A STRATIFIED PENALIZATION METHOD FOR SEMIPARAMETRIC VARIABLE LABELING OF MULTI-OUTPUT TIME-VARYING COEFFICIENT MODELS

Ting Zhang*, Weiliang Wang† and Yu Shao†

*University of Georgia and †Boston University

Supplementary Material

S1 An Iterative Algorithm

Here, we describe an iterative algorithm that can be used to compute the stratified penalized local linear estimator proposed in Section 3.1. For this, let $\Theta^{(0)} = \{\theta_{l,k,j}^{(0)}(t)\}_{l,k,j}$, for $t \in [0,1]$, denote an initial value, then by the local quadratic penalty approximation of Fan and Li (2001), we have

$$\frac{\partial f_{\lambda_{k,n}}(|\bar{\boldsymbol{\theta}}_{k,\cdot}|)}{\partial \boldsymbol{\theta}_{0,k,\cdot}(t)} \approx \frac{f'_{\lambda_{k,n}}(|\bar{\boldsymbol{\theta}}_{k,\cdot}^{(0)}|)}{|\bar{\boldsymbol{\theta}}_{k,\cdot}^{(0)}|} \bar{\boldsymbol{\theta}}_{k,\cdot} dt,$$

and

$$\frac{\partial g_{\tau_{k,n}}([\int_0^1\{|\boldsymbol{\theta}_{0,k,\cdot}(t)-\bar{\boldsymbol{\theta}}_{k,\cdot}|^2+|\boldsymbol{\theta}_{1,k,\cdot}(t)|^2\}dt]^{1/2})}{\partial\boldsymbol{\theta}_{0,k,\cdot}(t)}$$
$$\approx \frac{g'_{\tau_{k,n}}([\int_0^1\{|\boldsymbol{\theta}^{(0)}_{0,k,\cdot}(t)-\bar{\boldsymbol{\theta}}^{(0)}_{k,\cdot}|^2+|\boldsymbol{\theta}^{(0)}_{1,k,\cdot}(t)|^2\}dt]^{1/2})}{[\int_0^1\{|\boldsymbol{\theta}^{(0)}_{0,k,\cdot}(t)-\bar{\boldsymbol{\theta}}^{(0)}_{k,\cdot}|^2+|\boldsymbol{\theta}^{(0)}_{1,k,\cdot}(t)|^2\}dt]^{1/2}}\{\boldsymbol{\theta}_{0,k,\cdot}(t)-\bar{\boldsymbol{\theta}}_{k,\cdot}\}dt.$$

Write $\bar{\boldsymbol{\theta}}^{(0)}_{\cdot,j}=(\bar{\theta}^{(0)}_{1,j},\ldots,\bar{\theta}^{(0)}_{p,j})^\top$,

$$D^{(0)}_{c,n}=\operatorname{diag}\left\{\frac{1}{2nb_n}\cdot\frac{f'_{\lambda_{k,n}}(|\bar{\boldsymbol{\theta}}^{(0)}_{k,\cdot}|)}{|\bar{\boldsymbol{\theta}}^{(0)}_{k,\cdot}|},\ 1\le k\le p\right\},$$

$$D^{(0)}_{v,n}=\operatorname{diag}\left\{\frac{1}{2nb_n}\cdot\frac{g'_{\tau_{k,n}}([\int_0^1\{|\boldsymbol{\theta}^{(0)}_{0,k,\cdot}(t)-\bar{\boldsymbol{\theta}}^{(0)}_{k,\cdot}|^2+|\boldsymbol{\theta}^{(0)}_{1,k,\cdot}(t)|^2\}dt]^{1/2})}{[\int_0^1\{|\boldsymbol{\theta}^{(0)}_{0,k,\cdot}(t)-\bar{\boldsymbol{\theta}}^{(0)}_{k,\cdot}|^2+|\boldsymbol{\theta}^{(0)}_{1,k,\cdot}(t)|^2\}dt]^{1/2}},\ 1\le k\le p\right\},$$

and let

$$U_l(t)=\frac{1}{nb_n}\sum_{i=1}^n\boldsymbol{x}_{i,n}\boldsymbol{x}_{i,n}^\top\left(\frac{t_{i,n}-t}{b_n}\right)^lK\left(\frac{t_{i,n}-t}{b_n}\right),$$
$$V_{l,j}(t)=\frac{1}{nb_n}\sum_{i=1}^n y_{i,j,n}\boldsymbol{x}_{i,n}\left(\frac{t_{i,n}-t}{b_n}\right)^lK\left(\frac{t_{i,n}-t}{b_n}\right).$$

We first update the time-varying stratum $\boldsymbol{\iota}_{0,\cdot,j}(t)=\boldsymbol{\theta}_{0,\cdot,j}(t)-\bar{\boldsymbol{\theta}}^{(0)}_{\cdot,j}$ and $\boldsymbol{\iota}_{1,\cdot,j}(t)=\boldsymbol{\theta}_{1,\cdot,j}(t)$ by

$$\begin{bmatrix}\boldsymbol{\iota}^{(1)}_{0,\cdot,j}(t)\\ \boldsymbol{\iota}^{(1)}_{1,\cdot,j}(t)\end{bmatrix}=\left(\begin{bmatrix}U_0(t)&U_1(t)\\U_1(t)&U_2(t)\end{bmatrix}+\begin{bmatrix}D^{(0)}_{v,n}&\\&D^{(0)}_{v,n}\end{bmatrix}\right)^{-1}\begin{bmatrix}V_{0,j}(t)-\{U_0(t)+D^{(0)}_{c,n}\}\bar{\boldsymbol{\theta}}^{(0)}_{\cdot,j}\\ V_{1,j}(t)-U_1(t)\bar{\boldsymbol{\theta}}^{(0)}_{\cdot,j}\end{bmatrix}.$$

Note that the computation in this step can be made in parallel for $j=1,\ldots,d$. Then, we update the time-constant stratum by

$$\bar{\boldsymbol{\theta}}^{(1)}_{\cdot,j}=\left\{\int_0^1U_0(t)dt+D^{(0)}_{c,n}\right\}^{-1}\int_0^1\{V_{0,j}(t)-U_0(t)\boldsymbol{\iota}^{(1)}_{0,\cdot,j}(t)-U_1(t)\boldsymbol{\iota}^{(1)}_{1,\cdot,j}(t)\}dt,$$

where the term $D_{c,n}^{(0)}$ makes use of information from all outputs. Combining results from both strata, the updated estimate can now be computed as $\boldsymbol{\theta}_{0,\cdot,j}^{(1)}(t) = \bar{\boldsymbol{\theta}}_{\cdot,j}^{(1)} + \boldsymbol{\iota}_{0,\cdot,j}^{(1)}(t) - \int_0^1 \boldsymbol{\iota}_{0,\cdot,j}^{(1)}(t)dt$ and $\boldsymbol{\theta}_{1,\cdot,j}^{(1)}(t) = \boldsymbol{\iota}_{1,\cdot,j}^{(1)}(t)$. The proposed stratified penalized local linear estimator can then be obtained by iteratively repeat the aforementioned steps.

S2 Additional Simulation Results

In Table 1 of this supplementary material, we provide simulation results when $\sigma = 1$. A similar pattern can be observed as discussed in the main article.

S3 Technical Proofs

Here, we provide proofs for our results in Sections 2 and 3.

Proof. (Theorem 1) For data with nondegenerate local designs, we have $\sum_{i=1}^n x_{i,n}^2 K\{(t_{i,n} - t)/b_n\} > 0$ for each $t \in [0, 1]$. Then, for any given $\eta'(t)$, the $\eta(t)$ that minimizes (2.6) can be solved as

$$\tilde{\eta}(t) = \left\{ \sum_{i=1}^n x_{i,n}^2 K\left(\frac{t_{i,n} - t}{b_n}\right) \right\}^{-1} \left[\sum_{i=1}^n x_{i,n}\{y_{i,n} - x_{i,n}\eta'(t)(t_{i,n} - t)\} K\left(\frac{t_{i,n} - t}{b_n}\right) \right],$$

and we denote this relationship by $\tilde{\eta}(t) = H\{t, \eta'(t)\}$. Let $\{\eta^\star(t), \eta'^\star(t)\}$ be the minimizer of (2.6), then $\eta^\star(t) = H\{t, \eta'^\star(t)\}$, and it suffices to consider

Table 1: Simulation results for $\sigma = 1$, based on 100 realizations for each configuration.

Model	b_n	Method	Under-label	Correct	Over-label	MSE	LCR
2-2-16	0.1	SPLL	0.00	1.00	0.00	0.0093	1.0000
		Zhang15	0.04	0.93	0.03	0.0327	0.9920
		Gao19	0.00	0.72	0.28	0.0138	0.9835
	0.2	SPLL	0.00	1.00	0.00	0.0110	1.0000
		Zhang15	0.05	0.92	0.03	0.0311	0.9920
		Gao19	0.00	0.29	0.71	0.0106	0.9295
	0.3	SPLL	0.00	1.00	0.00	0.0232	1.0000
		Zhang15	0.06	0.91	0.03	0.0386	0.9920
		Gao19	0.00	0.15	0.85	0.0213	0.8820
5-5-10	0.1	SPLL	0.00	1.00	0.00	0.0197	1.0000
		Zhang15	0.04	0.94	0.02	0.0970	0.9805
		Gao19	0.93	0.03	0.04	0.1734	0.9315
	0.2	SPLL	0.00	1.00	0.00	0.0219	1.0000
		Zhang15	0.04	0.93	0.03	0.0989	0.9820
		Gao19	0.00	0.31	0.69	0.0229	0.9290
	0.3	SPLL	0.00	1.00	0.00	0.0376	1.0000
		Zhang15	0.04	0.87	0.09	0.1117	0.9800
		Gao19	0.00	0.17	0.83	0.0384	0.8870
2-8-10	0.1	SPLL	0.00	0.99	0.01	0.0110	0.9995
		Zhang15	0.04	0.94	0.02	0.0857	0.9825
		Gao19	0.00	0.80	0.20	0.0159	0.9880
	0.2	SPLL	0.00	1.00	0.00	0.0127	1.0000
		Zhang15	0.05	0.92	0.03	0.0841	0.9815
		Gao19	0.00	0.47	0.53	0.0127	0.9520
	0.3	SPLL	0.00	1.00	0.00	0.0244	1.0000
		Zhang15	0.05	0.91	0.04	0.0919	0.9810
		Gao19	0.00	0.22	0.78	0.0235	0.9145

the minimization problem

$$\min_{\eta' \in \mathcal{S}} \Omega_n([H\{t, \eta'(t)\}, \eta'(t)]_{t\in[0,1]}) + f_\tau\left(|\eta'|_{[0,1]}\right), \tag{S3.1}$$

where $\mathcal{S}$ is the class of all square integrable continuous functions on $[0, 1]$. Let Ω_n^0 denote the value of $\Omega_n([H\{t, \eta'(t)\}, \eta'(t)]_{t\in[0,1]})$ when $\eta'(t) = 0$ for all $t \in [0,1]$, then for any $\varepsilon_0 \in (0,1]$ and $\tau \geq \tau_0 > 0$,

$$f_\tau\left(|\eta'|_{[0,1]}\right) \geq f_\tau(0) + \tau|\eta'|_{[0,1]} \geq \varepsilon_0\tau_0|\eta'|_{[0,1]}.$$

Let $M_0 = (\varepsilon_0\tau_0)^{-1}\Omega_n^0$, then for any η' with $|\eta'|_{[0,1]} > M_0$, we have $f_\tau\left(|\eta'|_{[0,1]}\right) > \Omega_n^0$ and thus η' cannot be the minimizer of (S3.1). Therefore, the minimization problem (S3.1) is equivalent to

$$\min_{\eta' \in \mathcal{S},\ |\eta'|_{[0,1]} \leq M_0} \Omega_n([H\{t,\eta'(t)\},\eta'(t)]_{t\in[0,1]}) + f_\tau\left(|\eta'|_{[0,1]}\right).$$

For any $\eta'(t) \in L^2[0,1]$ with $|\eta'|_{[0,1]} \leq M_0$, we define

$$\eta'_\varepsilon(t) = \eta'^\star(t) + \varepsilon\left\{\eta'(t) - \eta'^\star(t)\right\}, \quad \varepsilon \in [0,1].$$

Then, by the definition of $\eta'^\star(t)$,

$$\Omega_n([H\{t,\eta'_\varepsilon(t)\},\eta'_\varepsilon(t)]_{t\in[0,1]}) + f_\tau\left(|\eta'_\varepsilon|_{[0,1]}\right) \geq \Omega_n([H\{t,\eta'^\star(t)\},\eta'^\star(t)]_{t\in[0,1]}) + f_\tau\left(|\eta'^\star|_{[0,1]}\right).$$

Since

$$f_\tau(|\eta'_\varepsilon|_{[0,1]}) \leq f_\tau(\varepsilon|\eta'|_{[0,1]} + (1-\varepsilon)|\eta'^\star|_{[0,1]}) \leq \varepsilon f_\tau(|\eta'|_{[0,1]}) + (1-\varepsilon)f_\tau(|\eta'^\star|_{[0,1]}),$$

we have

$$\Omega_n([H\{t,\eta'_\varepsilon(t)\},\eta'_\varepsilon(t)]_{t\in[0,1]}) - \Omega_n([H\{t,\eta'^\star(t)\},\eta'^\star(t)]_{t\in[0,1]}) + \varepsilon\{f_\tau(|\eta'|_{[0,1]}) - f_\tau(|\eta'^\star|_{[0,1]})\} \geq 0$$

holds for any $\varepsilon \in [0,1]$. This leads to the optimality condition

$$\lim_{\varepsilon\to 0+} \frac{\Omega_n([H\{t,\eta'_\varepsilon(t)\},\eta'_\varepsilon(t)]_{t\in[0,1]}) - \Omega_n([H\{t,\eta'^\star(t)\},\eta'^\star(t)]_{t\in[0,1]})}{\varepsilon} + f_\tau(|\eta'|_{[0,1]}) - f_\tau(|\eta'^\star|_{[0,1]}) \geq 0.$$

Denote the functional derivative as

$$\begin{aligned}
&\int_0^1 \frac{\delta \Omega_n}{\delta \eta'^\star}(t)\{\eta'(t)-\eta'^\star(t)\}dt \\
= & \lim_{\varepsilon\to 0+}\frac{1}{\varepsilon}\left\{\Omega_n([H\{t,\eta'_\varepsilon(t)\},\eta'_\varepsilon(t)]_{t\in[0,1]})-\Omega_n([H\{t,\eta'^\star(t)\},\eta'^\star(t)]_{t\in[0,1]})\right\} \\
= & \int_0^1\left[2\sum_{i=1}^n\left\{[x_{i,n}H\{t,\eta'^\star(t)\}+x_{i,n}\eta'^\star(t)(t_{i,n}-t)-y_{i,n}]K\left(\frac{t_{i,n}-t}{b_n}\right)x_{i,n}\right\}\frac{\partial H\{t,\eta'(t)\}}{\partial \eta'(t)}\right. \\
&\left.+2\sum_{i=1}^n\left\{[x_{i,n}H\{t,\eta'^\star(t)\}+x_{i,n}\eta'^\star(t)(t_{i,n}-t)-y_{i,n}]K\left(\frac{t_{i,n}-t}{b_n}\right)x_{i,n}(t_{i,n}-t)\right\}\right] \\
&\times\{\eta'(t)-\eta'^\star(t)\}dt,
\end{aligned}$$

where

$$\frac{\partial H\{t,\eta'(t)\}}{\partial \eta'(t)}=\left\{\sum_{i=1}^n x_{i,n}^2 K\left(\frac{t_{i,n}-t}{b_n}\right)\right\}^{-1}\left\{\sum_{i=1}^n x_{i,n}^2\,(t_{i,n}-t)\,K\left(\frac{t_{i,n}-t}{b_n}\right)\right\}.$$

Then, we can write

$$\int_0^1 \frac{\delta \Omega_n}{\delta \eta'^\star}(t)\{\eta'(t)-\eta'^\star(t)\}dt=\int_0^1\{C_0(t)+C_1(t)\eta'^\star(t)\}\{\eta'(t)-\eta'^\star(t)\}dt,$$

where $C_0(t)$ and $C_1(t)$ are constants that depend on t, either directly or through $\eta'^\star(t)$, but do not depend on $\eta(t)$, $\eta'(t)$ or τ. In this case, the optimality condition becomes

$$\int_0^1 \frac{\delta \Omega_n}{\delta \eta'^\star}(t)\{\eta'(t)-\eta'^\star(t)\}dt+f_\tau(|\eta'|_{[0,1]})-f_\tau(|\eta'^\star|_{[0,1]})\geq 0.$$

Note that

$$f_\tau(|\eta'^\star|_{[0,1]})\geq f_\tau(|\eta|_{[0,1]})+f'_\tau(|\eta|_{[0,1]})(|\eta'^\star|_{[0,1]}-|\eta'|_{[0,1]}),$$

the optimality condition then implies that

$$\int_0^1 \{C_0(t) + C_1(t)\eta'^{\star}(t)\}\{\eta'(t) - \eta'^{\star}(t)\}dt \geq f'_{\tau}(|\eta'|_{[0,1]})(|\eta'^{\star}|_{[0,1]} - |\eta'|_{[0,1]})$$

holds for all $\eta' \in \mathcal{S}$ with $|\eta'|_{[0,1]} \leq M_0$. In particular, if we let $\eta'(t) = 0$, for $t \in [0,1]$, then we have

$$\left(\sup_{t\in[0,1]} |C_0(t)|\right) |\eta'^{\star}|_{[0,1]} + \left(\sup_{t\in[0,1]} |C_1(t)|\right) (|\eta'^{\star}|_{[0,1]})^2 \geq f'_{\tau}(0) \cdot |\eta'^{\star}|_{[0,1]},$$

and thus

$$|\eta'^{\star}|_{[0,1]} \left\{ \sup_{t\in[0,1]} |C_0(t)| + M_0 \sup_{t\in[0,1]} |C_1(t)| - f'_{\tau}(0) \right\} \geq 0. \tag{S3.2}$$

For any $\tau > \tau_0$, we have $f'_{\tau}(0) \geq \varepsilon_0 \tau$, and as a result

$$\sup_{t\in[0,1]} |C_0(t)| + M_0 \sup_{t\in[0,1]} |C_1(t)| - f'_{\tau}(0) \leq 0$$

holds if

$$\tau \geq \frac{\sup_{t\in[0,1]} |C_0(t)| + M_0 \sup_{t\in[0,1]} |C_1(t)| - f'_{\tau}(0)}{\varepsilon_0}.$$

Therefore, in order for (S3.2) to hold, one must have $|\eta'^{\star}|_{[0,1]} = 0$. Since $\eta'^{\star} \in \mathcal{S}$, we have $\eta'^{\star}(t) = 0$, for $t \in [0,1]$, and the minimizer of (2.6) is then given by

$$\{\eta^{\star}(t), \eta'^{\star}(t)\} = \left[\left\{ \sum_{i=1}^{n} x_{i,n}^2 K\left(\frac{t_{i,n} - t}{b_n}\right) \right\}^{-1} \left\{ \sum_{i=1}^{n} x_{i,n} y_{i,n} K\left(\frac{t_{i,n} - t}{b_n}\right) \right\}, 0 \right],$$

which coincides with the local constant estimator (2.2), and the result follows. □

Here, we introduce the following notations for the proof of Theorems 2 and 3. For any $m, n \in \mathbb{R}$, we define $m \wedge n = \min(m, n)$ and $m \vee n = \max(m, n)$. For any matrix $\boldsymbol{A} = (a_{i,j})_{i,j}$, we use $\boldsymbol{a}_{\cdot,j}$ to denote its jth column and $\boldsymbol{a}_{i,\cdot}$ to denote the transpose of its ith row, and we write $\bar{\rho}(\boldsymbol{A}) = \sup\{|\boldsymbol{A}\boldsymbol{u}| : |\boldsymbol{u}| = 1\}$ and $\tilde{\rho}(\boldsymbol{A}) = \inf\{|\boldsymbol{A}\boldsymbol{u}| : |\boldsymbol{u}| = 1\}$. We use the matrix norm $|\boldsymbol{A}| = (\sum_{i,j} a_{i,j}^2)^{1/2}$, and for any matrix-valued function $\boldsymbol{f} : [0,1] \to \mathbb{R}^{s_1 \times s_2}$, we write $|\boldsymbol{f}|_{[0,1]} = \{\int_0^1 |\boldsymbol{f}(t)|^2 dt\}^{1/2}$. For the multi-output response vector $\boldsymbol{y}_{i,n} \in \mathbb{R}^d$, we write $\boldsymbol{y}_{i,n} = (y_{i,1,n}, \ldots, y_{i,d,n})^\top$, and similarly $\boldsymbol{e}_{i,n} = (e_{i,1,n}, \ldots, e_{i,d,n})^\top$. Let $\tilde{\boldsymbol{x}}_{i,n} = \boldsymbol{G}(t_{i,n}; \boldsymbol{\mathcal{F}}_i)$, $\tilde{\boldsymbol{e}}_{i,n} = \boldsymbol{H}(t_{i,n}; \boldsymbol{\mathcal{F}}_i)$ and $\boldsymbol{M}(\boldsymbol{G}, t) = E\{\boldsymbol{G}(t, \boldsymbol{\mathcal{F}}_i)\boldsymbol{G}(t, \boldsymbol{\mathcal{F}}_i)^\top\}$. We denote the true value by $\{\Theta^{\text{true}}(t)\}_{t\in[0,1]}$, and let C be a constant whose values may vary from place to place.

Proof. (Theorem 2) Let

$$\Psi_n(\{\Theta(t)\}_{t\in[0,1]}) = \sum_{k=1}^{p} g_{\tau_{k,n}}\left(\left[\int_0^1 \{|\boldsymbol{\theta}_{0,k,\cdot}(t) - \bar{\boldsymbol{\theta}}_k|^2 + |\boldsymbol{\theta}_{1,k,\cdot}(t)|^2\} dt\right]^{1/2}\right) + \Upsilon_n(\{\Theta(t)\}_{t\in[0,1]}) + \sum_{k=1}^{p} f_{\lambda_{k,n}}(|\bar{\boldsymbol{\theta}}_k|), \tag{S3.3}$$

$\phi_n = (nb_n)^{-1/2} + b_n^2$, and $V(t) = \{v_{\cdot,\cdot,\cdot}(t)\}_{l,j,k}$ be a 3-way tensor function. Then, by the proof in Theorem 1 of Fan and Li (2001), it suffices to show that, for any $\epsilon > 0$, there exists a large constant $Q^\star$ such that

$$\text{pr}\left\{\inf_{|V|_{[0,1]}=Q^\star} \Psi_n(\{\Theta^{\text{true}}(t) + \phi_n V(t)\}_{t\in[0,1]}) > \Psi_n(\{\Theta^{\text{true}}(t)\}_{t\in[0,1]})\right\} \geq 1 - \epsilon.$$

By (S3.3), we can write

$$\begin{aligned}&\Psi_n(\{\Theta^{\text{true}}(t)+\phi_n V(t)\}_{t\in[0,1]})-\Psi_n(\{\Theta^{\text{true}}(t)\}_{t\in[0,1]})\\=\;&I_n(\{V(t)\}_{t\in[0,1]})+II_n(\{V(t)\}_{t\in[0,1]})+III_n(\{V(t)\}_{t\in[0,1]}),\end{aligned}$$

where

$$\begin{aligned}I_n(\{V(t)\}_{t\in[0,1]})&=\Upsilon_n(\{\Theta^{\text{true}}(t)+\phi_n V(t)\}_{t\in[0,1]})-\Upsilon_n(\{\Theta^{\text{true}}(t)\}_{t\in[0,1]}),\\II_n(\{V(t)\}_{t\in[0,1]})&=\sum_{k=1}^{p} f_{\lambda_{k,n}}(|\bar{\boldsymbol{\theta}}_k^{\text{true}}+\phi_n\bar{\boldsymbol{v}}_k|)-\sum_{k=1}^{p} f_{\lambda_{k,n}}(|\bar{\boldsymbol{\theta}}_k^{\text{true}}|),\end{aligned}$$

and

$$\begin{aligned}&III_n(\{V(t)\}_{t\in[0,1]})\\=\;&\sum_{k=1}^{p} g_{\tau_{k,n}}\left(\left[\int_0^1\{|\boldsymbol{\theta}_{0,k,\cdot}^{\text{true}}(t)+\phi_n\boldsymbol{v}_{0,k,\cdot}(t)-\bar{\boldsymbol{\theta}}_k^{\text{true}}-\phi_n\bar{\boldsymbol{v}}_k|^2+|\boldsymbol{\theta}_{1,k,\cdot}^{\text{true}}(t)+\phi_n\boldsymbol{v}_{1,k,\cdot}(t)|^2\}dt\right]^{1/2}\right)\\&-\sum_{k=1}^{p} g_{\tau_{k,n}}\left(\left[\int_0^1\{|\boldsymbol{\theta}_{0,k,\cdot}^{\text{true}}(t)-\bar{\boldsymbol{\theta}}_k^{\text{true}}|^2+|\boldsymbol{\theta}_{1,k,\cdot}^{\text{true}}(t)|^2\}dt\right]^{1/2}\right).\end{aligned}$$

Here, we provide bounds for $I_n(\{V(t)\}_{t\in[0,1]})$, $II_n(\{V(t)\}_{t\in[0,1]})$ and $III_n(\{V(t)\}_{t\in[0,1]})$.

We first deal with the term $I_n(\{V(t)\}_{t\in[0,1]})$. For this, by Lemma 6 of Zhou and Wu (2010), for $s\in\{0,1,2\}$, we have

$$\left\|\frac{1}{nb_n}\sum_{i=1}^{n}\tilde{\boldsymbol{x}}_{i,n}\tilde{\boldsymbol{x}}_{i,n}^{\top}\left(\frac{t_{i,n}-t}{b_n}\right)^{s}K\left(\frac{t_{i,n}-t}{b_n}\right)-\boldsymbol{M}(\boldsymbol{G},t)\int_{-1}^{1}u^{s}K(u)du\right\|\le C\{(nb_n)^{-1/2}+b_n\}.$$

In addition, by (3.2), we have

$$\sup_{t\in[0,1]}\left|\sum_{i=1}^{n}(\boldsymbol{x}_{i,n}\boldsymbol{x}_{i,n}^{\top}-\tilde{\boldsymbol{x}}_{i,n}\tilde{\boldsymbol{x}}_{i,n}^{\top})K\left(\frac{t_{i,n}-t}{b_n}\right)\right|=O_p(1),\qquad\text{(S3.4)}$$

and

$$\sup_{t\in[0,1]}\left|\sum_{i=1}^{n}(\boldsymbol{x}_{i,n}\boldsymbol{e}_{i,n}^{\top}-\tilde{\boldsymbol{x}}_{i,n}\tilde{\boldsymbol{e}}_{i,n}^{\top})K\left(\frac{t_{i,n}-t}{b_n}\right)\right|=O_p(1). \tag{S3.5}$$

Let $\bar{\rho}_{\boldsymbol{M}} = \sup_{t\in[0,1]}\bar{\rho}\{\boldsymbol{M}(\boldsymbol{G},t)\}$, $\tilde{\rho}_{\boldsymbol{M}} = \inf_{t\in[0,1]}\tilde{\rho}\{\boldsymbol{M}(\boldsymbol{G},t)\}$, and $\kappa_s = \int_{-1}^{1}u^sK(u)du$. Then, $\kappa_0=1$, $\kappa_1=0$, and

$$\begin{aligned}
&\sum_{j=1}^{d}\int_0^1\sum_{i=1}^{n}\left[\boldsymbol{x}_{i,n}^{\top}\left\{\boldsymbol{v}_{0,\cdot,j}(t)+\boldsymbol{v}_{1,\cdot,j}(t)\left(\frac{t_{i,n}-t}{b_n}\right)\right\}\right]^2K\left(\frac{t_{i,n}-t}{b_n}\right)dt\\
=\;&\sum_{j=1}^{d}\int_0^1\sum_{i=1}^{n}\left\{\boldsymbol{v}_{0,\cdot,j}(t)^{\top}\boldsymbol{x}_{i,n}\boldsymbol{x}_{i,n}^{\top}\boldsymbol{v}_{0,\cdot,j}(t)+2\boldsymbol{v}_{0,\cdot,j}(t)^{\top}\boldsymbol{x}_{i,n}\boldsymbol{x}_{i,n}^{\top}\boldsymbol{v}_{1,\cdot,j}(t)\left(\frac{t_{i,n}-t}{b_n}\right)\right.\\
&\left.+\boldsymbol{v}_{1,\cdot,j}(t)^{\top}\boldsymbol{x}_{i,n}\boldsymbol{x}_{i,n}^{\top}\boldsymbol{v}_{1,\cdot,j}(t)\left(\frac{t_{i,n}-t}{b_n}\right)^2\right\}K\left(\frac{t_{i,n}-t}{b_n}\right)dt\\
=\;&nb_n\sum_{j=1}^{d}\int_0^1\{\boldsymbol{v}_{0,\cdot,j}(t)^{\top}\boldsymbol{M}(\boldsymbol{G},t)\boldsymbol{v}_{0,\cdot,j}(t)+\kappa_2\boldsymbol{v}_{1,\cdot,j}(t)^{\top}\boldsymbol{M}(\boldsymbol{G},t)\boldsymbol{v}_{1,\cdot,j}(t)\}dt\\
&+nb_n|V|_{[0,1]}^2O_p\{(nb_n)^{-1/2}+b_n\}\\
\geq\;&nb_n|V|_{[0,1]}^2[(\kappa_2\wedge 1)\tilde{\rho}_{\boldsymbol{M}}+O_p\{(nb_n)^{-1/2}+b_n\}].
\end{aligned}$$

On the other hand, by the Cauchy-Schwarz inequality and Lemma A.1 of

Zhang and Wu (2012), we have

$$
\begin{aligned}
& \left| \sum_{j=1}^{d} \int_0^1 \sum_{i=1}^{n} \boldsymbol{x}_{i,n}^\top \left\{ \boldsymbol{v}_{0,\cdot,j}(t) + \boldsymbol{v}_{1,\cdot,j}(t) \left(\frac{t_{i,n} - t}{b_n} \right) \right\} e_{i,j,n} K \left(\frac{t_{i,n} - t}{b_n} \right) dt \right| \\
\leq \; & \sum_{j=1}^{d} \left| \int_0^1 \boldsymbol{v}_{0,\cdot,j}(t)^\top \left\{ \sum_{i=1}^{n} \boldsymbol{x}_{i,n} e_{i,j,n} K \left(\frac{t_{i,n} - t}{b_n} \right) \right\} dt \right| \\
& + \sum_{j=1}^{d} \left| \int_0^1 \boldsymbol{v}_{1,\cdot,j}(t)^\top \left\{ \sum_{i=1}^{n} \boldsymbol{x}_{i,n} e_{i,j,n} \left(\frac{t_{i,n} - t}{b_n} \right) K \left(\frac{t_{i,n} - t}{b_n} \right) \right\} dt \right| \\
\leq \; & \left(\sum_{j=1}^{d} \int_0^1 |\boldsymbol{v}_{0,\cdot,j}(t)| dt \right) O_p\{1 + (nb_n)^{1/2}\} + \left(\sum_{j=1}^{d} \int_0^1 |\boldsymbol{v}_{1,\cdot,j}(t)| dt \right) O_p\{1 + (nb_n)^{1/2}\} \\
\leq \; & 2d|V|_{[0,1]} O_p\{(nb_n)^{1/2}\}.
\end{aligned}
$$

Since $-1 \leq (t_{i,n} - t)/b_n \leq 1$, we have

$$
\begin{aligned}
\left| \boldsymbol{\theta}_{0,\cdot,j}^{\text{true}}(t_{i,n}) - \boldsymbol{\theta}_{0,\cdot,j}^{\text{true}}(t) - \boldsymbol{\theta}_{1,\cdot,j}^{\text{true}}(t) \left(\frac{t_{i,n} - t}{b_n} \right) \right| &= \left| \frac{b_n^2}{2} \left(\frac{t_{i,n} - t}{b_n} \right)^2 \boldsymbol{b}''_{\cdot,j}(\xi) \right| \\
&\leq \frac{b_n^2}{2} \sup_{t \in [0,1]} |\boldsymbol{b}''_{\cdot,j}(t)| \qquad \text{(S3.6)}
\end{aligned}
$$

holds for some $\xi \in [0,1]$. Then, by a similar argument, we can obtain that

$$\begin{aligned}
&\left|\sum_{j=1}^{d}\int_0^1\sum_{i=1}^{n}\left\{\boldsymbol{\theta}_{0,\cdot,j}^{\mathrm{true}}(t_{i,n})-\boldsymbol{\theta}_{0,\cdot,j}^{\mathrm{true}}(t)-\boldsymbol{\theta}_{1,\cdot,j}^{\mathrm{true}}(t)\left(\frac{t_{i,n}-t}{b_n}\right)\right\}^{\top}\boldsymbol{x}_{i,n}\right.\\
&\qquad\qquad\left.\times\boldsymbol{x}_{i,n}^{\top}\left\{\boldsymbol{v}_{0,\cdot,j}(t)+\boldsymbol{v}_{1,\cdot,j}(t)\left(\frac{t_{i,n}-t}{b_n}\right)\right\}K\left(\frac{t_{i,n}-t}{b_n}\right)dt\right|\\
\leq\;& \sum_{j=1}^{d}\frac{b_n^2}{2}\sup_{t\in[0,1]}|\boldsymbol{b}''_{\cdot,j}(t)|\left[\left|\int_0^1\sum_{i=1}^{n}\tilde{\boldsymbol{x}}_{i,n}\tilde{\boldsymbol{x}}_{i,n}^{\top}\left\{\boldsymbol{v}_{0,\cdot,j}(t)+\boldsymbol{v}_{1,\cdot,j}(t)\left(\frac{t_{i,n}-t}{b_n}\right)\right\}K\left(\frac{t_{i,n}-t}{b_n}\right)dt\right|\right.\\
&\left.+\left|\int_0^1\sum_{i=1}^{n}(\boldsymbol{x}_{i,n}\boldsymbol{x}_{i,n}^{\top}-\tilde{\boldsymbol{x}}_{i,n}\tilde{\boldsymbol{x}}_{i,n}^{\top})\left\{\boldsymbol{v}_{0,\cdot,j}(t)+\boldsymbol{v}_{1,\cdot,j}(t)\left(\frac{t_{i,n}-t}{b_n}\right)\right\}K\left(\frac{t_{i,n}-t}{b_n}\right)dt\right|\right]\\
\leq\;& \sum_{j=1}^{d}\frac{b_n^2}{2}\sup_{t\in[0,1]}|\boldsymbol{b}''_{\cdot,j}(t)|\left(2nb_n|V|_{[0,1]}[\bar{\rho}_{\boldsymbol{M}}+O_p\{(nb_n)^{-1}+(nb_n)^{-1/2}+b_n\}]\right)\\
=\;& nb_n|V|_{[0,1]}\left[b_n^2\left\{\sum_{j=1}^{d}\sup_{t\in[0,1]}|\boldsymbol{b}''_{\cdot,j}(t)|\right\}\bar{\rho}_{\boldsymbol{M}}+O_p\{b_n^2(nb_n)^{-1/2}+b_n^3\}\right].
\end{aligned}$$

Combining the results above, we have

$$\begin{aligned}
&I_n(\{V(t)\}_{t\in[0,1]})\\
=\;& \phi_n^2\sum_{j=1}^{d}\int_0^1\sum_{i=1}^{n}\left[\boldsymbol{x}_{i,n}^{\top}\left\{\boldsymbol{v}_{0,\cdot,j}(t)+\boldsymbol{v}_{1,\cdot,j}(t)\left(\frac{t_{i,n}-t}{b_n}\right)\right\}\right]^2K\left(\frac{t_{i,n}-t}{b_n}\right)dt\\
&-2\phi_n\sum_{j=1}^{d}\int_0^1\sum_{i=1}^{n}\left[\boldsymbol{x}_{i,n}^{\top}\boldsymbol{\theta}_{0,\cdot,j}^{\mathrm{true}}(t_{i,n})+e_{i,j,n}-\boldsymbol{x}_{i,n}^{\top}\left\{\boldsymbol{\theta}_{0,\cdot,j}^{\mathrm{true}}(t)+\boldsymbol{\theta}_{1,\cdot,j}^{\mathrm{true}}(t)\left(\frac{t_{i,n}-t}{b_n}\right)\right\}\right]\\
&\qquad\qquad\times\left[\boldsymbol{x}_{i,n}^{\top}\left\{\boldsymbol{v}_{0,\cdot,j}(t)+\boldsymbol{v}_{1,\cdot,j}(t)\left(\frac{t_{i,n}-t}{b_n}\right)\right\}\right]K\left(\frac{t_{i,n}-t}{b_n}\right)dt\\
\geq\;& \phi_n^2nb_n|V|_{[0,1]}^2[(\kappa_2\wedge 1)\tilde{\rho}_{\boldsymbol{M}}+O_p\{(nb_n)^{-1/2}+b_n\}]\\
&-2\phi_nnb_n|V|_{[0,1]}\left[2dO_p\{(nb_n)^{-1/2}\}+b_n^2\left\{\sum_{j=1}^{d}\sup_{t\in[0,1]}\left|\boldsymbol{b}''_{\cdot,j}(t)\right|\right\}\bar{\rho}_{\boldsymbol{M}}+O_p\{b_n^2(nb_n)^{-1/2}+b_n^3\}\right]\\
\geq\;& \phi_n^2nb_n|V|_{[0,1]}^2[(\kappa_2\wedge 1)\tilde{\rho}_{\boldsymbol{M}}+O_p\{(nb_n)^{-1/2}+b_n\}]\\
&-2\phi_nnb_n|V|_{[0,1]}\left[b_n^2\left\{\sum_{j=1}^{d}\sup_{t\in[0,1]}|\boldsymbol{b}''_{\cdot,j}(t)|\right\}\bar{\rho}_{\boldsymbol{M}}+O_p\{(nb_n)^{-1/2}+b_n^3\}\right].
\end{aligned}$$

Now, we deal with the term $II_n(\{V(t)\}_{t\in[0,1]})$. For this, note that for $k \in \mathcal{D}_0 \cup \mathcal{D}_{v0}$, we have $|\bar{\boldsymbol{\theta}}_k^{\text{true}}| = 0$, and thus

$$II_n(\{V(t)\}_{t\in[0,1]}) \geq \sum_{k\in\mathcal{D}_c\cup\mathcal{D}_{v1}} f_{\lambda_{k,n}}(|\bar{\boldsymbol{\theta}}_k^{\text{true}} + \phi_n \bar{\boldsymbol{v}}_k|) - \sum_{k\in\mathcal{D}_c\cup\mathcal{D}_{v1}} f_{\lambda_{k,n}}(|\bar{\boldsymbol{\theta}}_k^{\text{true}}|).$$

Then, by the triangle inequality and Cauchy-Schwarz inequality, we have

$$\begin{aligned}
& \left| \sum_{k\in\mathcal{D}_c\cup\mathcal{D}_{v1}} f_{\lambda_{k,n}}(|\bar{\boldsymbol{\theta}}_k^{\text{true}} + \phi_n \bar{\boldsymbol{v}}_k|) - \sum_{k\in\mathcal{D}_c\cup\mathcal{D}_{v1}} f_{\lambda_{k,n}}(|\bar{\boldsymbol{\theta}}_k^{\text{true}}|) \right| \\
= & \left| \sum_{k\in\mathcal{D}_c\cup\mathcal{D}_{v1}} f'_{\lambda_{k,n}}(\xi_k) \left(|\bar{\boldsymbol{\theta}}_k^{\text{true}} + \phi_n \bar{\boldsymbol{v}}_k| - |\bar{\boldsymbol{\theta}}_k^{\text{true}}| \right) \right| \\
\leq & \left| \sum_{k\in\mathcal{D}_c\cup\mathcal{D}_{v1}} f'_{\lambda_{k,n}}(\xi_k) |\phi_n \bar{\boldsymbol{v}}_k| \right| \leq \phi_n |V|_{[0,1]} \sum_{k\in\mathcal{D}_c\cup\mathcal{D}_{v1}} \sup_{x\in\mathbb{R}} |f'_{\lambda_{k,n}}(x)|,
\end{aligned}$$

where ξ_k is a real number between $|\bar{\boldsymbol{\theta}}_k^{\text{true}} + \phi_n \bar{\boldsymbol{v}}_k|$ and $|\bar{\boldsymbol{\theta}}_k^{\text{true}}|$. Using a similar argument, for $k \in \mathcal{D}_0 \cup D_c$, we have $\boldsymbol{\theta}_{0,k,\cdot}^{\text{true}}(t) = \bar{\boldsymbol{\theta}}_k^{\text{true}}$ and $|\boldsymbol{\theta}_{1,k,\cdot}^{\text{true}}(t)| = 0$, and thus

$$\begin{aligned}
& III_n(\{V(t)\}_{t\in[0,1]}) \\
\geq & \sum_{k\in\mathcal{D}_v} g_{\tau_{k,n}} \left(\left[\int_0^1 \{ |\boldsymbol{\theta}_{0,k,\cdot}^{\text{true}}(t) + \phi_n \boldsymbol{v}_{0,k,\cdot}(t) - \bar{\boldsymbol{\theta}}_k^{\text{true}} - \phi_n \bar{\boldsymbol{v}}_k|^2 + |\boldsymbol{\theta}_{1,k,\cdot}^{\text{true}}(t) + \phi_n \boldsymbol{v}_{1,k,\cdot}(t)|^2 \} dt \right]^{1/2} \right) \\
& - \sum_{k\in\mathcal{D}_v} g_{\tau_{k,n}} \left(\left[\int_0^1 \{ |\boldsymbol{\theta}_{0,k,\cdot}^{\text{true}}(t) - \bar{\boldsymbol{\theta}}_k^{\text{true}}|^2 + |\boldsymbol{\theta}_{1,k,\cdot}^{\text{true}}(t)|^2 \} dt \right]^{1/2} \right),
\end{aligned}$$

where

$$\begin{aligned}
&\left|\sum_{k\in\mathcal{D}_v} g_{\tau_{k,n}}\left(\left[\int_0^1 \{|\boldsymbol{\theta}_{0,k,\cdot}^{\text{true}}(t)+\phi_n\boldsymbol{v}_{0,k,\cdot}(t)-\bar{\boldsymbol{\theta}}_k^{\text{true}}-\phi_n\bar{\boldsymbol{v}}_k|^2+|\boldsymbol{\theta}_{1,k,\cdot}^{\text{true}}(t)+\phi_n\boldsymbol{v}_{1,k,\cdot}(t)|^2\}dt\right]^{1/2}\right)\right.\\
&\left.-\sum_{k\in\mathcal{D}_v} g_{\tau_{k,n}}\left(\left[\int_0^1 \{|\boldsymbol{\theta}_{0,k,\cdot}^{\text{true}}(t)-\bar{\boldsymbol{\theta}}_k^{\text{true}}|^2+|\boldsymbol{\theta}_{1,k,\cdot}^{\text{true}}(t)|^2\}dt\right]^{1/2}\right)\right|\\
\leq\ &\phi_n|V|_{[0,1]}\sum_{k\in\mathcal{D}_v}\sup_{x\in\mathbb{R}}|g'_{\tau_{k,n}}(x)|.
\end{aligned}$$

Note that $\phi_n^2 nb_n = 1+nb_n^5$, $\phi_n nb_n^3 = (nb_n^5)^{1/2}+nb_n^5 = O(1+nb_n^5)$, and $\{(nb_n)^{-1/2}+b_n^2\}(\max_{k\in\mathcal{D}_c\cup\mathcal{D}_{v1}}\lambda_{k,n}+\max_{k\in\mathcal{D}_v}\tau_{k,n}) = O(1)$. Based on the above observations, we can conclude that, for sufficiently large $Q^\star$, the dominate term of $\Psi_n(\{\Theta^{\text{true}}(t)+\phi_nV(t)\}_{t\in[0,1]})-\Psi_n(\{\Theta^{\text{true}}(t)\}_{t\in[0,1]})$ is $\phi_n^2nb_n|V|_{[0,1]}^2(\kappa_2\wedge 1)\tilde{\rho}_{\boldsymbol{M}}>0$, and the result follows. □

Proof. (Theorem 3) Throughout this proof, we rewrite $\Theta_{0,\cdot,\cdot}(t)$ as a sum of its center and an additional time-varying component. To be more specific, let $\mathcal{C}\{\Theta(t)\} := (\bar{\boldsymbol{\theta}}_1,\ldots,\bar{\boldsymbol{\theta}}_p)^\top = (\gamma_{k,j})_{k,j}$ denote the center, then $\mathcal{V}_0\{\Theta(t)\} := \{\boldsymbol{\theta}_{0,1,\cdot}(t)-\bar{\boldsymbol{\theta}}_1,\ldots,\boldsymbol{\theta}_{0,p,\cdot}(t)-\bar{\boldsymbol{\theta}}_p\}^\top = (\iota_{0,k,j})_{0,k,j}$ and $\mathcal{V}_1\{\Theta(t)\} := \{\boldsymbol{\theta}_{1,1,\cdot}(t),\ldots,\boldsymbol{\theta}_{1,p,\cdot}(t)\}^\top = (\iota_{1,k,j})_{1,k,j}$ represent the time-varying component with respect to the center.

We can now rewrite (S3.3) as

$$
\begin{aligned}
&\Lambda_n(\{\mathcal{C}\{\Theta(t)\}\}_{t\in[0,1]},\{\mathcal{V}_0\{\Theta(t)\}\}_{t\in[0,1]},\{\mathcal{V}_1\{\Theta(t)\}\}_{t\in[0,1]})\\
=\;& \int_0^1 \sum_{i=1}^n \left| \boldsymbol{y}_{i,n} - \mathcal{C}\{\Theta(t)\}^\top \boldsymbol{x}_{i,n} - \mathcal{V}_0\{\Theta(t)\}^\top \boldsymbol{x}_{i,n} - \mathcal{V}_1\{\Theta(t)\}^\top \boldsymbol{x}_{i,n}\left(\frac{t_{i,n}-t}{b_n}\right)\right|^2 K\left(\frac{t_{i,n}-t}{b_n}\right) dt\\
&+\sum_{k=1}^p f_{\lambda_{k,n}}(|\boldsymbol{\gamma}_{k,\cdot}|) + \sum_{k=1}^p g_{\tau_{k,n}}\left(\left[\int_0^1 \{|\boldsymbol{\iota}_{0,k,\cdot}(t)|^2 + |\boldsymbol{\iota}_{1,k,\cdot}(t)|^2\} dt\right]^{1/2}\right)\\
&+\mathrm{tr}\left(\boldsymbol{\zeta}^\top \int_0^1 \mathcal{V}_0\{\Theta(t)\}dt\right)\\
=\;& \sum_{j=1}^d \int_0^1 \sum_{i=1}^n \left\{ y_{i,j,n} - \boldsymbol{\gamma}_{\cdot,j}^\top \boldsymbol{x}_{i,n} - \boldsymbol{\iota}_{0,\cdot,j}(t)^\top \boldsymbol{x}_{i,n} - \boldsymbol{\iota}_{1,\cdot,j}(t)^\top \boldsymbol{x}_{i,n}\left(\frac{t_{i,n}-t}{b_n}\right)\right\}^2 K\left(\frac{t_{i,n}-t}{b_n}\right) dt\\
&+\sum_{k=1}^p f_{\lambda_{k,n}}(|\boldsymbol{\gamma}_{k,\cdot}|) + \sum_{k=1}^p g_{\tau_{k,n}}\left(\left[\int_0^1 \{|\boldsymbol{\iota}_{0,k,\cdot}(t)|^2 + |\boldsymbol{\iota}_{1,k,\cdot}(t)|^2\} dt\right]^{1/2}\right)\\
&+\sum_{k=1}^p \sum_{j=1}^d \zeta_{k,j} \int_0^1 \iota_{0,k,j}(t) dt, \qquad \text{(S3.7)}
\end{aligned}
$$

where $\mathrm{tr}(\cdot)$ denotes the trace operator, and $\boldsymbol{\zeta} = (\zeta_{k,j})_{k,j} \in \mathbb{R}^{p\times d}$ represents the Lagrange multiplier that concerns the constraint $\int_0^1 \mathcal{V}_0\{\Theta(t)\}dt = 0$. The minimization problem (S3.7) relates to the normal equations:

$$
\begin{cases}
\partial \Lambda_n([\mathcal{C}\{\Theta(t)\}]_{t\in[0,1]},[\mathcal{V}_0\{\Theta(t)\}]_{t\in[0,1]},[\mathcal{V}_1\{\Theta(t)\}]_{t\in[0,1]})/\partial \gamma_{k,j} = 0;\\
\partial \Lambda_n([\mathcal{C}\{\Theta(t)\}]_{t\in[0,1]},[\mathcal{V}_0\{\Theta(t)\}]_{t\in[0,1]},[\mathcal{V}_1\{\Theta(t)\}]_{t\in[0,1]})/\partial \iota_{0,k,j}(t) = 0;\\
\partial \Lambda_n([\mathcal{C}\{\Theta(t)\}]_{t\in[0,1]},[\mathcal{V}_0\{\Theta(t)\}]_{t\in[0,1]},[\mathcal{V}_1\{\Theta(t)\}]_{t\in[0,1]})/\partial \iota_{1,k,j}(t) = 0.
\end{cases} \qquad \text{(S3.8)}
$$

For $s \in \{0,1\}$, let

$$
A_{k,j,s,n}(t) = \sum_{i=1}^n \left\{ y_{i,j,n} - \boldsymbol{\gamma}_{\cdot,j}^\top \boldsymbol{x}_{i,n} - \boldsymbol{\iota}_{0,\cdot,j}(t)^\top \boldsymbol{x}_{i,n} - \boldsymbol{\iota}_{1,\cdot,j}(t)^\top \boldsymbol{x}_{i,n}\left(\frac{t_{i,n}-t}{b_n}\right)\right\}\left(\frac{t_{i,n}-t}{b_n}\right)^s x_{i,k,n} K\left(\frac{t_{i,n}-t}{b_n}\right)
$$

Then, for any $j = 1, \ldots, d$, (S3.8) can be represented as

$$\begin{cases} \int_0^1 2A_{k,j,0,n}(t)dt = f'_{\lambda_{k,n}}(|\boldsymbol{\gamma}_{k,\cdot}|)\dfrac{\gamma_{k,j}}{|\boldsymbol{\gamma}_{k,\cdot}|}; \\ 2A_{k,j,0,n}(t) = g'_{\tau_{k,n}}\left(\left[\int_0^1 \{|\boldsymbol{\iota}_{0,k,\cdot}(t)|^2 + |\boldsymbol{\iota}_{1,k,\cdot}(t)|^2\}dt\right]^{1/2}\right)\dfrac{\iota_{0,k,j}(t)}{\left[\int_0^1 \{|\boldsymbol{\iota}_{0,k,\cdot}(t)|^2 + |\boldsymbol{\iota}_{1,k,\cdot}(t)|^2\}dt\right]^{1/2}} + \zeta_{k,j}; \\ 2A_{k,j,1,n}(t) = g'_{\tau_{k,n}}\left(\left[\int_0^1 \{|\boldsymbol{\iota}_{0,k,\cdot}(t)|^2 + |\boldsymbol{\iota}_{1,k,\cdot}(t)|^2\}dt\right]^{1/2}\right)\dfrac{\iota_{1,k,j}(t)}{\left[\int_0^1 \{|\boldsymbol{\iota}_{0,k,\cdot}(t)|^2 + |\boldsymbol{\iota}_{1,k,\cdot}(t)|^2\}dt\right]^{1/2}}. \end{cases}$$

Therefore, by observing the constraint that $\int_0^1 \boldsymbol{\iota}_{0,k,\cdot}(t)dt = 0$, we can obtain that

$$4\sum_{j=1}^{d}\left\{\int_0^1 A_{k,j,0,n}(t)dt\right\}^2 = \{f'_{\lambda_{k,n}}(|\boldsymbol{\gamma}_{k,\cdot}|)\}^2, \tag{S3.9}$$

and

$$4\sum_{j=1}^{d}\left\{\int_0^1 A_{k,j,0,n}(t)^2 + A_{k,j,1,n}(t)^2 dt\right\} = \left\{g'_{\tau_{k,n}}\left(\left[\int_0^1 \{|\boldsymbol{\iota}_{0,k,\cdot}(t)|^2 + |\boldsymbol{\iota}_{1,k,\cdot}(t)|^2\}dt\right]^{1/2}\right)\right\}^2 + \sum_{j=1}^{d}\zeta_{k,j}^2. \tag{S3.10}$$

Note that, by the proof of Lemma 6 in Zhou and Wu (2010),

$$\sup_{t\in[0,1]}\left|\sum_{i=1}^{n}\tilde{x}_{i,k,n}\tilde{\boldsymbol{x}}_{i,n}^{\top}\left(\frac{t_{i,n}-t}{b_n}\right)^s K\left(\frac{t_{i,n}-t}{b_n}\right)\right| = O_p(n^{1/2} + nb_n),$$

and by (S3.4) and (S3.5),

$$\sup_{t\in[0,1]}\left|\sum_{i=1}^{n}(x_{i,k,n}\boldsymbol{x}_{i,n}^{\top} - \tilde{x}_{i,k,n}\tilde{\boldsymbol{x}}_{i,n}^{\top})\left(\frac{t_{i,n}-t}{b_n}\right)^s K\left(\frac{t_{i,n}-t}{b_n}\right)\right| \le C\sum_{i=1}^{n}|x_{i,k,n}\boldsymbol{x}_{i,n}^{\top} - \tilde{x}_{i,k,n}\tilde{\boldsymbol{x}}_{i,n}^{\top}| = O_p(1),$$

we have

$$\sup_{t\in[0,1]}\left|\sum_{i=1}^{n}x_{i,k,n}\boldsymbol{x}_{i,n}^{\top}\left(\frac{t_{i,n}-t}{b_n}\right)^s K\left(\frac{t_{i,n}-t}{b_n}\right)\right| = O_p(1+n^{1/2}+nb_n) = O_p(n^{1/2}+nb_n).$$

Then, by Theorem 2 and the Cauchy-Schwarz inequality, we can obtain that

$$\begin{aligned}
&\int_0^1 \left[\sum_{i=1}^n x_{i,k,n}\boldsymbol{x}_{i,n}^\top\{\hat{\boldsymbol{\theta}}_{0,\cdot,j}(t)-\boldsymbol{\theta}_{0,\cdot,j}^{\text{true}}(t)\}\left(\frac{t_{i,n}-t}{b_n}\right)^s K\left(\frac{t_{i,n}-t}{b_n}\right)\right]^2 dt \\
\leq\;& \int_0^1 \left|\sum_{i=1}^n x_{i,k,n}\boldsymbol{x}_{i,n}^\top \left(\frac{t_{i,n}-t}{b_n}\right)^s K\left(\frac{t_{i,n}-t}{b_n}\right)\right|^2 |\hat{\boldsymbol{\theta}}_{0,\cdot,j}(t)-\boldsymbol{\theta}_{0,\cdot,j}^{\text{true}}(t)|^2 dt \\
=\;& O_p\{(n^{1/2}+nb_n)^2\}O_p[\{(nb_n)^{-1/2}+b_n^2\}^2] = O_p(nb_n+n^2b_n^6),
\end{aligned}$$

and

$$\int_0^1 \left[\sum_{i=1}^n x_{i,k,n}\boldsymbol{x}_{i,n}^\top\left\{\hat{\boldsymbol{\theta}}_{1,\cdot,j}(t)-\boldsymbol{\theta}_{1,\cdot,j}^{\text{true}}(t)\right\}\left(\frac{t_{i,n}-t}{b_n}\right)^{s+1} K\left(\frac{t_{i,n}-t}{b_n}\right)\right]^2 dt = O_p(nb_n+n^2b_n^6).$$

In addition, by (S3.6) and Lemma A.1 of Zhang and Wu (2012), we have

$$\left\|\sum_{i=1}^n \tilde{x}_{i,k,n}\tilde{e}_{i,j,n}\left(\frac{t_{i,n}-t}{b_n}\right)^s K\left(\frac{t_{i,n}-t}{b_n}\right)\right\| \leq C(nb_n)^{1/2},$$

and

$$\begin{aligned}
&\left\|\sum_{i=1}^n \tilde{x}_{i,k,n}\tilde{\boldsymbol{x}}_{i,n}^\top\left\{\boldsymbol{\theta}_{0,\cdot,j}^{\text{true}}(t_{i,n})-\boldsymbol{\theta}_{0,\cdot,j}^{\text{true}}(t)-\left(\frac{t_{i,n}-t}{b_n}\right)\boldsymbol{\theta}_{1,\cdot,j}^{\text{true}}(t)\right\}\left(\frac{t_{i,n}-t}{b_n}\right)^s K\left(\frac{t_{i,n}-t}{b_n}\right)\right\| \\
\leq\;& Cb_n^2\{(nb_n)^{1/2}+nb_n\}.
\end{aligned}$$

On the other hand, by a similar argument as in the proof of Theorem 2, we have

$$\left\|\sum_{i=1}^n x_{i,k,n}e_{i,j,n}\left(\frac{t_{i,n}-t}{b_n}\right)^s K\left(\frac{t_{i,n}-t}{b_n}\right)\right\| \leq C(nb_n)^{1/2},$$

and

$$\left\|\sum_{i=1}^n x_{i,k,n}\boldsymbol{x}_{i,n}^\top\left\{\boldsymbol{\theta}_{0,\cdot,j}^{\text{true}}(t_{i,n})-\boldsymbol{\theta}_{0,\cdot,j}^{\text{true}}(t)-\left(\frac{t_{i,n}-t}{b_n}\right)\boldsymbol{\theta}_{1,\cdot,j}^{\text{true}}(t)\right\}\left(\frac{i/n-t}{b_n}\right)^s K\left(\frac{i/n-t}{b_n}\right)\right\| = C(nb_n^3).$$

Combining the above results, we can obtain that

$$
\begin{aligned}
& \int_0^1 \hat{A}^2_{k,j,s,n}(t)dt \\
= & \int_0^1 \sum_{i=1}^n \left[\left\{y_{i,j,n} - \hat{\boldsymbol{\theta}}_{0,\cdot,j}(t)^\top \boldsymbol{x}_{i,n} - \hat{\boldsymbol{\theta}}_{1,\cdot,j}(t)^\top \boldsymbol{x}_{i,n}\left(\frac{t_{i,n}-t}{b_n}\right)\right\}\left(\frac{t_{i,n}-t}{b_n}\right)^s x_{i,k,n} K\left(\frac{t_{i,n}-t}{b_n}\right)\right]^2 dt \\
= & \int_0^1 \sum_{i=1}^n \left(\left[e_{i,j,n} - \boldsymbol{x}_{i,n}^\top\{\hat{\boldsymbol{\theta}}_{0,\cdot,j}(t) - \boldsymbol{\theta}^{\text{true}}_{0,\cdot,j}(t)\} - \boldsymbol{x}_{i,n}^\top\{\hat{\boldsymbol{\theta}}_{1,\cdot,j}(t) - \boldsymbol{\theta}^{\text{true}}_{1,\cdot,j}(t)\}\left(\frac{t_{i,n}-t}{b_n}\right)\right.\right. \\
& \left.\left. + \boldsymbol{x}_{i,n}^\top\left\{\boldsymbol{\theta}^{\text{true}}_{0,\cdot,j}(t_{i,n}) - \boldsymbol{\theta}^{\text{true}}_{0,\cdot,j}(t) - \left(\frac{t_{i,n}-t}{b_n}\right)\boldsymbol{\theta}^{\text{true}}_{1,\cdot,j}(t)\right\}\right]\left(\frac{t_{i,n}-t}{b_n}\right)^s x_{i,k,n} K\left(\frac{t_{i,n}-t}{b_n}\right)\right)^2 dt \\
= & O_p(nb_n + nb_n + n^2 b_n^6 + nb_n^3) = O_p(nb_n + n^2 b_n^6),
\end{aligned}
$$

and, as a result, the left-hand-side of (S3.10) becomes

$$
4\sum_{j=1}^d \left\{\int_0^1 A_{k,j,0,n}(t)^2 + A_{k,j,1,n}(t)^2 dt\right\} = O_p(nb_n + n^2 b_n^6).
$$

If $k \in \mathcal{D}_0 \cup \mathcal{D}_c$, then, by Theorem 2, $\left[\int_0^1 \{|\hat{\boldsymbol{\iota}}_{0,k,\cdot}(t)|^2 + |\hat{\boldsymbol{\iota}}_{1,k,\cdot}(t)|^2\}dt\right]^{1/2} = O_p\{(nb_n)^{-1/2} + b_n^2\}$. Under (P3) and the assumed conditions, we have

$$
\frac{\left\{g'_{\tau_{k,n}}\left(\left[\int_0^1 \{|\boldsymbol{\iota}_{0,k,\cdot}(t)|^2 + |\boldsymbol{\iota}_{1,k,\cdot}(t)|^2\}dt\right]^{1/2}\right)\right\}^2}{nb_n + n^2 b_n^6} \to \infty.
$$

Therefore, the solution will be at the point where the differentiability does not hold, which leads to

$$
\text{pr}\left\{\max_{k \in \mathcal{D}_0 \cup \mathcal{D}_c} \sup_{t \in [0,1]} |\hat{\boldsymbol{\iota}}_{0,k,\cdot}(t)| = 0 \text{ and } \max_{k \in \mathcal{D}_0 \cup \mathcal{D}_c} \sup_{t \in [0,1]} |\hat{\boldsymbol{\iota}}_{1,k,\cdot}(t)| = 0\right\} \to 1.
$$

For the second claim, note that

$$\begin{aligned}
&\int_0^1 \hat{A}_{k,j,0,n}(t)dt \\
=\ & \int_0^1 \sum_{i=1}^n \left\{ y_{i,j,n} - \hat{\boldsymbol{\theta}}_{0,\cdot,j}(t)^\top \boldsymbol{x}_{i,n} - \hat{\boldsymbol{\theta}}_{1,\cdot,j}(t)^\top \boldsymbol{x}_{i,n} \left(\frac{t_{i,n}-t}{b_n} \right) \right\} x_{i,k,n} K \left(\frac{t_{i,n}-t}{b_n} \right) dt \\
=\ & \int_0^1 \sum_{i=1}^n \Bigg[e_{i,j,n} - \boldsymbol{x}_{i,n}^\top \{\hat{\boldsymbol{\theta}}_{0,\cdot,j}(t) - \boldsymbol{\theta}_{0,\cdot,j}^{\text{true}}(t)\} - \boldsymbol{x}_{i,n}^\top \{\hat{\boldsymbol{\theta}}_{1,\cdot,j}(t) - \boldsymbol{\theta}_{1,\cdot,j}^{\text{true}}(t)\} \left(\frac{t_{i,n}-t}{b_n} \right) \\
& \qquad + \boldsymbol{x}_{i,n}^\top \left\{ \boldsymbol{\theta}_{0,\cdot,j}^{\text{true}}(t_{i,n}) - \boldsymbol{\theta}_{0,\cdot,j}^{\text{true}}(t) - \left(\frac{t_{i,n}-t}{b_n} \right) \boldsymbol{\theta}_{1,\cdot,j}^{\text{true}}(t) \right\} \Bigg] x_{i,k,n} K \left(\frac{t_{i,n}-t}{b_n} \right) dt \\
=\ & O_p[(nb_n)^{1/2} + (nb_n + n^{1/2})\{(nb_n)^{-1/2} + b_n^2\} + nb_n^3] = O_p\{nb_n^3 + (nb_n)^{1/2}\}.
\end{aligned}$$

Thus, by Theorem 2, $|\hat{\boldsymbol{\gamma}}_{k,\cdot}| = O_p\{(nb_n)^{-1/2} + b_n^2\}$ holds for $k \in \mathcal{D}_0$. On the other hand, note that

$$\frac{f'_{\lambda_{k,n}}(|\boldsymbol{\gamma}_{k,\cdot}|)}{nb_n^3 + (nb_n)^{1/2}} \to \infty,$$

the second claim follows by (S3.9) and a similar argument. □

Bibliography

FAN, J. AND LI, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, **96**, 1348–1360.

ZHANG, T. AND WU, W. B. (2012). Inference of time-varying regression models. *The Annals of Statistics*, **40**, 1376–1402.

ZHOU, Z. AND WU, W. B. (2010). Simultaneous inference of linear models with time varying coefficients. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **72**, 513–531.