\gamma}^*(\boldsymbol{\beta})\|} + \frac{\lambda_n}{n} \frac{\boldsymbol{\gamma}^*(\boldsymbol{\beta})'\tilde{\mathbf{D}}_2(\boldsymbol{\beta})\boldsymbol{\gamma}^*(\boldsymbol{\beta})}{\|\boldsymbol{\gamma}^*(\boldsymbol{\beta})\|} \right\} = O_p\Big(\sqrt{\frac{p_n}{n}}\Big). \quad \text{(S1.7)}$$

Note that here we are assuming that $\|\boldsymbol{\gamma}^*(\boldsymbol{\beta})\| \neq 0$. Observe that both terms inside the supremum in equation (S1.7) are nonnegative. Therefore,

$$\frac{\lambda_n}{n} \sup_{\boldsymbol{\beta}\in\mathfrak{B}} \frac{\boldsymbol{\gamma}^*(\boldsymbol{\beta})'\tilde{\mathbf{D}}_2(\boldsymbol{\beta})\boldsymbol{\gamma}^*(\boldsymbol{\beta})}{\|\boldsymbol{\gamma}^*(\boldsymbol{\beta})\|} = O_p\Big(\sqrt{\frac{p_n}{n}}\Big). \tag{S1.8}$$

Since

$$\tilde{\mathbf{D}}_2(\boldsymbol{\beta}) = \sum_{k=q_n+1}^{K_n} \frac{\mathbf{T}_2'\mathbf{d}_k\mathbf{d}_k'\mathbf{T}_2}{c_k^2(\boldsymbol{\beta})} = \sum_{k=q_n+1}^{K_n} \frac{\mathbf{T}_2'\mathbf{d}_k\mathbf{d}_k'\mathbf{T}_2}{(\mathbf{d}_k'\mathbf{T}_2\mathbf{T}_2'\boldsymbol{\beta})^2} = \sum_{k=q_n+1}^{K_n} \frac{\mathbf{d}_k^*\mathbf{d}_k^{*\prime}}{\{\mathbf{d}_k^{*\prime}\boldsymbol{\gamma}(\boldsymbol{\beta})\}^2},$$

where $\mathbf{d}_k^* = \mathbf{T}_2'\mathbf{d}_k$ and $\boldsymbol{\gamma}(\boldsymbol{\beta}) = \mathbf{T}_2'\boldsymbol{\beta}$, it follows from (S1.8) that

$$\frac{\lambda_n}{n} \sup_{\substack{q_n+1\leq k\leq K_n,\\ \boldsymbol{\beta}\in\mathfrak{B}}} \frac{\{\mathbf{d}_k^{*\prime}\boldsymbol{\gamma}^*(\boldsymbol{\beta})\}^2}{\{\mathbf{d}_k^{*\prime}\boldsymbol{\gamma}(\boldsymbol{\beta})\}^2\|\boldsymbol{\gamma}^*(\boldsymbol{\beta})\|} = O_p\Big(\sqrt{\frac{p_n}{n}}\Big).$$

On the other hand, since $\mathfrak{D}$ is a linear space spanned by $\mathbf{d}_{q_n+1}, \ldots, \mathbf{d}_{K_n}$ with orthonormal basis $\mathbf{T}_2$, for any unit vector $\mathbf{a}$ in $\mathfrak{D}$, there exist some $\tilde{\mathbf{d}}_j^* \in \{\mathbf{d}_k^*, q_n+1 \leq k \leq K_n\}$ such that $|\tilde{\mathbf{d}}_j^{*\prime}\mathbf{a}| > c_3$, for some constant $c_3 > 0$. Let $\tilde{\mathbf{d}}_j^*$ be such that $|\tilde{\mathbf{d}}_j^{*\prime}\boldsymbol{\gamma}^*(\boldsymbol{\beta})| > c_3\|\boldsymbol{\gamma}^*(\boldsymbol{\beta})\|$. Note that $|\tilde{\mathbf{d}}_j^{*\prime}\boldsymbol{\gamma}(\boldsymbol{\beta})| \leq \|\tilde{\mathbf{d}}_j^{*\prime}\|\|\boldsymbol{\gamma}(\boldsymbol{\beta})\|$. Then,

$$\begin{aligned}\frac{\|\boldsymbol{\gamma}^*(\boldsymbol{\beta})\|}{\|\boldsymbol{\gamma}(\boldsymbol{\beta})\|} &\leq c_3^{-1}|\tilde{\mathbf{d}}_j^{*\prime}\boldsymbol{\gamma}^*(\boldsymbol{\beta})| \times \frac{\|\tilde{\mathbf{d}}_j^{*\prime}\|}{|\tilde{\mathbf{d}}_j^{*\prime}\boldsymbol{\gamma}(\boldsymbol{\beta})|} \times \frac{|\tilde{\mathbf{d}}_j^{*\prime}\boldsymbol{\gamma}^*(\boldsymbol{\beta})|}{c_3\|\boldsymbol{\gamma}^*(\boldsymbol{\beta})\|} \\ &= \frac{\{\tilde{\mathbf{d}}_j^{*\prime}\boldsymbol{\gamma}^*(\boldsymbol{\beta})\}^2}{\{\tilde{\mathbf{d}}_j^{*\prime}\boldsymbol{\gamma}(\boldsymbol{\beta})\}^2\|\boldsymbol{\gamma}^*(\boldsymbol{\beta})\|}\|\tilde{\mathbf{d}}_j^{*\prime}\||\tilde{\mathbf{d}}_j^{*\prime}\boldsymbol{\gamma}(\boldsymbol{\beta})|O_p(1).\end{aligned} \tag{S1.9}$$

Since $\mathbf{T}_2'\boldsymbol{\beta}_0 = 0$ and $\boldsymbol{\gamma}(\boldsymbol{\beta}) = \mathbf{T}_2'\boldsymbol{\beta}$, for $\boldsymbol{\beta} \in \mathfrak{B}$, we have $\|\boldsymbol{\gamma}(\boldsymbol{\beta})\| \leq \delta_n\sqrt{p_n/n}$. Together with $\delta_n p_n/\lambda_n \to 0$, (S1.9) implies that with probability tending to 1,

$$\sup_{\boldsymbol{\beta}\in\mathfrak{B}} \frac{\|\boldsymbol{\gamma}^*(\boldsymbol{\beta})\|}{\|\boldsymbol{\gamma}(\boldsymbol{\beta})\|} = \sup_{\boldsymbol{\beta}\in\mathfrak{B}} \frac{\|\mathbf{T}_2'\mathbf{g}(\boldsymbol{\beta})\|}{\|\mathbf{T}_2'\boldsymbol{\beta}\|} = O_p\Big(\frac{\delta_n p_n}{\lambda_n}\Big) = o_p(1). \tag{S1.10}$$

This proves statement (b) in Lemma 1.

To show that with probability tending to 1, $\mathbf{g}(\cdot)$ is a mapping from the ball $\mathfrak{B}$ to itself, it suffices to show that

$$\mathbf{P}\left(\sup_{\boldsymbol{\beta}\in\mathfrak{B}} \|\mathbf{T}_1'\{\mathbf{g}(\boldsymbol{\beta})-\boldsymbol{\beta}_0\}\| \le \delta_n\sqrt{\frac{p_n}{n}}\right) \to 1.$$

In a similar, we rewrite equation (S1.2) as

$$\begin{aligned}\mathbf{T}_1'\{\mathbf{g}(\boldsymbol{\beta})-\boldsymbol{\beta}_0\}+&\frac{\lambda_n}{n}\mathbf{T}_1'\boldsymbol{\Sigma}_n^{-1}\mathbf{D}_1(\boldsymbol{\beta})\mathbf{g}(\boldsymbol{\beta})\\ &+\frac{\lambda_n}{n}\mathbf{T}_1'\boldsymbol{\Sigma}_n^{-1}\mathbf{D}_2(\boldsymbol{\beta})\mathbf{g}(\boldsymbol{\beta}) = \mathbf{T}_1'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}.\end{aligned}$$

Similar to equation (S1.6), we have

$$\sup_{\boldsymbol{\beta}\in\mathfrak{B}}\left\|\mathbf{T}_1'\{\mathbf{g}(\boldsymbol{\beta})-\boldsymbol{\beta}_0\}+\frac{\lambda_n}{n}\mathbf{T}_1'\boldsymbol{\Sigma}_n^{-1}\mathbf{D}_2(\boldsymbol{\beta})\mathbf{g}(\boldsymbol{\beta})\right\| = O_p\Big(\sqrt{\frac{p_n}{n}}\Big).$$

Observe that

$$\begin{aligned}\sup_{\boldsymbol{\beta}\in\mathfrak{B}}\left\|\frac{\lambda_n}{n}\mathbf{T}_1'\boldsymbol{\Sigma}_n^{-1}\mathbf{D}_2(\boldsymbol{\beta})\mathbf{g}(\boldsymbol{\beta})\right\| &= \sup_{\boldsymbol{\beta}\in\mathfrak{B}}\left\|\frac{\lambda_n}{n}\mathbf{T}_1'\boldsymbol{\Sigma}_n^{-1}\mathbf{T}_2\mathbf{T}_2'\mathbf{D}_2(\boldsymbol{\beta})\mathbf{T}_2\mathbf{T}_2'\mathbf{g}(\boldsymbol{\beta})\right\|\\ &\le \sup_{\boldsymbol{\beta}\in\mathfrak{B}}\left\|\frac{\lambda_n}{n}\tilde{\mathbf{D}}_2(\boldsymbol{\beta})\boldsymbol{\gamma}^*(\boldsymbol{\beta})\right\|\cdot\|\mathbf{T}_1'\boldsymbol{\Sigma}_n^{-1}\mathbf{T}_2\|\\ &= \sup_{\boldsymbol{\beta}\in\mathfrak{B}}\frac{\lambda_n}{n}\frac{\boldsymbol{\gamma}^*(\boldsymbol{\beta})'\tilde{\mathbf{D}}_2(\boldsymbol{\beta})\boldsymbol{\gamma}^*(\boldsymbol{\beta})}{\|\boldsymbol{\gamma}^*(\boldsymbol{\beta})\|}O_p(1)\\ &= O_p\Big(\sqrt{\frac{p_n}{n}}\Big).\end{aligned}$$

The last equation comes from (S1.8). Hence, we have

$$\sup_{\boldsymbol{\beta}\in\mathfrak{B}}\|\mathbf{T}_1'\{\mathbf{g}(\boldsymbol{\beta})-\boldsymbol{\beta}_0\}\| = O_p\Big(\sqrt{\frac{p_n}{n}}\Big).$$

It follows that

$$\mathbf{P}\left(\sup_{\boldsymbol{\beta}\in\mathfrak{B}} \|\mathbf{T}_1'\{\mathbf{g}(\boldsymbol{\beta})-\boldsymbol{\beta}_0\}\| \le \delta_n\sqrt{\frac{p_n}{n}}\right) \to 1. \qquad \text{(S1.11)}$$

On the other hand, the statement (S1.10) implies

$$\mathbf{P}\left(\sup_{\boldsymbol{\beta}\in\mathfrak{B}}\|\mathbf{T}_2'\{\mathbf{g}(\boldsymbol{\beta})-\boldsymbol{\beta}_0\}\|\leq\delta_n\sqrt{\frac{p_n}{n}}\right)\to 1. \tag{S1.12}$$

Hence, (S1.11) combined with (S1.12) yields

$$\mathbf{P}\left(\sup_{\boldsymbol{\beta}\in\mathfrak{B}}\|\mathbf{g}(\boldsymbol{\beta})-\boldsymbol{\beta}_0\|\leq\delta_n\sqrt{\frac{p_n}{n}}\right)\to 1.$$

This proves that $\mathbf{g}(\cdot)$ is a mapping from $\mathfrak{B}$ to itself with probability tending to 1. □

S1.2 Proof of Lemma 2

Proof. Recall that

$$\{\mathbf{X}_1'\mathbf{X}_1+\lambda_n\tilde{\mathbf{D}}(\mathbf{T}_1'\boldsymbol{\beta})\}\mathbf{f}(\mathbf{T}_1'\boldsymbol{\beta})=\mathbf{X}_1'\mathbf{y},$$

where $\mathbf{X}_1=\mathbf{X}\mathbf{T}_1$ and

$$\tilde{\mathbf{D}}(\mathbf{T}_1'\boldsymbol{\beta})=\mathbf{T}_1'\sum_{k=1}^{q_n}\frac{\mathbf{d}_k\mathbf{d}_k'}{\tilde{c}_k^2(\mathbf{T}_1'\boldsymbol{\beta})}\mathbf{T}_1 \quad\text{with}\quad \tilde{c}_k(\mathbf{T}_1'\boldsymbol{\beta})=\mathbf{d}_k'\mathbf{T}_1\mathbf{T}_1'\boldsymbol{\beta}.$$

Similarly, we have

$$\mathbf{f}(\mathbf{T}_1'\boldsymbol{\beta})-\boldsymbol{\theta}_0+\frac{\lambda_n}{n}\tilde{\boldsymbol{\Sigma}}_{n1}^{-1}\tilde{\mathbf{D}}(\mathbf{T}_1'\boldsymbol{\beta})\mathbf{f}(\mathbf{T}_1'\boldsymbol{\beta})=(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\boldsymbol{\varepsilon},$$

where $\boldsymbol{\theta}_0=\mathbf{T}_1'\boldsymbol{\beta}_0$. Recalling that $\mathfrak{B}_1=\{\mathbf{T}_1'\boldsymbol{\beta}\in\mathbb{R}^{m_n}:\|\mathbf{T}_1'\boldsymbol{\beta}-\mathbf{T}_1'\boldsymbol{\beta}_0\|\leq\delta_n\sqrt{p_n/n}\}$. It is straightforward to show that $\mathbf{f}(\cdot)$ is a mapping from $\mathfrak{B}_1$ to itself. In fact,

$$\|(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\boldsymbol{\varepsilon}\|=O_p\Big(\sqrt{\frac{m_n}{n}}\Big)=o_p\Big(\sqrt{\frac{p_n}{n}}\Big)$$

and

$$\sup_{\boldsymbol{\beta}\in\mathfrak{B}_1}\left\|\frac{\lambda_n}{n}\tilde{\boldsymbol{\Sigma}}_{n1}^{-1}\tilde{\mathbf{D}}(\mathbf{T}_1'\boldsymbol{\beta})\mathbf{f}(\mathbf{T}_1'\boldsymbol{\beta})\right\| \leq \frac{\lambda_n}{n}\|\tilde{\boldsymbol{\Sigma}}_{n1}^{-1}\|\sup_{\boldsymbol{\beta}\in\mathfrak{B}_1}\left\|\tilde{\mathbf{D}}(\mathbf{T}_1'\boldsymbol{\beta})\mathbf{f}(\mathbf{T}_1'\boldsymbol{\beta})\right\|$$
$$= O_p\Big(\frac{\lambda_n q_n\sqrt{m_n}}{nb_n^2}\Big).$$

Then, from assumption (A2), it follows that

$$\sup_{\boldsymbol{\beta}\in\mathfrak{B}_1}\left\|\frac{\lambda_n}{n}\tilde{\boldsymbol{\Sigma}}_{n1}^{-1}\tilde{\mathbf{D}}(\mathbf{T}_1'\boldsymbol{\beta})\mathbf{f}(\mathbf{T}_1'\boldsymbol{\beta})\right\| = o_p\Big(\sqrt{\frac{p_n}{n}}\Big).$$

Therefore,

$$\mathbf{P}\left(\sup_{\boldsymbol{\beta}\in\mathfrak{B}_1}\|\mathbf{f}(\mathbf{T}_1'\boldsymbol{\beta}) - \mathbf{T}_1'\boldsymbol{\beta}_0\| \leq \delta_n\sqrt{\frac{p_n}{n}}\right) \to 1.$$

This completes the proof that $\mathbf{f}(\cdot)$ is a mapping from $\mathfrak{B}_1$ to itself.

We next show that $\mathbf{f}(\cdot)$ is a contraction mapping. Since

$$\left\{\frac{1}{n}\mathbf{X}_1'\mathbf{X}_1 + \frac{\lambda_n}{n}\tilde{\mathbf{D}}(\mathbf{T}_1'\boldsymbol{\beta})\right\}\mathbf{f}(\mathbf{T}_1'\boldsymbol{\beta}) = \frac{1}{n}\mathbf{X}_1'\mathbf{y}, \tag{S1.13}$$

differentiating both sides of equation (S1.13) with respect to $\boldsymbol{\beta}'$ yields

$$\left\{\boldsymbol{\Sigma}_{n1} + \frac{\lambda_n}{n}\tilde{\mathbf{D}}(\mathbf{T}_1'\boldsymbol{\beta})\right\}\dot{\mathbf{f}}(\mathbf{T}_1'\boldsymbol{\beta})\mathbf{T}_1' = \frac{2\lambda_n}{n}\sum_{k=1}^{q_n}\frac{\mathbf{T}_1'\mathbf{d}_k\mathbf{d}_k'\mathbf{T}_1\mathbf{f}(\mathbf{T}_1'\boldsymbol{\beta})\mathbf{d}_k'\mathbf{T}_1\mathbf{T}_1'}{(\mathbf{d}_k'\mathbf{T}_1\mathbf{T}_1'\boldsymbol{\beta})^3}.$$

Hence, according to assumptions (A2) and (A3)

$$\sup_{\boldsymbol{\beta}\in\mathfrak{B}_1}\left\|\left\{\boldsymbol{\Sigma}_{n1} + \frac{\lambda_n}{n}\tilde{\mathbf{D}}(\mathbf{T}_1'\boldsymbol{\beta})\right\}\dot{\mathbf{f}}(\mathbf{T}_1'\boldsymbol{\beta})\mathbf{T}_1'\right\|$$
$$= \sup_{\boldsymbol{\beta}\in\mathfrak{B}_1}\left\|\frac{2\lambda_n}{n}\sum_{k=1}^{q_n}\frac{\mathbf{T}_1'\mathbf{d}_k\mathbf{d}_k'\mathbf{T}_1\mathbf{f}(\mathbf{T}_1'\boldsymbol{\beta})\mathbf{d}_k'\mathbf{T}_1\mathbf{T}_1'}{\tilde{c}_k^3(\mathbf{T}_1'\boldsymbol{\beta})}\right\|$$
$$= O_p\Big(\frac{\lambda_n q_n}{nb_n^3}\sqrt{m_n}\Big) = o_p(1).$$

Furthermore, since

$$\begin{aligned}&\sup_{\boldsymbol{\beta}\in\mathfrak{B}_1}\left\|\left\{\boldsymbol{\Sigma}_{n1}+\frac{\lambda_n}{n}\tilde{\mathbf{D}}(\mathbf{T}_1'\boldsymbol{\beta})\right\}\dot{\mathbf{f}}(\mathbf{T}_1'\boldsymbol{\beta})\mathbf{T}_1'\right\|\\&\geq\sup_{\boldsymbol{\beta}\in\mathfrak{B}_1}\frac{1}{C}\|\dot{\mathbf{f}}(\mathbf{T}_1'\boldsymbol{\beta})\mathbf{T}_1'\|-\sup_{\boldsymbol{\beta}\in\mathfrak{B}_1}\frac{\lambda_n}{n}\|\tilde{\mathbf{D}}(\mathbf{T}_1'\boldsymbol{\beta})\dot{\mathbf{f}}(\mathbf{T}_1'\boldsymbol{\beta})\mathbf{T}_1'\|,\end{aligned}$$

it follows from assumption (A2) that

$$\sup_{\boldsymbol{\beta}\in\mathfrak{B}_1}\|\dot{\mathbf{f}}(\mathbf{T}_1'\boldsymbol{\beta})\mathbf{T}_1'\|=\sup_{\boldsymbol{\beta}\in\mathfrak{B}_1}\|\dot{\mathbf{f}}(\mathbf{T}_1'\boldsymbol{\beta})\|=o_p(1).$$

Therefore, $\mathbf{f}(\cdot)$ is a contraction mapping from $\mathfrak{B}_1$ to itself. This indicates that there exists one unique fixed point of $\mathbf{f}(\cdot)$ in the region $\mathfrak{B}_1$ denoted as $\hat{\boldsymbol{\theta}}^\circ$ such that

$$\mathbf{f}(\hat{\boldsymbol{\theta}}^\circ)=\{\mathbf{X}_1'\mathbf{X}_1+\lambda_n\tilde{\mathbf{D}}(\hat{\boldsymbol{\theta}}^\circ)\}^{-1}\mathbf{X}_1'\mathbf{y}.$$

Hence, by the first order resolvent expansion formula $(\mathbf{H}+\boldsymbol{\Delta})^{-1}=\mathbf{H}^{-1}-\mathbf{H}^{-1}\boldsymbol{\Delta}(\mathbf{H}+\boldsymbol{\Delta})^{-1}$, we have

$$\begin{aligned}&\hat{\boldsymbol{\theta}}^\circ-\boldsymbol{\theta}_0\\&=\{\mathbf{X}_1'\mathbf{X}_1+\lambda_n\tilde{\mathbf{D}}(\hat{\boldsymbol{\theta}}^\circ)\}^{-1}\mathbf{X}_1'\mathbf{y}-\boldsymbol{\theta}_0\\&=(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{y}-\boldsymbol{\theta}_0-(\mathbf{X}_1'\mathbf{X}_1)^{-1}\lambda_n\tilde{\mathbf{D}}(\hat{\boldsymbol{\theta}}^\circ)\{\mathbf{X}_1'\mathbf{X}_1+\lambda_n\tilde{\mathbf{D}}(\hat{\boldsymbol{\theta}}^\circ)\}^{-1}\mathbf{X}_1'\mathbf{y}\\&=(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\boldsymbol{\varepsilon}-(\mathbf{X}_1'\mathbf{X}_1)^{-1}\lambda_n\tilde{\mathbf{D}}(\hat{\boldsymbol{\theta}}^\circ)\{\mathbf{X}_1'\mathbf{X}_1+\lambda_n\tilde{\mathbf{D}}(\hat{\boldsymbol{\theta}}^\circ)\}^{-1}\mathbf{X}_1'(\mathbf{X}_1\boldsymbol{\theta}_0+\boldsymbol{\varepsilon})\\&=(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\boldsymbol{\varepsilon}-\frac{\lambda_n}{n}\tilde{\boldsymbol{\Sigma}}_{n1}^{-1}\tilde{\mathbf{D}}(\hat{\boldsymbol{\theta}}^\circ)\left\{\mathbf{I}_{m_n}+\frac{\lambda_n}{n}\tilde{\boldsymbol{\Sigma}}_{n1}^{-1}\tilde{\mathbf{D}}(\hat{\boldsymbol{\theta}}^\circ)\right\}^{-1}\boldsymbol{\theta}_0\\&\qquad-\frac{\lambda_n}{n}\tilde{\boldsymbol{\Sigma}}_{n1}^{-1}\tilde{\mathbf{D}}(\hat{\boldsymbol{\theta}}^\circ)\left\{\mathbf{I}_{m_n}+\frac{\lambda_n}{n}\tilde{\boldsymbol{\Sigma}}_{n1}^{-1}\tilde{\mathbf{D}}(\hat{\boldsymbol{\theta}}^\circ)\right\}^{-1}\tilde{\boldsymbol{\Sigma}}_{n1}^{-1}\frac{\mathbf{X}_1'\boldsymbol{\varepsilon}}{n}.\end{aligned}$$

Therefore, we have

$$\begin{aligned}
&\sqrt{n}s_n^{-1}\mathbf{a}_n'(\hat{\boldsymbol\theta}^\circ-\boldsymbol\theta_0)\\
&= s_n^{-1}\mathbf{a}_n'\tilde{\boldsymbol\Sigma}_{n1}^{-1}\frac{\mathbf{X}_1'\boldsymbol\varepsilon}{\sqrt{n}} - \frac{\lambda_n}{\sqrt{n}}s_n^{-1}\mathbf{a}_n'\tilde{\boldsymbol\Sigma}_{n1}^{-1}\tilde{\mathbf{D}}(\hat{\boldsymbol\theta}^\circ)\left\{\mathbf{I}_{m_n}+\frac{\lambda_n}{n}\tilde{\boldsymbol\Sigma}_{n1}^{-1}\tilde{\mathbf{D}}(\hat{\boldsymbol\theta}^\circ)\right\}^{-1}\boldsymbol\theta_0\\
&\qquad - \frac{\lambda_n}{\sqrt{n}}s_n^{-1}\mathbf{a}_n'\tilde{\boldsymbol\Sigma}_{n1}^{-1}\tilde{\mathbf{D}}(\hat{\boldsymbol\theta}^\circ)\left\{\mathbf{I}_{m_n}+\frac{\lambda_n}{n}\tilde{\boldsymbol\Sigma}_{n1}^{-1}\tilde{\mathbf{D}}(\hat{\boldsymbol\theta}^\circ)\right\}^{-1}\tilde{\boldsymbol\Sigma}_{n1}^{-1}\frac{\mathbf{X}_1'\boldsymbol\varepsilon}{n}\\
&= s_n^{-1}\mathbf{a}_n'\tilde{\boldsymbol\Sigma}_{n1}^{-1}\frac{\mathbf{X}_1'\boldsymbol\varepsilon}{\sqrt{n}} - I_1 - I_2.
\end{aligned}$$

By assumption (A2) and the condition $\inf_{\boldsymbol\beta\in\mathfrak{B}_1}(\mathbf{d}_k'\mathbf{T}_1\mathbf{T}_1'\boldsymbol\beta)^2 \geq c_1(\mathbf{d}_k'\mathbf{T}_1\boldsymbol\theta_0)^2$, we have

$$\begin{aligned}
\|I_1\| &\leq \frac{\lambda_n}{\sqrt{n}}C^2\left\|\tilde{\mathbf{D}}(\hat{\boldsymbol\theta}^\circ)\left\{\mathbf{I}_{m_n}+\frac{\lambda_n}{n}\tilde{\boldsymbol\Sigma}_{n1}^{-1}\tilde{\mathbf{D}}(\hat{\boldsymbol\theta}^\circ)\right\}^{-1}\boldsymbol\theta_0\right\|\\
&\leq \frac{\lambda_n}{\sqrt{n}}C^2\|\tilde{\mathbf{D}}(\hat{\boldsymbol\theta}^\circ)\boldsymbol\theta_0\|\{1+o_p(1)\}\\
&\leq \frac{\lambda_n}{\sqrt{n}}C^2\left\|\sum_{k=1}^{q_n}\frac{\mathbf{T}_1'\mathbf{d}_k\mathbf{d}_k'\mathbf{T}_1\boldsymbol\theta_0}{(\mathbf{d}_k'\mathbf{T}_1\hat{\boldsymbol\theta}^\circ)^2}\right\|\{1+o_p(1)\}\\
&= O_p\Big(\frac{\lambda_n q_n}{\sqrt{n}b_n^2}\Big)\to 0.
\end{aligned}$$

On the other hand,

$$\|I_2\| \leq \frac{\lambda_n}{\sqrt{n}b_n^2}C^3 q_n\left\|\frac{\mathbf{X}_1'\boldsymbol\varepsilon}{n}\right\| = O_p\Big(\frac{\lambda_n q_n}{\sqrt{n}b_n^2}\Big)O_p\Big(\sqrt{\frac{m_n}{n}}\Big) = o_p(1).$$

As a result,

$$\sqrt{n}s_n^{-1}\mathbf{a}_n'(\hat{\boldsymbol\theta}^\circ-\boldsymbol\theta_0) = s_n^{-1}\mathbf{a}_n'\tilde{\boldsymbol\Sigma}_{n1}^{-1}\frac{\mathbf{X}_1'\boldsymbol\varepsilon}{\sqrt{n}} + o_p(1).$$

It follows from the Lindeberg-Feller central limit theorem that

$$\sqrt{n}s_n^{-1}\mathbf{a}_n'(\hat{\boldsymbol\theta}^\circ-\boldsymbol\theta_0)\to\mathcal{N}(0,1).$$

This completes the proof of Lemma 2. □

Proof of Theorem 1. Observe that Lemma 1 implies that

$$\mathbf{P}\left(\lim_{j\to\infty}\mathbf{T}_2{'}\hat{\boldsymbol{\beta}}^{(j)}=0\right)\to 1,\ \text{as } n\to\infty.$$

We show that with probability tending to 1, $\lim_{j\to\infty}\mathbf{T}_1{'}\hat{\boldsymbol{\beta}}^{(j)}$ exists. Since $\mathbf{d}_k'\mathbf{T}_1=\mathbf{0}$, for all $k=q_n+1,\ldots,K_n$ and

$$\mathbf{T}'\{\mathbf{X}'\mathbf{X}+\lambda_n\mathbf{D}(\boldsymbol{\beta})\}\mathbf{T}\mathbf{T}'\mathbf{g}(\boldsymbol{\beta})=\mathbf{T}'\mathbf{X}'\mathbf{y},$$

we have

$$\{\mathbf{X}_1'\mathbf{X}_1+\lambda_n\mathbf{T}_1'\mathbf{D}_1(\boldsymbol{\beta})\mathbf{T}_1\}\mathbf{T}_1'\mathbf{g}(\boldsymbol{\beta})+\{\mathbf{X}_1'\mathbf{X}_2+\lambda_n\mathbf{T}_1'\mathbf{D}_1(\boldsymbol{\beta})\mathbf{T}_2\}\mathbf{T}_2'\mathbf{g}(\boldsymbol{\beta})=\mathbf{X}_1'\mathbf{y},$$

$$\{\mathbf{X}_2'\mathbf{X}_1+\lambda_n\mathbf{T}_2'\mathbf{D}_1(\boldsymbol{\beta})\mathbf{T}_1\}\mathbf{T}_1'\mathbf{g}(\boldsymbol{\beta})+\{\mathbf{X}_2'\mathbf{X}_2+\lambda_n\mathbf{T}_2'\mathbf{D}(\boldsymbol{\beta})\mathbf{T}_2\}\mathbf{T}_2'\mathbf{g}(\boldsymbol{\beta})=\mathbf{X}_2'\mathbf{y}. \tag{S1.14}$$

Define $\mathbf{T}_2'\mathbf{g}(\boldsymbol{\beta})=\mathbf{0}$ if $\mathbf{T}_2'\boldsymbol{\beta}=\mathbf{0}$. Then $\mathbf{T}'\mathbf{g}(\boldsymbol{\beta})$ is continuous. In fact, from (S1.14), we have $\lim_{\mathbf{T}_2'\boldsymbol{\beta}\to 0}\mathbf{T}_2'\mathbf{g}(\boldsymbol{\beta})=\mathbf{0}$. On the other hand, since $\inf_{\boldsymbol{\beta}\in\mathfrak{B}_1}(\mathbf{d}_k'\mathbf{T}_1\mathbf{T}_1'\boldsymbol{\beta})^2\geq c_1(\mathbf{d}_k'\mathbf{T}_1\mathbf{T}_1'\boldsymbol{\beta}_0)^2$ holds for $1\leq k\leq q_n$ and assumption (A3) hold, we have

$$\begin{aligned}\mathbf{T}_1'\mathbf{D}_1(\boldsymbol{\beta})\mathbf{T}_1&=\sum_{k=1}^{q_n}\frac{\mathbf{T}_1'\mathbf{d}_k\mathbf{d}_k'\mathbf{T}_1}{(\mathbf{d}_k'\mathbf{T}_1\mathbf{T}_1'\boldsymbol{\beta}+\mathbf{d}_k'\mathbf{T}_2\mathbf{T}_2'\boldsymbol{\beta})^2}\\&\to\sum_{k=1}^{q_n}\frac{\mathbf{T}_1'\mathbf{d}_k\mathbf{d}_k'\mathbf{T}_1}{(\mathbf{d}_k'\mathbf{T}_1\mathbf{T}_1'\boldsymbol{\beta})^2}=\tilde{\mathbf{D}}(\mathbf{T}_1'\boldsymbol{\beta}),\ \text{as } \mathbf{T}_2'\boldsymbol{\beta}\to\mathbf{0},\end{aligned}$$

or equivalently

$$\|\mathbf{T}_1'\mathbf{D}_1(\boldsymbol{\beta})\mathbf{T}_1-\tilde{\mathbf{D}}(\mathbf{T}_1'\boldsymbol{\beta})\|\to 0,\ \text{as } \mathbf{T}_2'\boldsymbol{\beta}\to\mathbf{0}.$$

It follows from (S1.2) that

$$\lim_{\mathbf{T}_2'\boldsymbol{\beta}\to 0} \mathbf{T}_1'\mathbf{g}(\boldsymbol{\beta}) = \mathbf{f}(\mathbf{T}_1'\boldsymbol{\beta}).$$

Therefore,

$$\|\mathbf{T}_1'\mathbf{g}(\boldsymbol{\beta}^{(j)}) - \mathbf{f}(\mathbf{T}_1'\boldsymbol{\beta}^{(j)})\| \to 0, \text{ as } j \to \infty.$$

Recall that Lemma 2 shows that $\hat{\boldsymbol{\theta}}^{\circ}$ is the unique fixed point of $\mathbf{f}(\cdot)$ from $\mathfrak{B}_1$ to itself. Therefore, with probability tending to 1,

$$\begin{aligned}
\|\mathbf{T}_1'\hat{\boldsymbol{\beta}}^{(j+1)} - \hat{\boldsymbol{\theta}}^{\circ}\| &= \|\mathbf{T}_1'\mathbf{g}(\hat{\boldsymbol{\beta}}^{(j)}) - \hat{\boldsymbol{\theta}}^{\circ}\| \\
&\le \|\mathbf{T}_1'\mathbf{g}(\hat{\boldsymbol{\beta}}^{(j)}) - \mathbf{f}(\mathbf{T}_1'\hat{\boldsymbol{\beta}}^{(j)})\| + \|\mathbf{f}(\mathbf{T}_1'\hat{\boldsymbol{\beta}}^{(j)}) - \mathbf{f}(\hat{\boldsymbol{\theta}}^{\circ})\| \\
&\le \eta_j + \frac{1}{\tilde{C}}\|\mathbf{T}_1'\hat{\boldsymbol{\beta}}^{(j)} - \hat{\boldsymbol{\theta}}^{\circ}\|, \text{ for some constant } \tilde{C} > 1
\end{aligned}$$

where $\eta_j \to 0$ as $j \to \infty$. The last inequality is due to that $\mathbf{f}(\cdot)$ is a contraction mapping from $\mathfrak{B}_1$ to itself as stated in Lemma 2. Set $a_j = \|\mathbf{T}_1'\hat{\boldsymbol{\beta}}^{(j)} - \hat{\boldsymbol{\theta}}^{\circ}\|$. For any $\epsilon > 0$, there exists a $N > 0$, such that $|\eta_j| < \epsilon$ holds for all $j > N$. When $j > N$, we have that with probability tending to 1,

$$\begin{aligned}
a_{j+1} \le \eta_j + \frac{a_j}{\tilde{C}} &\le \eta_j + \frac{1}{\tilde{C}}(\eta_{j-1} + a_{j-1}/\tilde{C}) \\
&\le \frac{a_1}{\tilde{C}^j} + \frac{\eta_1}{\tilde{C}^{j-1}} + \cdots + \frac{\eta_N}{\tilde{C}^{j-N}} + \frac{\eta_{N+1}}{\tilde{C}^{j-N-1}} + \cdots + \frac{\eta_{j-1}}{\tilde{C}} + \eta_j \\
&\le M_1 \frac{1}{\tilde{C}^{j-N}} + \epsilon M_2 + 2\epsilon \cdot \frac{1}{\tilde{C}^{j-N}} \longrightarrow 0, \text{ as } j \to \infty,
\end{aligned}$$

for some constant $M_1 > 0$ and $M_2 > 0$. This proves that

$$\mathbf{P}\left(\lim_{j\to\infty} \mathbf{T}_1{}'\hat{\boldsymbol{\beta}}^{(j)} = \hat{\boldsymbol{\theta}}^{\circ}\right) \to 1, \text{ as } n \to \infty.$$

Since from Lemma 2,

$$\sqrt{n} s_n^{-1} \mathbf{a}_n' (\hat{\boldsymbol{\theta}}^\circ - \mathbf{T}_1' \boldsymbol{\beta}_0) \to \mathcal{N}(0,1)$$

with probability tending to 1, we have

$$\sqrt{n} s_n^{-1} \mathbf{a}_n' (\mathbf{T}_1' \hat{\boldsymbol{\beta}} - \mathbf{T}_1' \boldsymbol{\beta}_0) \to \mathcal{N}(0,1),$$

with probability tending to 1. This completes the proof of Theorem 1. □