Supplement to "Endogeneity in High Dimensions"

Jianqing Fan Yuan Liao

Princeton University University of Maryland

Abstract

This supplementary material contains the proof of Theorem 7.1 in the main paper.

The following lemma is useful.

Lemma 0.1. *For a general loss function L_n, suppose it is twice differentiable in a neighborhood of $\boldsymbol{\beta}_0$. Assume*

$$\max_{l\notin S}\left|\frac{\partial L_n(\boldsymbol{\beta}_0)}{\partial \beta_l}\right| = o_p(P_n'(0^+)), \tag{0.1}$$

and there is a neighborhood $U \subset \mathbb{R}^s$ of $\boldsymbol{\beta}_{0S}$ such that

$$\sup_{\beta_S\in U}\max_{l\notin S, j\in S}\left|\frac{\partial^2 L_n(\boldsymbol{\beta}_S, 0)}{\partial \beta_l \partial \beta_j}\right| = o_p\left(\frac{P_n'(0^+)}{\sqrt{s}k_n}\right). \tag{0.2}$$

where we denote $P_n'(0^+) = \liminf_{t\to 0^+} P_n'(t)$ and $k_n = a_n + \sqrt{s}P_n'(d_n)$ as given in Theorem B.1 of the main paper. In addition, assume $\sqrt{s}k_n = o(P_n'(0^+))$. Then Condition A in Theorem B.2 of the main paper is satisfied.

Proof. If L_n is continuously differentiable in a neighborhood of $\boldsymbol{\beta}_0$, by the mean value theorem, there exists $\lambda \in (0,1)$ such that for $\mathbf{h} = \lambda\boldsymbol{\beta} + (1-\lambda)\mathbb{T}\boldsymbol{\beta}$,

$$\begin{aligned} Q_n(\mathbb{T}\boldsymbol{\beta}) - Q_n(\boldsymbol{\beta}) &= \sum_{l\notin S}\frac{\partial L_n(\mathbf{h})}{\partial \beta_l}(-\beta_l) - \sum_{l\notin S} P_n'(|h_l|)|\beta_l| \\ &\leq \sum_{l\notin S}\left(\left|\frac{\partial L_n(\mathbf{h})}{\partial \beta_l}\right| - P_n'(|h_l|)\right)|\beta_l|, \end{aligned} \tag{0.3}$$

where we used the fact that $\mathrm{sgn}(h_l) = \mathrm{sgn}(\beta_l)$ for $l \notin S$. Suppose we have (which we will prove later)

$$\max_{l\notin S}\left|\frac{\partial L_n(\widehat{\boldsymbol{\beta}})}{\partial \beta_l}\right| = o_p(P_n'(0^+)), \tag{0.4}$$

the following event then holds with probability approaching one:

$$A_n = \left\{ \max_{l \notin S} \left| \frac{\partial L_n(\widehat{\boldsymbol{\beta}})}{\partial \beta_l} \right| < P_n'(0^+)/2 \right\}.$$

Conditioning on the event A_n, by the continuity, there is a neighborhood $\mathcal{H}$ of $\widehat{\boldsymbol{\beta}}$ such that for any $\boldsymbol{\beta} \in \mathcal{H}$, $\max_{l \notin S} |\partial L_n(\boldsymbol{\beta})/\partial \beta_l| < P_n'(0^+)/2$. Also by the continuity of $P_n'(\cdot)$ in a right-neighborhood of 0, we can make $\mathcal{H}$ sufficiently small so that there is a $\delta > 0$ sufficiently small, $\max_{\beta \in \mathcal{H}, l \notin S} |\beta_l| < \delta$ and

$$\max_{l \notin S} \max_{\beta \in \mathcal{H}} |\partial L_n(\boldsymbol{\beta})/\partial \beta_l| < P_n'(\delta)/2.$$

In addition, since $\mathbb{T}(\widehat{\boldsymbol{\beta}}) = \widehat{\boldsymbol{\beta}}$ and $\|\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}\|^2 = \|\mathbb{T}\boldsymbol{\beta} - \boldsymbol{\beta}\|^2 + \|\mathbb{T}\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}\|^2 \geq \|\mathbb{T}\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}\|^2$, we have $\boldsymbol{\beta} \in \mathcal{H}$ implies that both $\mathbb{T}\boldsymbol{\beta}$ and $\mathbf{h} = (\boldsymbol{\beta}_S^T, \lambda \boldsymbol{\beta}_N^T)^T$ are also inside $\mathcal{H}$, where $\mathbf{h}$ and $\lambda \in (0, 1)$ are defined as before. This then implies that $\max_{l \notin S} |\partial L_n(\mathbf{h})/\partial \beta_l| < P_n'(\delta)/2$ uniformly over $\boldsymbol{\beta} \in \mathcal{H}$, and $\max_{\beta \in \mathcal{H}, l \notin S} |\beta_l| < \delta$. Hence, the non-increasingness of $P_n'(\cdot)$ implies

$$\begin{aligned} \max_{l \notin S} \left| \frac{\partial L_n(\mathbf{h})}{\partial \beta_l} \right| - P_n'(|h_l|) &= \max_{l \notin S} \left| \frac{\partial L_n(\mathbf{h})}{\partial \beta_l} \right| - P_n'(\lambda |\beta_l|) \\ &\leq \max_{l \notin S} \left| \frac{\partial L_n(\mathbf{h})}{\partial \beta_l} \right| - P_n'(\delta) < 0. \end{aligned}$$

This with (0.3) implies the inequality (B.6) of the main paper. Because $P(A_n) \to 1$, we have our conclusion.

It then remains to prove (0.4). By the triangular inequality,

$$\max_{l \notin S} \left| \frac{\partial L_n(\widehat{\boldsymbol{\beta}})}{\partial \beta_l} \right| \leq \max_{l \notin S} \left| \frac{\partial L_n(\widehat{\boldsymbol{\beta}})}{\partial \beta_l} - \frac{\partial L_n(\boldsymbol{\beta}_0)}{\partial \beta_l} \right| + \max_{l \notin S} \left| \frac{\partial L_n(\boldsymbol{\beta}_0)}{\partial \beta_l} \right|.$$

By assumption, $\max_{l \notin S} |\partial L_n(\boldsymbol{\beta}_0)/\partial \beta_l| = o_p(P_n'(0^+))$. For the first term on the right hand side, apply the mean value theorem (note that $\widehat{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}_0$ only differ in the coordinates in S),

$$\begin{aligned} \max_{l \notin S} \left| \frac{\partial L_n(\widehat{\boldsymbol{\beta}})}{\partial \beta_l} - \frac{\partial L_n(\boldsymbol{\beta}_0)}{\partial \beta_l} \right| &\leq \max_{l \notin S} \left| \sum_{j \in S} \frac{\partial^2 L_n(\tilde{\boldsymbol{\beta}})}{\partial \beta_l \partial \beta_j} (\hat{\beta}_j - \beta_{0j}) \right| \\ \leq \max_{l \notin S, j \in S} \left| \frac{\partial^2 L_n(\tilde{\boldsymbol{\beta}})}{\partial \beta_l \partial \beta_j} \right| \sqrt{s} \|\widehat{\boldsymbol{\beta}}_S - \boldsymbol{\beta}_{0S}\| \equiv \xi. \end{aligned} \tag{0.5}$$

where $\tilde{\boldsymbol{\beta}} = (\tilde{\boldsymbol{\beta}}_S^T, 0)^T$ lies on the line segment joining $\widehat{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}_0$, and we used the Cauchy-

Schwarz inequality that $\|\widehat{\boldsymbol{\beta}}_S - \boldsymbol{\beta}_{0S}\|_1 \leq \sqrt{s}\|\widehat{\boldsymbol{\beta}}_S - \boldsymbol{\beta}_{0S}\|$. From the assumption, there is a fixed neighborhood $\boldsymbol{\beta}_{0S} \in U \subset \mathcal{B}$ where the radius of U does not shrink as n increases, such that

$$\zeta \equiv \sup_{\beta_S \in U} \max_{l \notin S, j \in S} \left| \frac{\partial^2 L_n(\boldsymbol{\beta}_S, 0)}{\partial \beta_l \partial \beta_j} \right| = o_p\left(\frac{P_n'(0^+)}{\sqrt{s} k_n} \right). \tag{0.6}$$

Hence for any $\epsilon > 0$,

$$P\left(\max_{l \notin S, j \in S} \left| \frac{\partial^2 L_n(\tilde{\boldsymbol{\beta}})}{\partial \beta_l \partial \beta_j} \right| > \frac{\epsilon P_n'(0^+)}{\sqrt{s} k_n} \right) \leq P(\zeta > \frac{\epsilon P_n'(0^+)}{\sqrt{s} k_n}) + P(\widehat{\boldsymbol{\beta}}_S \notin U) = o(1).$$

This implies that $\max_{l \notin S, j \in S} |\partial^2 L_n(\tilde{\boldsymbol{\beta}})/\partial \beta_l \partial \beta_j| = o_p(P_n'(0^+)/(\sqrt{s} k_n))$, which yields that for ξ defined in (0.5), $\xi = o_p(\sqrt{s} k_n P_n'(0^+)/(k_n \sqrt{s})) = o_p(P_n'(0^+))$. This proves (0.4).

□

Verifying conditions in Theorems B.1 and B.2 for the smoothed FGMM

For simplicity, we focus on the linear model where $g(Y, \mathbf{X}^T\boldsymbol{\beta}) = Y - \mathbf{X}^T\boldsymbol{\beta}$. Generalization to a more general nonlinear model is straightforward.

Proof. First of all, we restrict on the oracle space. For any $\boldsymbol{\beta} = (\boldsymbol{\beta}_S, 0)$, $L_K(\boldsymbol{\beta}_S, 0)$ is given by

$$l(\boldsymbol{\beta}_S) = \sum_{j \in S} K\left(\frac{\beta_j^2}{h_n}\right) \left[\frac{1}{\widehat{\mathrm{var}}(V_j)} \left(\frac{1}{n} \sum_{i=1}^n (Y_i - \mathbf{X}_{Si}^T \boldsymbol{\beta}_S) F_{ij} \right)^2 + \frac{1}{\widehat{\mathrm{var}}(\tilde{V}_j)} \left(\frac{1}{n} \sum_{i=1}^n (Y_i - \mathbf{X}_{Si}^T \boldsymbol{\beta}_S) Z_{ij} \right)^2 \right]$$

We have $\nabla_S l(\boldsymbol{\beta}_{0S}) = \partial 1 + \partial 2$, where

$$\partial 1 = -\frac{1}{n} \sum_{i=1}^n \mathbf{X}_{iS} \mathbf{V}_{iS}^T \mathbf{J}_S \frac{1}{n} \sum_{i=1}^n \varepsilon_i \mathbf{F}_{Si} - \frac{1}{n} \sum_{i=1}^n \mathbf{X}_{iS} \mathbf{Z}_{iS}^T \widetilde{\mathbf{J}}_S \frac{1}{n} \sum_{i=1}^n \varepsilon_i \mathbf{Z}_{Si},$$

with $\mathbf{J}_S = \mathrm{diag}\{\widehat{\mathrm{var}}(F_{S,1})^{-1} K(\beta_{0S,1}^2/h_n), \cdots, \widehat{\mathrm{var}}(F_{S,s})^{-1} K(\beta_{0S,s}^2/h_n)\}$, and
$\widetilde{\mathbf{J}}_S = \mathrm{diag}\{\widehat{\mathrm{var}}(Z_{S,1})^{-1} K(\beta_{0S,1}^2/h_n), \cdots, \widehat{\mathrm{var}}(Z_{S,s})^{-1} K(\beta_{0S,s}^2/h_n)\}$. By assumption, $\|E\mathbf{X}_S \mathbf{F}_S^T\|_2 + \|E\mathbf{X}_S \mathbf{Z}_S^T\|_2 = O(1)$ and $E(\varepsilon|\mathbf{W}) = 0$. Therefore, $\partial 1 = O_p(\sqrt{s \log p/n})$.

$$\partial 2 = (K'(\beta_{0S,1}^2/h_n) 2\beta_{0S,1} r_1/h_n, \cdots, K'(\beta_{0S,s}^2/h_n) 2\beta_{0S,s} r_s/h_n)^T,$$

where $r_j = (\frac{1}{n}\sum_{i=1}^n \varepsilon_i F_{Sj,i})^2/\widehat{\mathrm{var}}(F_{Sj}) + (\frac{1}{n}\sum_{i=1}^n \varepsilon_i Z_{Sj,i})^2/\widehat{\mathrm{var}}(Z_{Sj})$. We have $\max_{j \leq p} |r_j| = O_p(s \log p/n)$. Also, for each $j \in S$, and $t_j = \beta_{0j}^2/h_n$,

$$\left| K'\left(\frac{\beta_{0S,j}^2}{h_n}\right) \frac{\beta_{0S,j}}{h_n} \right| \leq \frac{|K'(t_j) t_j|}{\min_{k \in S} |\beta_{0k}|} \leq O_p(\frac{e^{-t_j} t_j}{\min_{k \in S} |\beta_{0k}|}) = O_p(\frac{e^{-t_j/2}}{\min_{k \in S} |\beta_{0k}|})$$

For any small $\gamma > 0$, if $h_n^{1-\gamma} = o(\min_{k\in S}|\beta_{0k}|^2)$, then $\frac{\max_{j\in S} e^{-t_j/2}}{\min_{k\in S}|\beta_{0k}|} = O(\exp(-\frac{1}{4h_n^r})) = o(1)$. This implies $\partial 2 = O_p(\sqrt{s\log p/n})$.

For the Hessian matrix, $\nabla^2 l(\boldsymbol{\beta}_{0S}) = \boldsymbol{\Sigma} + \mathbf{G}_1 + \mathbf{G}_2 + \mathbf{G}_3 + \mathbf{G}_4$. We look at these terms one after another. $\mathbf{G}_1 = (K'(\beta_{0S,1}^2/h_n)2\beta_{0S,1}\boldsymbol{\alpha}_1/h_n, \cdots, K'(\beta_{0S,s}^2/h_n)2\beta_{0S,s}\boldsymbol{\alpha}_s/h_n)$, where

$$\boldsymbol{\alpha}_j = -\frac{1}{\widehat{\text{var}}(F_{S,j})}(\frac{2}{n}\sum_{i=1}^n \varepsilon_i F_{Sj,i})(\frac{1}{n}\sum_{i=1}^n F_{Sj,i}\mathbf{X}_{Si}) - \frac{1}{\widehat{\text{var}}(Z_{S,j})}(\frac{2}{n}\sum_{i=1}^n \varepsilon_i Z_{Sj,i})(\frac{1}{n}\sum_{i=1}^n Z_{Sj,i}\mathbf{X}_{Si}).$$

We have $\max_{j\leq S}\|\alpha_j\| = O_p(\sqrt{s\log p/n})$. Thus for $t_j = \beta_{0j}^2/h_n$, because $s\sqrt{\log p/n} = O(1)$,

$$\|\mathbf{G}_1\|_F = O_p(s\sqrt{\frac{\log p}{n}})\max_{j\in S}\frac{|K'(t_j)t_j|}{\min_{k\in S}|\beta_{0k}|} = O_p(s\sqrt{\frac{\log p}{n}}\exp(-\frac{1}{4h_n^r})) = o(1).$$

We also have $\mathbf{G}_2 = \text{diag}\{K''(\beta_{0S,1}^2/h_n)4\beta_{0S,1}^2 r_1/h_n^2, \cdots, K''(\beta_{0S,s}^2/h_n)4\beta_{0S,s}^2 r_s/h_n^2\}$. As before, $\max_{j\leq s}|r_j| = O_p(s\log p/n)$. Thus

$$\|\mathbf{G}_2\|_2 = O_p(\frac{\max_{j\in S}K''(t_j)t_j^2}{\min_{k\in S}|\beta_{0k}|^2}\frac{s\log p}{n}) = O_p(\frac{s\log p}{n}\frac{e^{-t_j}}{\min_{k\in S}|\beta_{0k}|^2}) = o(1),$$

the last equality follows since we have shown $\frac{e^{-t_j/2}}{\min_{k\in S}|\beta_{0k}|} = o(1)$. Moreover,

$$\mathbf{G}_3 = \text{diag}\left\{K'\left(\frac{\beta_{0S,1}^2}{h_n}\right)\frac{2r_1}{h_n}, \cdots, K'\left(\frac{\beta_{0S,s}^2}{h_n}\right)\frac{2r_s}{h_n}\right\}$$

We have, $\|\mathbf{G}_3\|_2 = O_p(\frac{\max_{j\in S} e^{-t_j/2}}{\min_{k\in S}|\beta_{0k}|}\frac{s\log p}{n\min_{k\in S}|\beta_{0k}|}) = o_p(1)$ since $s\log p/n = o(\min_{k\in s}|\beta_{0k}|)$.

$$\mathbf{G}_4 = \frac{2}{n}\sum_{i=1}^n \mathbf{X}_{Si}\mathbf{F}_{Si}^T\boldsymbol{\Delta}_1\frac{1}{n}\sum_{i=1}^n \mathbf{F}_{Si}\mathbf{X}_{Si}^T + \frac{2}{n}\sum_{i=1}^n \mathbf{X}_{Si}\mathbf{Z}_{Si}^T\boldsymbol{\Delta}_2\frac{1}{n}\sum_{i=1}^n \mathbf{Z}_{Si}\mathbf{X}_{Si}^T$$

where

$$\boldsymbol{\Delta}_1 = \text{diag}\left\{\frac{K(\beta_{0S,j}^2/h_n)-1}{\widehat{\text{var}}(F_{S,j})}\right\}_{j\leq s}, \quad \boldsymbol{\Delta}_2 = \text{diag}\left\{\frac{K(\beta_{0S,j}^2/h_n)-1}{\widehat{\text{var}}(Z_{S,j})}\right\}_{j\leq s}.$$

Note that $\max_{j\leq s}|K(\beta_{0S,j}^2/h_n)-1| = o(1)$, which gives $\|\mathbf{G}_4\|_2 = o_p(1)$. Finally,

$$\boldsymbol{\Sigma} = \frac{2}{n}\sum_{i=1}^n \mathbf{X}_{Si}\mathbf{F}_{Si}^T\boldsymbol{\Omega}_S\frac{1}{n}\sum_{i=1}^n \mathbf{F}_{Si}\mathbf{X}_{Si}^T + \frac{2}{n}\sum_{i=1}^n \mathbf{X}_{Si}\mathbf{Z}_{Si}^T\tilde{\boldsymbol{\Omega}}\frac{1}{n}\sum_{i=1}^n \mathbf{Z}_{Si}\mathbf{X}_{Si}^T,$$

where $\boldsymbol{\Omega} = \text{diag}\{\widehat{\text{var}}(F_{S,j})^{-1}\}_{j\leq s}$ and $\tilde{\boldsymbol{\Omega}} = \text{diag}\{\widehat{\text{var}}(Z_{S,j})^{-1}\}_{j\leq s}$. Therefore all the eigenvalues of $\mathbf{\Sigma}$ are bounded way from zero. This implies $\lambda_{\min}(\nabla^2 l(\boldsymbol{\beta}_{0S}))$ is bounded away from zero. Thus we have verified all the conditions in Theorem B.1 of the main paper.

To verify the conditions of Theorem B.2, it suffices to verify those in Lemma 0.1 above. More concretely, we verify

$$\max_{l\notin S}\left|\frac{\partial L_K(\boldsymbol{\beta}_0)}{\partial \beta_l}\right| = o_p(P_n(0^+)), \tag{0.7}$$

and there is a neighborhood U of $\boldsymbol{\beta}_{0S}$, such that

$$\sup_{\beta_S\in U}\max_{l\notin S,k\in S}\left|\frac{\partial^2 L_K(\boldsymbol{\beta}_S,0)}{\partial\beta_l\partial\beta_k}\right| = o_p\left(P_n'(0^+)\left(\frac{n}{s^2\log s}\right)^{1/2}\right). \tag{0.8}$$

For (0.7), note that $\beta_{0l}=0$ when $l\notin S$. Thus we have

$$\frac{\partial L_K(\boldsymbol{\beta}_0)}{\partial\beta_l} = -\sum_{j\in S}K\left(\frac{\beta_{0j}^2}{h_n}\right)\left[\frac{2}{\mathrm{var}(V_j)}\left(\frac{1}{n}\sum_{i=1}^n\varepsilon_i F_{ij}\right)\left(\frac{1}{n}\sum_{i=1}^n X_{il}F_{ij}\right)\right.$$

$$\left.+\frac{2}{\mathrm{var}(\tilde{V}_j)}\left(\frac{1}{n}\sum_{i=1}^n\varepsilon_i Z_{ij}\right)\left(\frac{1}{n}\sum_{i=1}^n X_{il}Z_{ij}\right)\right]$$

Because $s\sqrt{\log p/n}\max_{l\notin S,j\in S}|EX_lV_j| = o(P_n'(0^+))$,

$$\max_{l\notin S}\left|\frac{\partial L_K(\boldsymbol{\beta}_0)}{\partial\beta_l}\right| \le \sum_{j\in S}\sqrt{\frac{\log s}{n}}O_p(|EX_lV_j|+|EX_l\tilde{V}_j|)+\frac{s(\log p)^2}{n} = o_p(P_n'(0^+)).$$

As for (0.8), we know that for $k\notin S$, $\partial^2 L_K(\boldsymbol{\beta}_S,0)/\partial\beta_l\partial\beta_k = a_{1,lk}+a_{2,lk}+a_{3,lk}$ where

$$a_{1,lk} = -K'\left(\frac{\beta_{Sk}^2}{h_n}\right)\frac{2\beta_{Sk}}{h_n}\left[\frac{2}{\widehat{\mathrm{var}}(F_l)}\left(\frac{1}{n}\sum_{i=1}^n(Y_i-\mathbf{X}_{Si}^T\boldsymbol{\beta}_S)F_{il}\right)\left(\frac{1}{n}\sum_{i=1}^n X_{ik}F_{il}\right)\right.$$

$$\left.+\frac{2}{\widehat{\mathrm{var}}(Z_l)}\left(\frac{1}{n}\sum_{i=1}^n(Y_i-\mathbf{X}_{Si}^T\boldsymbol{\beta}_S Z_{il}\right)\left(\frac{1}{n}\sum_{i=1}^n X_{ik}Z_{il}\right)\right]$$

There is a neighborhood U of $\boldsymbol{\beta}_{0S}$ that doest not shrink with n, such that uniformly in $l,k\le p$, $\sup_{\beta_S\in U}|a_{1,lk}|$ is $O_p(1)$. In addition, $a_{3,lk}=a_{1,kl}$. Finally,

$$a_{2,lk} = \frac{2}{n}\sum_{i=1}^n X_{ik}\mathbf{F}_{Si}^T\mathbf{J}_1(\boldsymbol{\beta}_S)\frac{1}{n}\sum_{i=1}^n X_{il}\mathbf{F}_{Si}+\frac{2}{n}\sum_{i=1}^n X_{ik}\mathbf{Z}_{Si}^T\mathbf{J}_2(\boldsymbol{\beta}_S)\frac{1}{n}\sum_{i=1}^n X_{il}\mathbf{Z}_{Si}$$

where $\mathbf{J}_1(\boldsymbol{\beta}_S) = \mathrm{diag}\{K(\beta_{S,j}^2/h_n)/\widehat{\mathrm{var}}(F_{S,j})\}_{j\le s}$, and $\mathbf{J}_2(\boldsymbol{\beta}_S) = \mathrm{diag}\{K(\beta_{S,j}^2/h_n)/\widehat{\mathrm{var}}(Z_{S,j})\}_{j\leq s}$. Therefore,

$$\sup_{\beta_S\in U}\max_{l,k}|a_{2,lk}| \leq O_p(1)\max_{k\leq p}\|\frac{1}{n}\sum_{i=1}^{n}X_{ik}\mathbf{F}_{Si}\|^2 = o_p(1) + O_p(1)\max_{k\leq p}\|EX_k\mathbf{F}_S\|^2$$

$$= o_p(1) + O_p(1)\max_{k\leq p}\lambda_{\max}[E(X_k\mathbf{F}_S)E(X_k\mathbf{F}_S^T)] \leq o_p(1) + O_p(1)\max_{k\leq p}\lambda_{\max}[E(X_k^2\mathbf{F}_S\mathbf{F}_S^T)],$$

where we used the fact that $\lambda_{\max}(E\mathbf{A}E\mathbf{A}^T) \leq \lambda_{\max}(E\mathbf{A}\mathbf{A}^T)$ for any random vector $\mathbf{A}$ that has finite expectation. This implies (0.8).

Proof of Theorem 7.1

We now can apply Theorem B.2 to conclude that there is a local minimizer $\widehat{\boldsymbol{\beta}}' = (\widehat{\boldsymbol{\beta}}'_S, \widehat{\boldsymbol{\beta}}'_N)$ of $Q_K(\cdot)$ such that $P(\widehat{\boldsymbol{\beta}}'_N = 0) \to 1$. In addition, for an arbitrarily small $\epsilon > 0$, the local minimizer $\widehat{\boldsymbol{\beta}}'$ is strict with probability at least $1-\epsilon$ for all large n. The remainder of the proof is the same as that of Theorem 4.1 (iii).

□