Supplementary Materials for
Linear or Nonlinear? Automatic Structure Discovery for Partially Linear Models

Hao Helen Zhang*, Guang Cheng and Yufeng Liu

The supplementary material contains the proof of Theorem 2.

Appendix 4. Proof of Theorem 2

It suffices to show that, with probability tending to one,

$$\mathcal{P}_{1j}\widehat{g} = 0 \Longleftrightarrow \mathcal{P}_{1j}g_0 = 0, \tag{8.18}$$

$$\widehat{\beta}_j = 0 \Longleftrightarrow \beta_j^0 = 0 \tag{8.19}$$

for $j = 1, \ldots, d$. Without loss of generality, we focus on the case $d = 2$, i.e. $g(x_1, x_2) = b + \beta_1 k_1(x_1) + \beta_2 k_1(x_2) + g_{11}(x_1) + g_{12}(x_2)$, where $g_{1j}(x_j) \in \mathcal{S}_{per,j}$, in the proof. Note that in this case the sample size n is m^2 since we assume $n_1 = n_2 = m$. We have three major steps in the proof.

Step I: Formulation

Let $\mathbf{\Sigma} = \{R_1(x_{i,1}, x_{k,1})\}_{i,k=1}^m$ be the $m \times m$ marginal Gram matrix corresponding to the reproducing kernel for $\mathcal{S}_{per}$. Let $\mathbf{1}_m$ be a vector of m ones. Assuming the observations are permuted appropriately, we can write the $n \times n$ Gram matrix $\mathbf{R}_{11} = \mathbf{\Sigma} \odot (\mathbf{1}_m \mathbf{1}_m')$ and $\mathbf{R}_{12} = (\mathbf{1}_m \mathbf{1}_m') \odot \mathbf{\Sigma}$, where $\odot$ stands for the Kronecker product between two matrices. Let $\{\boldsymbol{\xi}_1 = \mathbf{1}_m, \boldsymbol{\xi}_2, \ldots, \boldsymbol{\xi}_m\}$ be an orthonormal (with respect to the inner product $< \cdot >_m$ in $\mathcal{R}^m$) eigensystem of $\mathbf{\Sigma}$ with corresponding eigenvalues $m\eta_1, \ldots, m\eta_m$ where $\eta_1 = (720m^4)^{-1}$. Thus, we have

$$< \boldsymbol{\xi}_1, \boldsymbol{\xi}_j >_m = 0 \Longrightarrow \frac{1}{m}\sum_{i=1}^m \xi_{ij} = 0 \ \text{ for } j \geq 2,$$

$$< \boldsymbol{\xi}_j, \boldsymbol{\xi}_j >_m = 1 \Longrightarrow \frac{1}{m}\sum_{i=1}^m \xi_{ij}^2 = 1 \ \text{ for } j \geq 1.$$

From Utreras (1983), we know that $\eta_2 \geq \eta_3 \geq \ldots \geq \eta_m$ and $\eta_i \sim i^{-4}$ for $i \geq 2$.

Let Υ be the $m \times m$ matrix with $\{\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_m\}$ as its columns. We then define a $n \times n$ matrix $\mathbf{O} = \Upsilon \odot \Upsilon$. It is easy to verify that the columns of $\mathbf{O}$, i.e. $\{\widetilde{\boldsymbol{\xi}}_i : i = 1, 2, \ldots, n\}$, form an eigensystem for each of $\mathbf{R}_{11}$ and $\mathbf{R}_{12}$. We next rearrange the columns of $\mathbf{O}$ to form $\{\boldsymbol{\zeta}_{1j}, \ldots, \boldsymbol{\zeta}_{nj}\}$ so that their first m elements are those corresponding to nonzero eigenvalues for $\mathbf{R}_{1j}$ and the rest $(n-m)$ elements are given by the remaining $\widetilde{\boldsymbol{\xi}}_i$ for $j = 1, 2$. The corresponding eigenvalues are then $\eta_{ij} = n\eta_i$ for $i = 1, \ldots, m$ and zero otherwise. It is clear that $\{\widetilde{\boldsymbol{\xi}}_1, \ldots, \widetilde{\boldsymbol{\xi}}_n\}$ is also an orthonormal basis in $\mathbb{R}^n$ with respect to the inner product $< \mathbf{u}, \mathbf{v} >_n$. Thus we have $\mathbf{O}'\mathbf{O} = n\mathbf{I}$ and $\mathbf{O}\mathbf{O}' = n\mathbf{I}$.

Recall that our estimate $(\widehat{\boldsymbol{\beta}}, \widehat{g}_{11}, \widehat{g}_{12})$ is obtained by minimizing

$$\frac{1}{n}\left(\mathbf{y} - \mathbf{T}\boldsymbol{\beta} - \mathbf{R}_{\mathbf{w}_1,\boldsymbol{\theta}}\mathbf{c}\right)'\left(\mathbf{y} - \mathbf{T}\boldsymbol{\beta} - \mathbf{R}_{\mathbf{w}_1,\boldsymbol{\theta}}\mathbf{c}\right) + \lambda_1 \sum_{j=1}^{d} w_{0j}|\beta_j| + \tau_0 \mathbf{c}'\mathbf{R}_{\mathbf{w}_1,\boldsymbol{\theta}}\mathbf{c} + \tau_1 \sum_{j=1}^{d} w_{1j}\theta_j, \tag{8.20}$$

over $(\boldsymbol{\beta}, \mathbf{c}, \boldsymbol{\theta})$, see (5.3). For simplicity, we hold $\tau_0 = 1$. By using the special construction of $\mathbf{O}$, i.e. $\mathbf{O}\mathbf{O}' = n\mathbf{I}$, we can rewrite (8.20) as

$$\left(\mathbf{z} - \mathbf{O}'\mathbf{T}\boldsymbol{\beta}/n - \mathbf{D}_\theta \mathbf{s}\right)'\left(\mathbf{z} - \mathbf{O}'\mathbf{T}\boldsymbol{\beta}/n - \mathbf{D}_\theta \mathbf{s}\right) + \lambda_1 \sum_{j=1}^{d} w_{0j}|\beta_j| + \mathbf{s}'\mathbf{D}_\theta \mathbf{s} + \tau_1 \sum_{j=1}^{d} w_{1j}\theta_j, \tag{8.21}$$

where $\mathbf{z} = (1/n)\mathbf{O}'\mathbf{y}$, $\mathbf{s} = \mathbf{O}'\mathbf{c}$, $\mathbf{D}_\theta = \sum_{j=1}^{d} \theta_j w_{1j}^{-1}\mathbf{D}_j$ and $\mathbf{D}_j = (1/n^2)\mathbf{O}'\mathbf{R}_{1j}\mathbf{O}$ is a diagonal $n \times n$ matrix with diagonal elements η_{ij}. We further write $\mathbf{O}'\mathbf{T}\boldsymbol{\beta}/n = (b, 0, 0, \ldots)' + \mathbf{O}'\boldsymbol{t}_1\beta_1/n + \mathbf{O}'\boldsymbol{t}_2\beta_2/n$, where $\mathbf{T} = (\mathbf{1}_n, \boldsymbol{t}_1, \boldsymbol{t}_2)$ and

$$\boldsymbol{t}_1 = (1/m - 1/2, 2/m - 1/2, \ldots, 1 - 1/2)' \otimes \mathbf{1}_m, \tag{8.22}$$

$$\boldsymbol{t}_2 = \mathbf{1}_m \otimes (1/m - 1/2, 2/m - 1/2, \ldots, 1 - 1/2)'. \tag{8.23}$$

Due to the orthogonality of basis $\{\boldsymbol{\zeta}_{1j}, \ldots, \boldsymbol{\zeta}_{nj}\}$ for any j, we can further write (8.21) as

$$L(\mathbf{s}, \boldsymbol{\beta}, \boldsymbol{\theta}) = (z_{11} - b - t_{1,11}\beta_1 - t_{2,11}\beta_2 - \theta_1\eta_{11}s_{11})^2 + \left(\sum_{i=2}^{m}\sum_{j=1}^{d} + \sum_{i=1}^{1}\sum_{j=2}^{d}\right)$$
$$(z_{ij} - t_{1,ij}\beta_1 - t_{2,ij}\beta_2 - \theta_j\eta_{ij}s_{ij})^2 + \sum_{i=1}^{m}\sum_{j=1}^{d} \eta_{ij}\theta_j w_{1j}^{-1} s_{ij}^2 + \lambda_1 \sum_{j=1}^{d} w_{0j}|\beta_j| + \tau_1 \sum_{j=1}^{d} w_{1j}\theta_j, \tag{8.24}$$

where $t_{1,ij} = \boldsymbol{\zeta}_{ij}'\boldsymbol{t}_1/n$, $t_{2,ij} = \boldsymbol{\zeta}_{ij}'\boldsymbol{t}_2/n$, $z_{ij} = \boldsymbol{\zeta}_{ij}'\mathbf{y}/n$ and $s_{ij} = \boldsymbol{\zeta}_{ij}'\mathbf{c}$.

Note that our estimate $(\widehat{\boldsymbol{\beta}}, \widehat{g}_{11}, \widehat{g}_{12})$ are related to the minimizer of (8.24), denoted by $(\widehat{\boldsymbol{\beta}}, \widehat{\mathbf{s}}, \widehat{\boldsymbol{\theta}})$, as shown in (5.2). Thus, we first analyze $(\widehat{\boldsymbol{\beta}}, \widehat{\mathbf{s}}, \widehat{\boldsymbol{\theta}})$. Straightforward calculation shows that $\widehat{s}_{11} = 0$ and $z_{11} - \widehat{b} - t_{1,11}\widehat{\beta}_1 - t_{2,11}\widehat{\beta}_2 = 0$. Thus, we only need to consider minimizing

$$\begin{aligned} L_1(\mathbf{s}, \beta_1, \beta_2, \boldsymbol{\theta}) = & \left(\sum_{i=2}^{m}\sum_{j=1}^{d} + \sum_{i=1}^{1}\sum_{j=2}^{d} \right) \left[\left(z_{ij} - t_{1,ij}\beta_1 - t_{2,ij}\beta_2 - \theta_j w_{1j}^{-1}\eta_{ij}s_{ij} \right)^2 + \eta_{ij}\theta_j s_{ij}^2 \right] \\ & + \lambda_1 \sum_{j=1}^{d} w_{0j}|\beta_j| + \tau_1 \sum_{j=1}^{d} w_{1j}\theta_j, \end{aligned} \tag{8.25}$$

We minimize $L_1(\mathbf{s}, \beta_1, \beta_2, \boldsymbol{\theta})$ in two steps. Given fixed $(\beta_1, \beta_2, \boldsymbol{\theta})$, we first minimize L_1 over $\mathbf{s}$. Since L_1 is a convex function in $\mathbf{s}$, we can obtain the minimizer

$$\widehat{s}_{ij}(\beta_1, \beta_2, \boldsymbol{\theta}) = \frac{z_{ij} - t_{1,ij}\beta_1 - t_{2,ij}\beta_2}{1 + \theta_j \eta_{ij}/w_{1j}}. \tag{8.26}$$

Plugging (8.26) into (8.25), we obtain $L_1(\widehat{\mathbf{s}}(\beta_1, \beta_2, \boldsymbol{\theta}), \beta_1, \beta_2, \boldsymbol{\theta})$, denoted as $L_2(\beta_1, \beta_2, \boldsymbol{\theta})$:

$$\begin{aligned} L_2(\beta_1, \beta_2, \boldsymbol{\theta}) = & \left(\sum_{i=2}^{m}\sum_{j=1}^{d} + \sum_{i=1}^{1}\sum_{j=2}^{d} \right) \left[\frac{(z_{ij} - t_{1,ij}\beta_1 - t_{2,ij}\beta_2)^2}{(1 + \theta_j \eta_{ij}/w_{1j})} \right] + \lambda_1 \sum_{j=1}^{d} w_{0j}|\beta_j| \\ & + \tau_1 \sum_{j=1}^{d} w_{1j}\theta_j \end{aligned} \tag{8.27}$$

Step 2: Prove $\mathcal{P}_{1j}\widehat{g} = 0 \Longleftrightarrow \mathcal{P}_{1j}g_0 = 0$

In this step we consider selection consistency for $\mathcal{P}_{1j}g$. We first verify that $L_2(\beta_1, \beta_2, \boldsymbol{\theta})$ in (8.27) is convex in $\boldsymbol{\theta}$ for any fixed values of β_1 and β_2 by obtaining that

$$\begin{aligned} \frac{\partial^2 L_2(\beta_1, \beta_2, \boldsymbol{\theta})}{\partial \theta_j^2} &= 2 \left(\sum_{i=2}^{m}\sum_{j=1}^{d} + \sum_{i=1}^{1}\sum_{j=2}^{d} \right) \left[\frac{\eta_{ij}^2 (z_{ij} - t_{1,ij}\beta_1 - t_{2,ij}\beta_2)^2}{(1 + \theta_j \eta_{ij}/w_{1j})^3} \right] > 0, \\ \frac{\partial^2 L_2(\beta_1, \beta_2, \boldsymbol{\theta})}{\partial \theta_j \theta_k} &= 0 \ \text{ for } j \neq k. \end{aligned}$$

By the above convexity, we know $\widehat{\theta}_j = 0$ if and only if

$$\left(\frac{\partial}{\partial \theta_j} |_{\theta_j = 0} \right) L_2(\widehat{\beta}_1, \widehat{\beta}_2, \boldsymbol{\theta}) \geq 0,$$

which is equivalent to

$$U_1 \equiv \sum_{i=2}^{m} \eta_{i1}(z_{i1} - t_{1,i1}\widehat{\beta}_1 - t_{2,i1}\widehat{\beta}_2)^2 \leq \tau_1 w_{11}^2, \tag{8.28}$$

$$U_j \equiv \sum_{i=1}^{m} \eta_{ij}(z_{ij} - t_{1,ij}\widehat{\beta}_1 - t_{2,ij}\widehat{\beta}_2)^2 \leq \tau_1 w_{1j}^2 \quad \text{for } j \geq 2. \tag{8.29}$$

We define $a_{ij} = \boldsymbol{\zeta}_{ij}'\mathbf{G}_1/n$, where $\mathbf{G}_1 = (G_1(\mathbf{x}_1), \ldots, G_1(\mathbf{x}_n))'$ and $G_1(\mathbf{x}_i) = \sum_{j=1}^{d} g_{1j}^0(x_{ij})$. Combining the fact that $z_{ij} = \boldsymbol{\zeta}_{ij}'\mathbf{y}/n$, we have the following equation:

$$z_{ij} - t_{1,ij}\beta_1^0 - t_{2,ij}\beta_2^0 = a_{ij} + e_{ij}, \tag{8.30}$$

where $e_{ij} \overset{i.i.d.}{\sim} N(0, \sigma^2/n)$ for $1 \le i \le m$ and $1 \le j \le d$. Thus, (8.28) and (8.29) become

$$U_1 = \sum_{i=2}^{m} \eta_{i1}\left(t_{1,i1}(\beta_1^0 - \widehat{\beta}_1) + t_{2,i1}(\beta_2^0 - \widehat{\beta}_2) + e_{i1} + a_{i1}\right)^2, \tag{8.31}$$

$$U_j = \sum_{i=1}^{m} \eta_{ij}\left(t_{1,ij}(\beta_1^0 - \widehat{\beta}_1) + t_{2,ij}(\beta_2^0 - \widehat{\beta}_2) + e_{ij} + a_{ij}\right)^2 \tag{8.32}$$

by considering (8.30).

In the below, without loss of generality, we assume that $g_{12}^0(x_{i2}) = 0$ for $i = 1, \ldots, n$. We first show "$\mathcal{P}_{12}g_0 = 0 \Longrightarrow \mathcal{P}_{12}\widehat{g} = 0$". To show $\mathcal{P}_{12}\widehat{g} = 0$, it suffices to show

$$P(U_2 > \tau_1 w_{12}^2) \to 0. \tag{8.33}$$

based on the above analysis and (5.2). Note that $\mathcal{P}_{12}g_0 = 0$ implies $a_{i2} = 0$ for all $1 \le i \le m$. Thus, we have

$$\begin{aligned}
P(U_2 > \tau_1 w_{12}) &= P\left(\sum_{i=1}^{m} \eta_{i2}\left(t_{1,i2}(\beta_1^0 - \widehat{\beta}_1) + t_{2,i2}(\beta_2^0 - \widehat{\beta}_2) + e_{i2}\right)^2 > \tau_1 w_{12}^2\right) \\
&\le P\left(\sum_{i=1}^{m} \eta_{i2}\left[t_{1,i2}^2(\widehat{\beta}_1 - \beta_1^0)^2 + t_{2,i2}^2(\widehat{\beta}_2 - \beta_2^0)^2 + e_{i2}^2\right] > \tau_1 w_{12}^2/3\right) \\
&\le \sum_{k=1}^{2} P\left(\sum_{i=1}^{m} \eta_{i2} t_{k,i2}^2(\widehat{\beta}_k - \beta_k^0)^2 > \tau_1 w_{12}^2/9\right) + P\left(\sum_{i=1}^{m} \eta_{i2} e_{i2}^2 > \tau_1 w_{12}^2/9\right). \quad (8.34)
\end{aligned}$$

The first inequality in the above follows from the Cauchy-Schwarz inequality. For $k = 1, 2$, we have

$$\begin{aligned}
\sum_{i=1}^{m} \eta_{i2} t_{k,i2}^2 &\le \sqrt{\sum_{i=1}^{m} \eta_{i2}^2}\sqrt{\sum_{i=1}^{m} t_{k,i2}^4} \le \sqrt{\sum_{i=1}^{m} (n\eta_i)^2}\sqrt{\sum_{i=1}^{m} (\boldsymbol{\zeta}_{i2}'\boldsymbol{t}_k/n)^4} \\
&\le n^{-1} \times \sqrt{\sum_{i=1}^{m} \|\boldsymbol{\zeta}_{i2}\|^4 \|\boldsymbol{t}_k\|^4} \\
&\le n^{-1} \times O(n^{5/4}) = O(n^{1/4}) \qquad (8.35)
\end{aligned}$$

by considering $\eta_1 = (720m^4)^{-1}$, $\eta_i \sim i^{-4}$ for $i = 2, \ldots, m$, and Holder's inequality. By adapting the arguments in Lemma 8.1, we can show

$$\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| = O_P(n^{-1/5}). \tag{8.36}$$

Now we focus on the first two probabilities in (8.34). Combining (8.35), (8.36) and the condition that $n^{3/20}\tau_1 w_{12}^2 \to \infty$, we can show that they converge to zero. Let $V_2 = \sum_{i=1}^m \eta_{i2} e_{i2}^2$. Since e_{i2} follows $N(0, \sigma^2/n)$ as discussed above, we have

$$E(nV_2) \sim \sigma^2 \text{ and } Var(nV_2) \sim \sigma^4. \tag{8.37}$$

As for the third probability in (8.34), we have

$$\begin{aligned} P(V_2 > \tau_1 w_{12}^2/9) &\leq P\left(|nV_2 - EnV_2| > n\tau_1 w_{12}^2/9 - EnV_2\right) \\ &\leq \frac{Var(nV_2)}{(n\tau_1 w_{12}^2/9 - EnV_2)^2} \to 0 \end{aligned}$$

where the second inequality follows from the Chebyshev's inequality and the condition that $n\tau_1 w_{12}^2 \to \infty$. This completes the proof of (8.33), thus shows "$\mathcal{P}_{12} g_0 = 0 \Longrightarrow \mathcal{P}_{12}\widehat{g} = 0$".

Next we prove "$\mathcal{P}_{12}\widehat{g} = 0 \Longrightarrow \mathcal{P}_{12} g_0 = 0$" by showing the equivalent statement "$\mathcal{P}_{12} g_0 \neq 0 \Longrightarrow \mathcal{P}_{12}\widehat{g} \neq 0$". To show $\mathcal{P}_{12}\widehat{g} \neq 0$, it suffices to show

$$P(U_2 \leq \tau_1 w_{12}^2) \to 0 \tag{8.38}$$

based on the previous discussions. We first establish the following inequalities:

$$\begin{aligned} P(U_2 \leq \tau_1 w_{12}^2) &\leq P(|U_2 - EW_2| \geq EW_2 - \tau_1 w_{12}^2) \\ &\leq P(|U_2 - W_2| \geq (EW_2 - \tau_1 w_{12}^2)/2) + P(|W_2 - EW_2| \geq (EW_2 - \tau_1 w_{12}^2)/2) \\ &\leq I + II, \end{aligned}$$

where $W_2 = \sum_{i=1}^m \eta_{i2}(e_{i2} + a_{i2})^2$. By the Cauchy-Schwartz inequality, we have

$$|U_2 - W_2| \leq 4W_2 + 3\sum_{k=1}^2 \sum_{i=1}^m \eta_{i2} t_{k,i2}^2 (\widehat{\beta}_k - \beta_k^0)^2.$$

Thus, the term I can be further bounded by

$$\begin{aligned} I &\leq P(W_2 \geq (EW_2 - \tau_1 w_{12}^2)/16) + \sum_{k=1}^2 P\left(\sum_{i=1}^m \eta_{i2} t_{k,i2}^2 (\widehat{\beta}_k - \beta_k^0)^2 \geq (EW_2 - \tau_1 w_{12}^2)/24\right) \\ &\leq I_1 + I_2. \end{aligned}$$

To analyze the order of I_1, I_2 and II, we need to study the order of EW_2 and $VarW_2$. Note that $\mathcal{P}_{12}g_0 \neq 0$ implies $a_{i_0 2} \neq 0$ for some $1 \leq i_0 \leq m$. Thus, we have

$$E(W_2) \geq E(\eta_{i_0 2}(e_{i_0 2} + a_{i_0 2})^2) \geq \eta_{i_0 2} a_{i_0 2}^2, \tag{8.39}$$

$$\begin{aligned} Var(W_2) &= \sum_{i=1}^m \eta_{i2}^2 Var((e_{i2} + a_{i2})^2) = \sum_{i=1}^m \eta_{i2}^2 (4n^{-1} a_{i2}^2 \sigma^2 + 2n^{-2}\sigma^4) \\ &\leq 4n^{-1}\sigma^2 \sum_{i=1}^m a_{i2}^2 + 2n^{-2}\sigma^4 \leq 4n^{-1}\sigma^2 \|\mathcal{P}_{12}g_0\|_2 + 2n^{-2}\sigma^2 = O(n^{-1}) \end{aligned} \tag{8.40}$$

By (8.39) and Lemma 8.1, we know $(EW_2 - \tau_1 w_{12}^2)$ is bounded away from zero. Then, by Chebyshev's inequality, we have

$$II \lesssim \frac{Var(W_2)}{(EW_2 - \tau_1 w_{12}^2)^2} \to 0$$

by (8.40). As for the term I_2, we can also show it converges to zero by considering (8.35) and (8.36). For the term I_1, we have

$$I_2 = P(16(W_2 - EW_2) \geq -\tau_1 w_{12}^2 - 15EW_2) \lesssim \frac{Var(W_2)}{(\tau_1 w_{12}^2 + 15EW_2)^2} \to 0$$

since $(EW_2 + \tau_1 w_{12}^2)$ is bounded away from zero and $Var(W_2) = O(n^{-1})$.

Step 3: Prove $\widehat{\beta}_j = 0 \Longleftrightarrow \beta_j^0 = 0$

In this step we consider selection consistency for β_j. Without loss of generality, we assume that $\beta_2^0 = 0$. First, we rewrite (8.27) as $Q(\beta_1, \beta_2, \boldsymbol{\theta}) + \lambda_1 \sum_{j=1}^d w_{0j}|\beta_j| + \tau_1 \sum_{j=1}^d w_{1j}\theta_j$. Applying the Taylor expansion to (8.27), we have

$$\begin{aligned} \frac{\partial L_2(\beta_1, \beta_2, \widehat{\boldsymbol{\theta}})}{\partial \beta_2} &= \frac{\partial Q(\beta_1, \beta_2, \widehat{\boldsymbol{\theta}})}{\partial \beta_2} + \lambda_1 w_{02}\text{sign}(\beta_2) \\ &= \frac{\partial Q(\beta_1^0, \beta_2^0, \widehat{\boldsymbol{\theta}})}{\partial \beta_2} + \frac{\partial^2 Q(\beta_1^0, \beta_2^0, \widehat{\boldsymbol{\theta}})}{\partial \beta_1 \partial \beta_2}(\beta_1 - \beta_1^0) + \frac{\partial^2 Q(\beta_1^0, \beta_2^0, \widehat{\boldsymbol{\theta}})}{\partial \beta_2^2}(\beta_2 - \beta_2^0) \\ &\quad + \lambda_1 w_{02}\text{sign}(\beta_2). \end{aligned} \tag{8.41}$$

Recall that $\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| = O_P(n^{-1/5})$ by (8.36). Thus, in the below, we only consider β_1 and β_2 satisfying $|\beta_1 - \beta_1^0| = O_P(n^{-1/5})$ and $|\beta_2 - \beta_2^0| = O_P(n^{-1/5})$.

By (8.30), the first term in (8.41) can be written as

$$\begin{aligned} &-2\left(\sum_{i=2}^m \sum_{j=1}^d + \sum_{i=1}^1 \sum_{j=2}^d\right)\left[\frac{(a_{ij} + e_{ij})t_{2,ij}}{1 + \widehat{\theta}_j \eta_{ij}/w_{1j}}\right] \\ &= -2\left(\sum_{i=2}^m \sum_{j=1}^d + \sum_{i=1}^1 \sum_{j=2}^d\right)\left[\frac{\mathbf{G}_1' \boldsymbol{\zeta}_{ij}\boldsymbol{\zeta}_{ij}' \boldsymbol{t}_2 + \boldsymbol{\epsilon}' \boldsymbol{\zeta}_{ij}\boldsymbol{\zeta}_{ij}' \boldsymbol{t}_2}{n^2(1 + \widehat{\theta}_j \eta_{ij}/w_{1j})}\right] \\ &= O_P(n^{-1/2}), \end{aligned} \tag{8.42}$$

where the last equality follows from the orthogonality of the constructed $\{\boldsymbol{\zeta}_{ij}\}$ and Lindeberger-Feller theorem. As for the second term of (8.41), we have

$$
\begin{aligned}
\frac{\partial^2 Q(\beta_1^0,\beta_2^0,\widehat{\boldsymbol{\theta}})}{\partial\beta_1\partial\beta_2}(\beta_1-\beta_1^0) &= 2\left(\left[\sum_{i=2}^{m}\sum_{j=1}^{d}+\sum_{i=1}^{1}\sum_{j=2}^{d}\right]\frac{t_{1,ij}t_{2,ij}}{1+\widehat{\theta}_j\eta_{ij}/w_{1j}}(\beta_1-\beta_1^0)\right)\\
&= 2\left(\left[\sum_{i=2}^{m}\sum_{j=1}^{d}+\sum_{i=1}^{1}\sum_{j=2}^{d}\right]\frac{\boldsymbol{t}_1'\boldsymbol{\zeta}_{ij}\boldsymbol{\zeta}_{ij}'\boldsymbol{t}_2}{n^2(1+\widehat{\theta}_j\eta_{ij}/w_{1j})}(\beta_1-\beta_1^0)\right)\\
&\leq O(n^{-1})O_P(n^{-1/5}) = O_P(n^{-6/5}),
\end{aligned}
$$

where the last inequality follows from the orthogonality of the constructed $\{\boldsymbol{\zeta}_{ij}\}$ and the forms of $\boldsymbol{t}_1$ and $\boldsymbol{t}_2$, i.e. (8.22) and (8.23). By applying similar analysis to the third term in (8.41), we know its order is also $O_P(n^{-1/5})$. In summary, we have

$$
\frac{\partial L_2(\beta_1,\beta_2,\widehat{\boldsymbol{\theta}})}{\partial\beta_2} = O_P(n^{-1/5}) + \lambda_1 w_{02}\mathrm{sign}(\beta_2). \tag{8.43}
$$

We first show "$\beta_2^0 = 0 \Longrightarrow \widehat{\beta}_2 = 0$". If $\beta_2^0 = 0$, then the range of β_2 in (8.43) is $(-Cn^{-1/5}, Cn^{-1/5})$ for some $C > 0$. By the assumed condition that $n^{1/5}\lambda_1 w_{02} \to \infty$, we can conclude that $\partial L_2(\beta_1,\beta_2,\widehat{\boldsymbol{\theta}})/\partial\beta_2 < 0$ for $\beta_2 \in (-Cn^{-1/5}, 0)$ and $\partial L_2(\beta_1,\beta_2,\widehat{\boldsymbol{\theta}})/\partial\beta_2 > 0$ for $\beta_2 \in (0, Cn^{-1/5})$. In other words, we have

$$
L_2(\beta_1, 0, \widehat{\boldsymbol{\theta}}) = \min_{|\beta_2|\leq Cn^{-1/5}} L_2(\beta_1,\beta_2,\widehat{\boldsymbol{\theta}}) \text{ with probability tending to one,}
$$

which implies $\widehat{\beta}_2 = 0$. We next show "$\widehat{\beta}_2 = 0 \Longrightarrow \beta_2^0 = 0$" by showing the equivalent statement that "$\beta_2^0 \neq 0 \Longrightarrow \widehat{\beta}_2 \neq 0$". For simplicity, we assume $\beta_2^0 = 1$ which means that $\beta_2 \in (1 - Cn^{-1/5}, 1 + Cn^{-1/5})$. Then, by considering the condition $n^{1/5}\lambda_1 w_{02} \to \infty$ in (8.43), we have $\partial L_2(\beta_1,\beta_2,\widehat{\boldsymbol{\theta}})/\partial\beta_2 > 0$ which implies that $\widehat{\beta}_2 > 0$. This completes the whole proof. □