Penalized spline estimation in varying coefficient models with censored data

Supplementary material

HENDRICKX, Kim; JANSSEN, Paul & VERHASSELT, Anneleen (2018) Penalized spline estimation in varying coefficient models with censored data. In: TEST, 27(4), p. 871-895.

DOI: 10.1007/s11749-017-0574-y
Handle: http://hdl.handle.net/1942/25678

Noname manuscript No.
(will be inserted by the editor)

Penalized spline estimation in varying coefficient models with censored data

Supplementary Material

K. Hendrickx[1] · **P. Janssen**[1] · **A. Verhasselt**[1]

Received: date / Accepted: date

We prove the asymptotic results (Theorem 1 and Theorem 2) of Section 5 of the manuscript [3]. Throughout, sections refer to the main manuscript.

1 Definitions and properties

Definition 1 For a real valued matrix $\mathbf{A}$ of dimension $m_A \times n_A$, the 2-norm of $\mathbf{A}$ is given by $\|\mathbf{A}\|_2 = \sup_{\mathbf{x}\neq 0} \frac{\|\mathbf{A}\mathbf{x}\|_2}{\|\mathbf{x}\|_2}$, with $\mathbf{x} \in I\!R^{n_A}$ and $\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^{n_A} x_i^2}$. This norm is equal to $\sqrt{\zeta_{\max}(\mathbf{A}'\mathbf{A})}$ where $\zeta_{\max}$ is the largest eigenvalue of $\mathbf{A}'\mathbf{A}$.

Definition 2 For sequences of positive numbers r_n and s_n, $r_n \lesssim s_n$ means that $s_n^{-1} r_n$ is bounded and $r_n \asymp s_n$ means that $s_n^{-1} r_n$ and $r_n^{-1} s_n$ are bounded.

Definition 3 For a real valued function f on $\mathcal{U}$ and a vector valued function $\mathbf{g} = (g_1, ..., g_d)$ on $\mathcal{U}^d$, the L_∞-norm is given by:

$$\|f\|_\infty = \sup_{u \in \mathcal{U}} \mid f(u) \mid, \quad \|\mathbf{g}\|_\infty = \max_{1 \leq p \leq d} \|g_p\|_\infty.$$

Our estimation technique relies on properties of B-splines. For a detailed description of B-splines we refer to [2] or [6].

Property 1 $B_{pl}(u_p; q_p) \geq 0$ and $\sum_{l=1}^{m_p} B_{pl}(u_p; q_p) = 1$.

K. Hendrickx
Hasselt University, I-BioStat, Agoralaan, B3590 Diepenbeek, Belgium.
Tel.: +32 11 26 82 81
E-mail: kim.hendrickx@uhasselt.be

[1]Hasselt University

Property 2 There exists positive constants N_7, N_8 and coefficients $\alpha_{pl} \in \mathrm{R}$ such that

$$m_p^{-1} N_7 \sum_{l=1}^{m_p} \alpha_{pl}^2 \leq \int_{\mathcal{U}} \{\sum_{l=1}^{m_p} \alpha_{pl} B_{pl}(u_p; q_p)\}^2 du \leq m_p^{-1} N_8 \sum_{l=1}^{m_p} \alpha_{pl}^2.$$

Property 3 $\int_{\mathcal{U}} B_{pl}(u; q_p) du = O(m_p^{-1})$.

Property 4 $\|g\|_\infty \lesssim m_p^{-1/2} \|g\|_{L_2}$ for $g \in \mathrm{G}(q_p, \boldsymbol{\xi}_p)$, where $\mathrm{G}(q_p, \boldsymbol{\xi}_p)$ is the space of spline functions on $\mathcal{U}_p$ with fixed degree q_p and knot sequence $\boldsymbol{\xi}_p$.

We use as notations $\hat{\boldsymbol{\alpha}}_j, \boldsymbol{\alpha}_j^*$ and $\tilde{\boldsymbol{\alpha}}_j$ for methods $j = 1, 2$ (described in Section 4 of [3]), when we replace $\mathbf{Y}$ in expression

$$\hat{\boldsymbol{\alpha}} = \left(\mathbf{R}'\mathbf{R} + \mathbf{Q}_{\boldsymbol{\lambda}}\right)^{-1} \mathbf{R}'\mathbf{Y}.$$

by $\hat{\mathbf{Y}}_j^* = (\hat{Y}_{j1}^*, \ldots, \hat{Y}_{jn}^*)'$, $\mathbf{Y}_j^* = (Y_{j1}^*, \ldots, Y_{jn}^*)'$, and $\mathbf{M}_j = (M_{j1}, \ldots, M_{jn})'$ with $M_{ji} = E(Y_{ji}^* | \mathbf{U}_i, X_i)$ for $i = 1, \ldots, n$ respectively. Similar notations hold for $\hat{\boldsymbol{\beta}}_j = (\hat{\beta}_{j1}, \ldots, \hat{\beta}_{jd})'$, $\boldsymbol{\beta}_j^* = (\beta_{j1}^*, \ldots, \beta_{jd}^*)'$ and $\tilde{\boldsymbol{\beta}}_j = (\tilde{\beta}_{j1}, \ldots, \tilde{\beta}_{jd})'$.

1.1 Proof of Theorem 1, Part 1

The proof of the first result stated in Theorem 1 relies on the maximal distance between the Y_{1i}^* and $\hat{Y}_{1i}^*$, derived in Lemma 1.

Lemma 1 $\max_{1 \leq i \leq n} \mid \hat{Y}_{1i}^* - Y_{1i}^* \mid =$

$$O_p\left(\sup_{\mathbf{u},\mathbf{x}} \left\{ \tau_1(\mathbf{u}, \mathbf{x}) \sup_{t \leq \tau_1(\mathbf{u},\mathbf{x})} |\hat{G}(t|\mathbf{u}, \mathbf{x}) - G(t|\mathbf{u}, \mathbf{x})| + \kappa(\mathbf{u}, \mathbf{x}) \right\}\right).$$

Proof (Proof of Lemma 1)
Since $\mid \hat{Y}_{1i}^* - Y_{1i}^* \mid =$

$$\mid \hat{Y}_{1i}^* - Y_{1i}^* \mid 1_{\{Z_i \leq \tau_1(\mathbf{U}_i, \mathbf{X}_i)\}} + \mid \hat{Y}_{1i}^* - Y_{1i}^* \mid 1_{\{Z_i > \tau_1(\mathbf{U}_i, \mathbf{X}_i)\}},$$

we consider two cases and prove the following results,

$$\max_{1 \leq i \leq n} \left\{ \mid \hat{Y}_{1i}^* - Y_{1i}^* \mid 1_{\{Z_i \leq \tau_1(\mathbf{U}_i, \mathbf{X}_i)\}} \right\} \lesssim \sup_{\mathbf{u},\mathbf{x}} \left(\tau_1(\mathbf{u}, \mathbf{x}) \sup_{t \leq \tau_1(\mathbf{u},\mathbf{x})} |\hat{G}(t|\mathbf{u}, \mathbf{x}) - G(t|\mathbf{u}, \mathbf{x})| \right), \quad (1)$$

$$\max_{1 \leq i \leq n} \left\{ \mid \hat{Y}_{1i}^* - Y_{1i}^* \mid 1_{\{Z_i > \tau_1(\mathbf{U}_i, \mathbf{X}_i)\}} \right\} \lesssim \sup_{\mathbf{u},\mathbf{x}} \kappa(\mathbf{u}, \mathbf{x}). \quad (2)$$

For (1) we start by the triangle inequality,

$$\begin{aligned}
&\mid \hat{Y}^*_{1i} - Y^*_{1i} \mid 1_{\{Z_i \le \tau_1(\mathbf{U}_i,\mathbf{X}_i)\}} \le\mid \Delta_i\{\hat{\varphi}_1(\mathbf{U}_i,\mathbf{X}_i,Z_i) - \varphi_1(\mathbf{U}_i,\mathbf{X}_i,Z_i)\} \\
&\qquad\qquad\qquad + (1-\Delta_i)\{\hat{\psi}_1(\mathbf{U}_i,\mathbf{X}_i,Z_i) - \psi_1(\mathbf{U}_i,\mathbf{X}_i,Z_i)\} \mid \\
&\le\mid \hat{\varphi}_1(\mathbf{U}_i,\mathbf{X}_i,Z_i) - \varphi_1(\mathbf{U}_i,\mathbf{X}_i,Z_i) \mid + \mid \hat{\psi}_1(\mathbf{U}_i,\mathbf{X}_i,Z_i) - \psi_1(\mathbf{U}_i,\mathbf{X}_i,Z_i) \mid .
\end{aligned}$$

We derive the order bound for $\mid \hat{\varphi}_1(\mathbf{U}_i,\mathbf{X}_i,Z_i) - \varphi_1(\mathbf{U}_i,\mathbf{X}_i,Z_i) \mid$, similar result holds if we replace φ_1 and $\hat{\varphi}_1$ by ψ_1 and $\hat{\psi}_1$ respectively.

$$\begin{aligned}
&|\hat{\varphi}_1(\mathbf{U}_i,\mathbf{X}_i,Z_i) - \varphi_1(\mathbf{U}_i,\mathbf{X}_i,Z_i)| \\
&\le \left|(1+\gamma)\left\{\int_0^{Z_i} \frac{1}{\hat{G}(t|\mathbf{U}_i,\mathbf{X}_i)}dt - \int_0^{Z_i} \frac{1}{G(t|\mathbf{U}_i,\mathbf{X}_i)}dt\right\}\right| \\
&\qquad + \left|\frac{\gamma Z_i}{\hat{G}(Z_i|\mathbf{U}_i,\mathbf{X}_i)} - \frac{\gamma Z_i}{G(Z_i|\mathbf{U}_i,\mathbf{X}_i)}\right| \\
&\le \left|(1+\gamma)\int_0^{Z_i} \frac{\hat{G}(t|\mathbf{U}_i,\mathbf{X}_i) - G(t|\mathbf{U}_i,\mathbf{X}_i)}{G(t|\mathbf{U}_i,\mathbf{X}_i)\hat{G}(t|\mathbf{U}_i,\mathbf{X}_i)}dt\right| \\
&\qquad + \left|\frac{\gamma Z_i\{\hat{G}(Z_i|\mathbf{U}_i,\mathbf{X}_i) - G(Z_i|\mathbf{U}_i,\mathbf{X}_i)\}}{G(Z_i|\mathbf{U}_i,\mathbf{X}_i)\hat{G}(Z_i|\mathbf{U}_i,\mathbf{X}_i)}\right| \\
&\le |1+\gamma| \sup_{t\le\tau_1(\mathbf{U}_i,\mathbf{X}_i)} \Big\{\mid \hat{G}(t|\mathbf{U}_i,\mathbf{X}_i) - G(t|\mathbf{U}_i,\mathbf{X}_i) \mid\Big\} \\
&\qquad\qquad \times \int_0^{\tau_1(\mathbf{U}_i,\mathbf{X}_i)} \frac{G(t|\mathbf{U}_i,\mathbf{X}_i)}{\hat{G}(t|\mathbf{U}_i,\mathbf{X}_i)} \frac{1}{G(t|\mathbf{U}_i,\mathbf{X}_i)^2}dt \\
&\quad + |\gamma|\tau_1(\mathbf{U}_i,\mathbf{X}_i) \sup_{t\le\tau_1(\mathbf{U}_i,\mathbf{X}_i)} \Big\{\mid \hat{G}(t|\mathbf{U}_i,\mathbf{X}_i) - G(t|\mathbf{U}_i,\mathbf{X}_i) \mid\Big\} \\
&\qquad\qquad \times \sup_{t\le\tau_1(\mathbf{U}_i,\mathbf{X}_i)} \left\{\frac{1}{G(t|\mathbf{U}_i,\mathbf{X}_i)^2} \frac{G(t|\mathbf{U}_i,\mathbf{X}_i)}{\hat{G}(t|\mathbf{U}_i,\mathbf{X}_i)}\right\}.
\end{aligned}$$

From the uniform convergence of $\hat{G}$ we have:

$$\sup_{t\le\tau_1(\mathbf{U}_i,\mathbf{X}_i)} \frac{G(t|\mathbf{U}_i,\mathbf{X}_i)}{\hat{G}(t|\mathbf{U}_i,\mathbf{X}_i)} = 1 + o_p(1).$$

Also $\inf_{t\le\tau_1(\mathbf{U}_i,\mathbf{X}_i)}\{G(t|\mathbf{U}_i,\mathbf{X}_i)\} > 0$, therefore,

$$\begin{aligned}
&\mid \hat{\varphi}_1(\mathbf{U}_i,\mathbf{X}_i,Z_i) - \varphi_1(\mathbf{U}_i,\mathbf{X}_i,Z_i) \mid \\
&\qquad = O_p\Big(\tau_1(\mathbf{U}_i,\mathbf{X}_i) \sup_{t\le\tau_1(\mathbf{U}_i,\mathbf{X}_i)} \mid \hat{G}(t|\mathbf{U}_i,\mathbf{X}_i) - G(t|\mathbf{U}_i,\mathbf{X}_i) \mid \Big).
\end{aligned}$$

For (2) we have

$$
\begin{aligned}
&E\{\mid \hat{Y}^*_{1i} - Y^*_{1i} \mid 1_{\{Z_i > \tau_1(\mathbf{U}_i, \mathbf{X}_i)\}}\} \\
&\quad \le E\left[E\left\{ \max_{\phi=\varphi_1,\psi_1} 1_{\{Z_i > \tau_1(\mathbf{U}_i, \mathbf{X}_i)\}} \mid Z_i - \phi(U_i, \mathbf{X}_i, Z_i) \mid \Big| \mathbf{U}_i, \mathbf{X}_i \right\} \right] \\
&\quad \le \sup_{\mathbf{u},\mathbf{x}} \kappa(\mathbf{u}, \mathbf{x}).
\end{aligned}
$$

By combining (1) and (2), the result of Lemma 1 follows.

Proof (Proof of Theorem 1, Part 1)
Since

$$
\|\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1\|_{L_2} \le \|\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}^*_1\|_{L_2} + \|\boldsymbol{\beta}^*_1 - \tilde{\boldsymbol{\beta}}_1\|_{L_2} + \|\tilde{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1\|_{L_2},
$$

the result follows by showing that

$$
\begin{aligned}
&\|\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}^*_1\|_{L_2} && (3)\\
&\quad = O_p\left(\sup_{\mathbf{u},\mathbf{x}} \left\{ \tau_1(\mathbf{u},\mathbf{x}) \sup_{t \le \tau_1(\mathbf{u},\mathbf{x})} |\hat{G}(t|\mathbf{u},\mathbf{x}) - G(t|\mathbf{u},\mathbf{x})| + \kappa(\mathbf{u},\mathbf{x}) \right\} \right), \\
&\|\boldsymbol{\beta}^*_1 - \tilde{\boldsymbol{\beta}}_1\|_{L_2} = O_p\left(n^{-1/2} m_{\max}^{1/2} \right), && (4)\\
&\|\tilde{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1\|_{L_2} = O_p\left(n^{-1} m_{\max}^{3/2} \lambda_{\max} + \rho_n \right). && (5)
\end{aligned}
$$

We start with the proof of (3). By Property 2 it suffices to show that

$$
\begin{aligned}
&\|\hat{\boldsymbol{\alpha}}_1 - \boldsymbol{\alpha}^*_1\|_2 = \\
&O_p\left(m_{\max}^{1/2} \left(\sup_{\mathbf{u},\mathbf{x}} \left\{ \tau_1(\mathbf{u},\mathbf{x}) \sup_{t \le \tau_1(\mathbf{u},\mathbf{x})} |\hat{G}(t|\mathbf{u},\mathbf{x}) - G(t|\mathbf{u},\mathbf{x})| + \kappa(\mathbf{u},\mathbf{x}) \right\} \right) \right).
\end{aligned}
$$

From [1] we have

$$
\begin{aligned}
&\hat{\boldsymbol{\alpha}}_1 - \boldsymbol{\alpha}^*_1 \\
&= \left\{ (\mathbf{R}'\mathbf{R})^{-1} - (\mathbf{R}'\mathbf{R})^{-1}\mathbf{Q}_{\boldsymbol{\lambda}}(\mathbf{R}'\mathbf{R})^{-1} + o_p(n^{-1}m_{\max}^{3/2}\lambda_{\max})(\mathbf{R}'\mathbf{R})^{-1} \right\} \\
&\qquad\qquad \times \sum_{i=1}^n \mathbf{R}_i(\hat{Y}^*_{1i} - Y^*_{1i}) \\
&= \hat{\boldsymbol{\alpha}}_{1,reg} - \boldsymbol{\alpha}^*_{reg} - \left\{ (\mathbf{R}'\mathbf{R})^{-1}\mathbf{Q}_{\boldsymbol{\lambda}}(\mathbf{R}'\mathbf{R})^{-1} + o_p(n^{-1}m_{\max}^{3/2}\lambda_{\max})(\mathbf{R}'\mathbf{R})^{-1} \right\} \\
&\qquad\qquad \times \sum_{i=1}^n \mathbf{R}_i(\hat{Y}^*_{1i} - Y^*_{1i}) \\
&= \left\{ 1 - (\mathbf{R}'\mathbf{R})^{-1}\mathbf{Q}_{\boldsymbol{\lambda}} + o_p(n^{-1}m_{\max}^{3/2}\lambda_{\max}) \right\} \left(\hat{\boldsymbol{\alpha}}_{1,reg} - \boldsymbol{\alpha}^*_{reg} \right),
\end{aligned}
$$

where $\hat{\boldsymbol{\alpha}}_{1,reg}$ and $\boldsymbol{\alpha}^*_{reg}$ denote the regular B-spline estimator (i.e. $\lambda_0 = \ldots = \lambda_d = 0$). Consequently

$$\begin{aligned}&\|\hat{\boldsymbol{\alpha}}_1 - \boldsymbol{\alpha}_1^*\|_2 \\ &\qquad \leq \left\{1 + \|(\mathbf{R}'\mathbf{R})^{-1}\|_2 \|\mathbf{Q}_{\boldsymbol{\lambda}}\|_2 + o_p(n^{-1}m_{\max}^{3/2}\lambda_{\max})\right\} \|\hat{\boldsymbol{\alpha}}_{1,reg} - \boldsymbol{\alpha}^*_{1,reg}\|_2.\end{aligned}$$

From Lemma 1 in [1] we know that except on an event whose probability tends to zero, $\|(\mathbf{R}'\mathbf{R})^{-1}\|_2 \|\mathbf{Q}_{\boldsymbol{\lambda}}\|_2 = O_p(n^{-1}m_{\max}^{3/2}\lambda_{\max})$. Furthermore,

$$\begin{aligned}&\|\hat{\boldsymbol{\alpha}}_{1,reg} - \boldsymbol{\alpha}^*_{1,reg}\|_2^2 = (\hat{\mathbf{Y}}_1^* - \mathbf{Y}_1^*)'\mathbf{R}(\mathbf{R}'\mathbf{R})^{-1}(\mathbf{R}'\mathbf{R})^{-1}\mathbf{R}'(\hat{\mathbf{Y}}_1^* - \mathbf{Y}_1^*) \\ &= (n^{-1}m_{\max})^2(\hat{\mathbf{Y}}_1^* - \mathbf{Y}_1^*)'\mathbf{R}\big(n^{-1}m_{\max}\mathbf{R}'\mathbf{R}\big)^{-1}\big(n^{-1}m_{\max}\mathbf{R}'\mathbf{R}\big)^{-1}\mathbf{R}'(\hat{\mathbf{Y}}_1^* - \mathbf{Y}_1^*).\end{aligned}$$

and since all eigenvalues of $n^{-1}m_{\max}\mathbf{R}'\mathbf{R}$ fall between positive constants, we have $\|n^{-1}m_{\max}\mathbf{R}'\mathbf{R}\|_2 \asymp 1$ and thus

$$\begin{aligned}&\|\hat{\boldsymbol{\alpha}}_{1,reg} - \boldsymbol{\alpha}^*_{1,reg}\|_2^2 = (\hat{\mathbf{Y}}_1^* - \hat{\mathbf{Y}}_1^*)'\mathbf{R}(\mathbf{R}'\mathbf{R})^{-1}(\mathbf{R}'\mathbf{R})^{-1}\mathbf{R}'(\hat{\mathbf{Y}}_1^* - \mathbf{Y}_1^*) \\ &\asymp n^{-1}m_{\max}(\hat{\mathbf{Y}}_1^* - \mathbf{Y}_1^*)'(\hat{\mathbf{Y}}_1^* - \mathbf{Y}_1^*) \\ &\lesssim m_{\max}\left(\sup_{\mathbf{u},\mathbf{x}}\left\{\tau_1(\mathbf{u},\mathbf{x}) \sup_{t \leq \tau_1(\mathbf{u},\mathbf{x})} |\hat{G}(t|\mathbf{u},\mathbf{x}) - G(t|\mathbf{u},\mathbf{x})| + \kappa(\mathbf{u},\mathbf{x})\right\}\right)^2.\end{aligned}$$

In the last step, we use the result of Lemma 1 and the inequality

$$\sqrt{(\hat{\mathbf{Y}}_1^* - \mathbf{Y}_1^*)'(\hat{\mathbf{Y}}_1^* - \mathbf{Y}_1^*)} = \|\hat{\mathbf{Y}}_1^* - \mathbf{Y}_1^*\|_2 \leq \sqrt{n}\max_{1\leq i\leq n}|\hat{Y}_{1i}^* - Y_{1i}^*|.$$

We continue with the proof of (4). Using similar arguments as is the proof of (3), we have

$$\begin{aligned}&\|\boldsymbol{\alpha}_1^* - \tilde{\boldsymbol{\alpha}}_1\|_2 \\ &\leq \left\{1 + \|(\mathbf{R}'\mathbf{R})^{-1}\|_2 \|\mathbf{Q}_{\boldsymbol{\lambda}}\|_2 + o_p(n^{-1}m_{\max}^{3/2}\lambda_{\max})\right\} \|\boldsymbol{\alpha}^*_{1,reg} - \tilde{\boldsymbol{\alpha}}_{1,reg}\|_2, \end{aligned} \tag{6}$$

and

$$\begin{aligned}&\|\boldsymbol{\alpha}^*_{1,reg} - \tilde{\boldsymbol{\alpha}}_{1,reg}\|_2^2 \\ &= (n^{-1}m_{\max})^2(\mathbf{Y}_1^* - \mathbf{M}_1)'\mathbf{R}(n^{-1}m_{\max}\mathbf{R}'\mathbf{R})^{-1}(n^{-1}m_{\max}\mathbf{R}'\mathbf{R})^{-1}\mathbf{R}'(\mathbf{Y}_1^* - \mathbf{M}_1).\end{aligned}$$

By Assumption A.3,

$$
\begin{aligned}
&E\left\{(\mathbf{Y}_1^*-\mathbf{M}_1)'\mathbf{R}\mathbf{R}'(\mathbf{Y}_1^*-\mathbf{M}_1)\right\}\\
&= E\left[\left\{\sum_{i=1}^{n}\mathbf{R}_i(Y_{1i}^*-M_{1i})\right\}'\left\{(\sum_{i=1}^{n}\mathbf{R}_i(Y_{1i}^*-M_{1i})\right\}\right]\\
&= E\left\{\sum_{p,l}\sum_{i,j=1}^{n}X_{ip}X_{jp}B_{pl}(U_{ip};q_p)B_{pl}(U_{jp};q_p)(Y_{1i}^*-M_{1i})(Y_{1j}^*-M_{1j})\right\}\\
&\lesssim \sum_{p,l}\left[\sum_{i=1}^{n}E\left\{B_{pl}^2(U_{ip};q_p)^2(Y_{1i}^*-M_{1i})^2\right\}\right.\\
&\qquad\left.+\sum_{i\neq j}E\left\{B_{pl}(U_{ip};q_p)B_{pl}(U_{jp};q_p)(Y_{1i}^*-M_{1i})(Y_{1j}^*-M_{1j})\right\}\right].
\end{aligned}
$$

By the independence of the observations, Assumption A.5 and Properties 2 and 3 of B-splines it follows that, using the law of the total expectation,

$$
\begin{aligned}
&E\left\{B_{pl}^2(U_{ip};q_p)(Y_{1i}^*-M_{1i})^2\right\}\lesssim E\{B_{pl}^2(U_{ip};q_p)\}\lesssim m_p^{-1}=O(m_{\max}^{-1}),\\
&E\{B_{pl}(U_{ip};q_p)B_{pl}(U_{jp};q_p)(Y_{1i}^*-M_{1i})(Y_{1j}^*-M_{1j})\}\\
&\qquad = E\{B_{pl}(U_{ip};q_p)(Y_{1i}^*-M_{1i})\}E\{B_{pl}(U_{jp};q_p)(Y_{1j}^*-M_{1j})\}=0.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
E\left\{(\mathbf{Y}_1^*-\mathbf{M}_1)'\mathbf{R}\mathbf{R}'(\mathbf{Y}_1^*-\mathbf{M}_1)\right\}&=O(n),\\
(\mathbf{Y}_1^*-\mathbf{M}_1)'\mathbf{R}\mathbf{R}'(\mathbf{Y}_1^*-\mathbf{M}_1)&=O_p(n),
\end{aligned}
$$

such that

$$
\|\boldsymbol{\alpha}_{1,reg}^*-\tilde{\boldsymbol{\alpha}}_{1,reg}\|_2^2=O_p\left(n^{-1}m_{\max}^2\right). \tag{7}
$$

Combining (6) and (7) gives,

$$
\begin{aligned}
\|\boldsymbol{\alpha}_1^*-\tilde{\boldsymbol{\alpha}}_1\|_2^2&=O_p\left(n^{-1}m_{\max}^2\left(1+n^{-1}m_{\max}^{3/2}\lambda_{\max}\right)^2\right)=O_p(n^{-1}m_{\max}^2),\\
\|\boldsymbol{\beta}_1^*-\tilde{\boldsymbol{\beta}}_1\|_{L_2}^2&\asymp\frac{1}{m_{\max}}\|\boldsymbol{\alpha}_1^*-\tilde{\boldsymbol{\alpha}}_1\|_2^2=O_p\left(n^{-1}m_{\max}\right),
\end{aligned}
$$

where we use Assumption A.6 and B-spline Property 2. From the proof of Theorem 1 in [1], we have,

$$
\|\tilde{\boldsymbol{\beta}}_1-\boldsymbol{\beta}\|_{L_2}=O_p\left(n^{-1}m_{\max}^{3/2}\lambda_{\max}+\rho_n\right),
$$

and (5) follows immediately.

1.2 Proof of Theorem 1, Part 2

To prove Part 2 of Theorem 1, we can repeat the proof of Part 1 of Theorem 1 but now using Lemma 2 instead of Lemma 1 giving the maximal distance between Y_2^* and $\hat{Y}_2^*$. The proof of Lemma 2 needs two further lemmas: Lemma 3 on the uniform consistency of the initial estimators $\hat{m}_1$ and $\hat{\sigma}_1$ as estimators for m and σ; and Lemma 4 on the uniform consistency of $\hat{F}$ as estimator of F. The proof of Lemma 3 is included, that of Lemma 4 follows along the lines of a similar result (in the kernel estimation context) in [7]. The details of the proof of Lemma 4 are not given but we do give and prove, in Lemma 5, the key result that is needed to modify their result to our P-spline setting.

Lemma 2 *If Assumptions A, B and C hold,*

$$\max_{1\le i\le n} \mid \hat{Y}_{2i}^* - Y_{2i}^* \mid = O_p(a_n) = o_p(1),$$

where $a_n = n^{-1/2}(\log n)^{1/2} + n^{-1}m_{max}^{3/2}\lambda_{max} + \rho_n + m_{max}^{-1/2}\left(\sup_{\mathbf{u},\mathbf{x}}\left\{\tau_1(\mathbf{u},\mathbf{x})\sup_{t\le\tau_1(\mathbf{u},\mathbf{x})}|\hat{G}(t|\mathbf{u},\mathbf{x}) - G(t|\mathbf{u},\mathbf{x})| + \kappa(\mathbf{u},\mathbf{x}) + \kappa_\sigma(\mathbf{u},\mathbf{x})\right\}\right)$.

Method 2 uses (8) and (10) as initial estimates for $m(\mathbf{u},\mathbf{x})$ and $\sigma^2(\mathbf{u},\mathbf{x})$. We therefore need, in the proof of Theorem 1, Part 2, the consistency results given in Lemma 3.

Lemma 3 *Under Assumptions A, B.1 and B.2, we have*

$$(a)\sup_{\mathbf{u},\mathbf{x}}|\hat{m}_1(\mathbf{u},\mathbf{x}) - m(\mathbf{u},\mathbf{x})| = O_p\Big(n^{-1/2} + n^{-1}m_{max}^{3/2}\lambda_{max} + \rho_n + m_{max}^{-1/2}\big(\sup_{\mathbf{u},\mathbf{x}}\big\{\tau_1(\mathbf{u},\mathbf{x})\sup_{t\le\tau_1(\mathbf{u},\mathbf{x})}|\hat{G}(t|\mathbf{u},\mathbf{x}) - G(t|\mathbf{u},\mathbf{x})| + \kappa(\mathbf{u},\mathbf{x})\big\}\big)\Big).$$

$$(b)\max_{1\le i\le n} \mid \hat{Y}_{1i,\sigma^2}^* - Y_{1i,\sigma^2}^* \mid = O_p\Big(n^{-1/2} + n^{-1}m_{max}^{3/2}\lambda_{max} + \rho_n + \sup_{\mathbf{u},\mathbf{x}}\big\{\tau_1(\mathbf{u},\mathbf{x})\sup_{t\le\tau_1(\mathbf{u},\mathbf{x})}|\hat{G}(t|\mathbf{u},\mathbf{x}) - G(t|\mathbf{u},\mathbf{x})| + m_{max}^{-1/2}\kappa(\mathbf{u},\mathbf{x}) + \kappa_\sigma(\mathbf{u},\mathbf{x})\big\}\Big),$$

where $Y_{1i,\sigma^2}^* = \dfrac{\Delta_i(Z_i - m(\mathbf{U}_i,\mathbf{X}_i))^2}{G(Z_i|\mathbf{U}_i,\mathbf{X}_i)}$.

$$(c)\sup_{\mathbf{u},\mathbf{x}}|\hat{\sigma}_1(\mathbf{u},\mathbf{x}) - \sigma(\mathbf{u},\mathbf{x})| = O_p\Big(n^{-1/2} + n^{-1}m_{max}^{3/2}\lambda_{max} + \rho_n + m_{max}^{-1/2}\Big(\sup_{\mathbf{u},\mathbf{x}}\big\{\tau_1(\mathbf{u},\mathbf{x})\sup_{t\le\tau_1(\mathbf{u},\mathbf{x})}|\hat{G}(t|\mathbf{u},\mathbf{x}) - G(t|\mathbf{u},\mathbf{x})| + m_{max}^{-1/2}\kappa(\mathbf{u},\mathbf{x}) + \kappa_\sigma(\mathbf{u},\mathbf{x})\big\}\Big)\Big).$$

Proof (Proof of Lemma 3(a))
Since the X_p are bounded (see Assumption A.3), we have,

$$\sup_{\mathbf{u},\mathbf{x}} |\hat{m}_1(\mathbf{u},\mathbf{x}) - m(\mathbf{u},\mathbf{x})| \lesssim \sum_{p=1}^{d} \|\hat{\beta}_{1p} - \beta_p\|_{L_\infty}$$
$$\leq \sum_{p=1}^{d} \|\hat{\beta}_{1p} - \tilde{\beta}_{1p}\|_{L_\infty} + \sum_{p=1}^{d} \|\tilde{\beta}_{1p} - \beta_p\|_{L_\infty}.$$

By Property 4, we have $\|\hat{\beta}_{1p} - \tilde{\beta}_{1p}\|_{L_\infty} \lesssim m_{\max}^{-1/2}\|\hat{\beta}_{1p} - \tilde{\beta}_{1p}\|_{L_2}$. Using the intermediate results stated in the proof of Theorem 1, part 1, we obtain that

$$\|\hat{\beta}_{1p} - \tilde{\beta}_{1p}\|_{L_\infty} = O_p\Bigg(n^{-1/2} + m_{\max}^{-1/2}\left(\sup_{\mathbf{u},\mathbf{x}}\left\{\tau_1(\mathbf{u},\mathbf{x}) \sup_{t\leq\tau_1(\mathbf{u},\mathbf{x})} |\hat{G}(t|\mathbf{u},\mathbf{x}) - G(t|\mathbf{u},\mathbf{x})| + \kappa(\mathbf{u},\mathbf{x})\right\}\right)\Bigg).$$

By Lemma A.10 of [5], we have

$$\|\tilde{\boldsymbol{\beta}}_{1,reg} - \boldsymbol{\beta}\|_{L_\infty} = O_p(\rho_n),$$

where $\tilde{\beta}_{1p,reg}(u_p) = \mathbf{B}(u_p)(\mathbf{R}'\mathbf{R})\mathbf{RM}$ is the expectation of the regular spline estimator (i.e. $\lambda_1 = \ldots = \lambda_d = 0$). From the proof of Theorem 2 in [1], we have that

$$\tilde{\boldsymbol{\beta}}_1 = \left(1 - O_p(n^{-1}m_{\max}^{3/2}\lambda_{\max})\right)\tilde{\boldsymbol{\beta}}_{1,reg}.$$

Since each spline $\tilde{\beta}_p$ is a continuous function on the compact set $\mathcal{U}_p$, each spline $\tilde{\beta}_p$ is bounded and $\|\tilde{\boldsymbol{\beta}}_{1,reg}\|_{L_\infty} = O_P(1)$. We therefore conclude that

$$\|\tilde{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}\|_{L_\infty} = O_p(\rho_n + n^{-1}m_{\max}^{3/2}\lambda_{\max}).$$

The result of Lemma 3(a) now follows.

Proof of Lemma 3(b)
Lemma 3(b) is for $\sigma(\mathbf{u},\mathbf{x})$ what Lemma 1 is for $m(\mathbf{u},\mathbf{x})$. Again we consider two cases: Z_i exceeds or does not exceed $\tau_1(\mathbf{U}_i,\mathbf{X}_i)$. Suppose first that $Z_i \leq \tau_1(\mathbf{U}_i,\mathbf{X}_i)$, then we write

$$| \hat{Y}^*_{1i,\sigma^2} - Y^*_{1i,\sigma^2} |$$
$$\leq \left|\hat{m}_1^2(\mathbf{U}_i,\mathbf{X}_i) - m^2(\mathbf{U}_i,\mathbf{X}_i)\right| + 2Z_i\,|\hat{m}_1(\mathbf{U}_i,\mathbf{X}_i) - m(\mathbf{U}_i,\mathbf{X}_i)|$$
$$+ (Z_i - m(\mathbf{U}_i,\mathbf{X}_i))^2\,|\hat{G}(Z_i|\mathbf{U}_i,\mathbf{X}_i) - G(Z_i|\mathbf{U}_i,\mathbf{X}_i)|.$$

Since $\hat{m}^2(\mathbf{u},\mathbf{x}) - m^2(\mathbf{u},\mathbf{x}) = \{\hat{m}(\mathbf{u},\mathbf{x}) - m(\mathbf{u},\mathbf{x})\}\{\hat{m}(\mathbf{u},\mathbf{x}) + m(\mathbf{u},\mathbf{x})\}$, we get from the uniform convergence of $\hat{m}(\mathbf{u},\mathbf{x})$ to $m(\mathbf{u},\mathbf{x})$, that the rate of the first and second term on the right-hand side are both equal to the rate obtained in

Lemma 3(a). The third term on the right hand side is bounded in probability by $\sup_{t\le\tau_1(\mathbf{U}_i,\mathbf{X}_i)}|\hat{G}(t|\mathbf{U}_i,\mathbf{X}_i)-G(t|\mathbf{U}_i,\mathbf{X}_i)|$.

Next, suppose $Z_i>\tau_1(\mathbf{U}_i,\mathbf{X}_i)$, then we can write

$$|\hat{Y}^*_{1i,\sigma^2}-Y^*_{1i,\sigma^2}|\le|\hat{Y}^*_{1i,\sigma^2}-\tilde{Y}^*_{1i,\sigma^2}|+|\tilde{Y}^*_{1i,\sigma^2}-Y^*_{1i,\sigma^2}|,$$

where $\tilde{Y}^*_{1i,\sigma^2}=Y^*_{1i,\sigma^2}1_{\{Z_i\le\tau_1(\mathbf{U}_i,\mathbf{X}_i)\}}+(Z_i-m^2(\mathbf{U}_i,\mathbf{X}_i))^2 1_{\{Z_i>\tau_1(\mathbf{U}_i,\mathbf{X}_i)\}}$. Analogue to the second part of the proof of Lemma 1, we use κ_σ to bound the difference between $\hat{Y}^*_{1i,\sigma^2}$ and Y^*_{1i,σ^2} in the truncation area. For the estimation of the mean of Y, the transformation formula when Z_i lies in the truncation area is Z_i, whereas in this case, the transformation formula is $(Z_i-\hat{m}_1(\mathbf{U}_i,\mathbf{X}_i))^2$ and therefore also involves an estimator $\hat{m}_1$. The variable $\tilde{Y}^*_{1i,\sigma^2}$ is introduced to make the transition from $\hat{Y}^*_{1i,\sigma^2}\equiv(Z_i-\hat{m}_1(\mathbf{U}_i,\mathbf{X}_i))^2$ via $\tilde{Y}^*_{1i,\sigma^2}\equiv(Z_i-m(\mathbf{U}_i,\mathbf{X}_i))^2$ to Y^*_{1i,σ^2}. We get

$$E|\tilde{Y}^*_{1i,\sigma^2}-Y^*_{1i,\sigma^2}|\le\sup_{\mathbf{u},\mathbf{x}}\kappa_\sigma(\mathbf{u},\mathbf{x}),$$

and

$$\begin{aligned}
&|\hat{Y}^*_{1i,\sigma^2}-\tilde{Y}^*_{1i,\sigma^2}|\\
&\le 2Z_i\,|\hat{m}_1(\mathbf{U}_i,\mathbf{X}_i)-m(\mathbf{U}_i,\mathbf{X}_i)|+\left|\hat{m}_1^2(\mathbf{U}_i,\mathbf{X}_i)-m^2(\mathbf{U}_i,\mathbf{X}_i)\right|\\
&=O_p\Bigg(n^{-1/2}+n^{-1}m_{\max}^{3/2}\lambda_{\max}+\rho_n\\
&\quad+m_{\max}^{-1/2}\left(\sup_{\mathbf{u},\mathbf{x}}\left\{\tau_1(\mathbf{u},\mathbf{x})\sup_{t\le\tau_1(\mathbf{u},\mathbf{x})}|\hat{G}(t|\mathbf{u},\mathbf{x})-G(t|\mathbf{u},\mathbf{x})|+\kappa(\mathbf{u},\mathbf{x})\right\}\right)\Bigg).
\end{aligned}$$

Proof of Lemma 3(c)
Following the same steps as in the proof of Theorem 1, Part 1, we can, using the result of Lemma 3(b), derive the L_2-distance between $\hat{\sigma}^2$ and σ^2. Analogous to Lemma 3(a), the L_∞-distance then follows. Since $\hat{\sigma}_1-\sigma=(\hat{\sigma}_1^2-\sigma^2)/(\hat{\sigma}_1+\sigma)$, it follows from the convergence of $\hat{\sigma}_1^2(\mathbf{u},\mathbf{x})$ to $\sigma^2(\mathbf{u},\mathbf{x})>0$, that the rate is maintained for $\hat{\sigma}_1-\sigma$.

Lemma 4 *If assumptions A, B and C hold, then, for $t<S$, we have*

$$\begin{aligned}
&\hat{F}(t)-F(t)=O_p\Bigg(n^{-1/2}(\log n)^{1/2}+n^{-1}m_{max}^{3/2}\lambda_{max}+\rho_n+\\
&m_{max}^{-1/2}\left[\sup_{\mathbf{u},\mathbf{x}}\left\{\tau_1(\mathbf{u},\mathbf{x})\sup_{t\le\tau_1(\mathbf{u},\mathbf{x})}|\hat{G}(t|\mathbf{u},\mathbf{x})-G(t|\mathbf{u},\mathbf{x})|+\kappa(\mathbf{u},\mathbf{x})+\kappa_\sigma(\mathbf{u},\mathbf{x})\right\}\right]\Bigg).
\end{aligned}$$

Lemma 5 *Suppose $\beta_p\in C^r([a_p,b_p])$ for each $p=1,\dots,d$. Then under Assumptions A and B, we have*

$$\begin{aligned}
&\|\hat{\boldsymbol{\beta}}_1^{(v)}-\boldsymbol{\beta}^{(v)}\|_{L_\infty}=O_p\Bigg(n^{-1/2}m_{max}^{v}+n^{-1}m_{max}^{3/2}\lambda_{max}+m_{max}^{v-r}\\
&+m_{max}^{v-1/2}\left[\sup_{\mathbf{u},\mathbf{x}}\left\{\tau_1(\mathbf{u},\mathbf{x})\sup_{t\le\tau_1(\mathbf{u},\mathbf{x})}|\hat{G}(t|\mathbf{u},\mathbf{x})-G(t|\mathbf{u},\mathbf{x})|+\kappa(\mathbf{u},\mathbf{x})\right\}+\rho_n\right]\Bigg),
\end{aligned}$$

where $\boldsymbol{\beta}^{(v)} = \left(\frac{\partial^v \beta_1}{\partial u_1^v}, \ldots, \frac{\partial^v \beta_d}{\partial u_d^v}\right)'$ *and* $\hat{\boldsymbol{\beta}}_1^{(v)} = \left(\frac{\partial^v \hat{\beta}_{11}}{\partial u_1^v}, \ldots, \frac{\partial^v \hat{\beta}_{1d}}{\partial u_d^v}\right)'$ *are the vectors of the v-th order derivative functions for* $v = 0, \ldots, r-1$.

Proof (Proof of Lemma 5)
We first note that the v-th derivative of the B-spline function $\hat{\boldsymbol{\beta}}_{1p}(u_p) = \sum_{l=1}^{m_p} \hat{\alpha}_{1p,l} B_{pl}(u_p, q_p)$ of degree q_p is a B-spline function of degree $q_p - v$ given by (see [2])

$$\hat{\boldsymbol{\beta}}_1^{(v)} = K_p^v \, \mathbf{b}(u_p, q-v)' \mathbf{D}_v \hat{\boldsymbol{\alpha}}_{1p}, \tag{8}$$

where $\mathbf{b}(u_p, q-v) = (B_{1p}(u_p, q_p - v), \ldots, B_{m_p-1,p}(u_p, q_p - v))'$ is the vector of the $K_p + q_p - v$ B-spline basis functions of degree $q_p - v$ with knots $\boldsymbol{\xi}_p$, i.e. for $v = 1$, we have

$$\begin{aligned}\hat{\beta}_{1p}^{(1)}(u_p) &= K_p \sum_{l=1}^{m_p-1} (\hat{\alpha}_{1p,l-1} - \hat{\alpha}_{1p,l}) B_{pl}(u_p, q_p - 1) = K_p \mathbf{b}(u_p, q-1)' \mathbf{D}_1 \hat{\boldsymbol{\alpha}}_{1p} \\ &= K_p \left(\mathbf{b}(u_p, q-1)' \hat{\boldsymbol{\alpha}}_{1[-1]} - \mathbf{b}(u_p, q-1)' \hat{\boldsymbol{\alpha}}_{1[-m]}\right),\end{aligned}$$

where $\hat{\boldsymbol{\alpha}}_{1[-1]} = (\hat{\boldsymbol{\alpha}}_{12}, \ldots \hat{\boldsymbol{\alpha}}_{1m})$, $\boldsymbol{\alpha}_{1[-m]} = (\hat{\boldsymbol{\alpha}}_{11}, \ldots \hat{\boldsymbol{\alpha}}_{1,m-1})$. Representation (8) implies that the v-th derivative of β_p is again a spline function with coefficient vector $K_p \mathbf{D}_v \hat{\boldsymbol{\alpha}}_{1p}$. As a consequence we have, using Property 2, that

$$\|\hat{\boldsymbol{\beta}}_1^{(v)} - \tilde{\boldsymbol{\beta}}_1^{(v)}\|_{L_2} = O_p(m_{\max}^{v-1/2} \|\hat{\boldsymbol{\alpha}}_1 - \tilde{\boldsymbol{\alpha}}_1\|_2). \tag{9}$$

We now use the fact that there exists a spline function (see Corollary 6.21 and (2.120) of Theorem 2.59 in [6]) $\zeta_p(u_p) = \sum_{l=1}^{m_p} c_{pl} B_{pl}(u_p, q_p)$ of degree q_p with equidistant knots $\boldsymbol{\xi}_p$ and coefficient vector $\mathbf{c}_p = (c_{1p}, \ldots, c_{m_p p})'$ such that

$$\|\tilde{\boldsymbol{\beta}}_1^{(v)} - \boldsymbol{\zeta}^{(v)}\|_{L_2} = O_p(m_{\max}^v \rho_n + n^{-1} m_{\max}^{3/2} \lambda_{\max}). \tag{10}$$

To show the validity of (10), we proceed as follows. By Lemma A.7 of [5], we have that $\|\tilde{\boldsymbol{\alpha}}_{1,reg} - \mathbf{c}\|_2 = O(m_{\max}^{1/2} \rho_n)$, using a similar argument as before we find, $\|\tilde{\boldsymbol{\beta}}_{1,reg}^{(v)} - \boldsymbol{\zeta}^{(v)}\|_{L_2} = O_p(m_{\max}^v \rho_n)$. Using the relationship

$$\tilde{\boldsymbol{\beta}}_1^{(v)} = \left(1 - O_p(n^{-1} m_{\max}^{3/2} \lambda_{\max})\right) \tilde{\boldsymbol{\beta}}_{1,reg}^{(v)},$$

and the fact that $\boldsymbol{\beta}_{1,reg}^{(v)}$ is bounded on a compact region, we have $\|\boldsymbol{\beta}_{1,reg}^{(v)}\|_{L_2} = O_p(1)$ and (10) follows. Also note ([6]) that ζ_p satisfies

$$\|\beta_p^{(v)} - \zeta_p^{(v)}\|_{L_\infty} = O(m_p^{v-r}). \tag{11}$$

The rates in (9)-(11) provide the key for the proof. Indeed

$$\|\hat{\boldsymbol{\beta}}_1^{(v)} - \boldsymbol{\beta}^{(v)}\|_{L_\infty} \leq \|\hat{\boldsymbol{\beta}}_1^{(v)} - \boldsymbol{\zeta}^{(v)}\|_{L_\infty} + \|\boldsymbol{\zeta}^{(v)} - \boldsymbol{\beta}^{(v)}\|_{L_\infty}. \tag{12}$$

For the second term in (12) we use (11). For the first term, note that

$$\|\hat{\boldsymbol{\beta}}_1^{(v)} - \boldsymbol{\zeta}^{(v)}\|_{L_\infty} \lesssim m_{\max}^{-1/2} \|\hat{\boldsymbol{\beta}}_1^{(v)} - \boldsymbol{\zeta}^{(v)}\|_{L_2} \tag{13}$$

and that

$$\begin{aligned}\|\hat{\boldsymbol{\beta}}_1^{(v)} - \boldsymbol{\zeta}^{(v)}\|_{L_2} &\le \|\hat{\boldsymbol{\beta}}_1^{(v)} - \tilde{\boldsymbol{\beta}}^{(v)}\|_{L_2} + \|\tilde{\boldsymbol{\beta}}_1^{(v)} - \boldsymbol{\zeta}^{(v)}\|_{L_2} \\ &= O_p(m_{\max}^{v-1/2} \|\hat{\boldsymbol{\alpha}}_1 - \tilde{\boldsymbol{\alpha}}_1\|_2 + m_{\max}^{v} \rho_n + n^{-1} m_{\max}^{3/2} \lambda_{\max}). \quad (14)\end{aligned}$$

The result now follows from the rate obtained for $\|\hat{\boldsymbol{\alpha}}_1 - \tilde{\boldsymbol{\alpha}}_1\|_2$ in Theorem 1, Part 1 in combination with (9)-(14).

Proof (Proof of Lemma 2)
We first note that $\sup_{\mathbf{u},\mathbf{x}} |\hat{m}_1(\mathbf{u},\mathbf{x}) - m(\mathbf{u},\mathbf{x})|$ and $\sup_{\mathbf{u},\mathbf{x}} |\hat{\sigma}_1(\mathbf{u},\mathbf{x}) - \sigma(\mathbf{u},\mathbf{x})|$ are both $O_p(a_n)$ by Lemma 3. We write

$$\begin{aligned}\hat{Y}_{2i}^* - Y_{2i}^* &= \hat{m}_1(\mathbf{U}_i,\mathbf{X}_i) - m(\mathbf{U}_i,\mathbf{X}_i) \\ &\quad + \frac{\hat{\sigma}_1(\mathbf{U}_i,\mathbf{X}_i)}{1-\hat{F}(\hat{E}_i^T)} \int_{\hat{E}_i^T}^{\hat{S}_i} s d\hat{F}(s) - \frac{\sigma(\mathbf{U}_i,\mathbf{X}_i)}{1-F(E_i^T)} \int_{E_i^T}^{S_i} s dF(s) \\ &= \{\hat{m}_1(\mathbf{U}_i,\mathbf{X}_i) - m(\mathbf{U}_i,\mathbf{X}_i)\} \qquad (15)\\ &\quad + \frac{\hat{\sigma}_1(\mathbf{U}_i,\mathbf{X}_i) - \sigma(\mathbf{U}_i,\mathbf{X}_i)}{1-\hat{F}(\hat{E}_i^T)} \int_{\hat{E}_i^T}^{\hat{S}_i} s d\hat{F}(s) \qquad (16)\\ &\quad + \frac{\sigma(\mathbf{U}_i,\mathbf{X}_i)\{\hat{F}(\hat{E}_i^T) - F(E_i^T)\}}{\{1-\hat{F}(\hat{E}_i^T)\}\{1-F(E_i^T)\}} \int_{\hat{E}_i^T}^{\hat{S}_i} s d\hat{F}(s) \qquad (17)\\ &\quad + \frac{\sigma(\mathbf{U}_i,\mathbf{X}_i)}{1-F(E_i^T)} \left\{ \int_{\hat{E}_i^T}^{E_i^T} s d\hat{F}(s) + \int_{E_i^T}^{S_i} s d(\hat{F}(s) - F(s)) + \int_{S_i}^{\hat{S}_i} s d\hat{F}(s) \right\}. \qquad (18)\end{aligned}$$

We first consider the three integrals in (18). Using integration by part, we have

$$\begin{aligned}&\int_{\hat{E}_i^T}^{E_i^T} s d\hat{F}(s) = E_i^T \hat{F}(E_i^T) - \hat{E}_i^T \hat{F}(\hat{E}_i^T) - \int_{\hat{E}_i^T}^{E_i^T} \hat{F}(s) ds \\ &= E_i^T \{\hat{F}(E_i^T) - F(E_i^T)\} + \{E_i^T F(E_i^T) - \hat{E}_i^T F(E_i^T)\} + \hat{E}_i^T \{F(E_i^T) - \hat{F}(\hat{E}_i^T)\} \\ &\qquad - \int_{\hat{E}_i^T}^{E_i^T} \hat{F}(s) ds. \qquad (19)\end{aligned}$$

For the first term of (19), using Lemma 4, we conclude that

$$\left| E_i^T \{\hat{F}(E_i^T) - F(E_i^T)\} \right| = |E_i^T| O_p(a_n) = O_p(a_n),$$

since $|E_i^T| \leq \{\sigma(\mathbf{U}_i,\mathbf{X}_i)\}^{-1}\{|\min(Z_i,\tau_2(\mathbf{U}_i,\mathbf{X}_i))|+|m(\mathbf{U}_i,\mathbf{X}_i)|\} < \infty$. To get a consistency rate for the second and the fourth term of (19), note that

$$
\begin{aligned}
&\hat{E}_i^T - E_i^T \\
&= \frac{\min(\tau_2(\mathbf{U}_i,\mathbf{X}_i),Z_i) - \hat{m}_1(\mathbf{U}_i,\mathbf{X}_i)}{\hat{\sigma}_1(\mathbf{U}_i,\mathbf{X}_i)} - \frac{\min(\tau_2(\mathbf{U}_i,\mathbf{X}_i),Z_i) - m(\mathbf{U}_i,\mathbf{X}_i)}{\sigma(\mathbf{U}_i,\mathbf{X}_i)} \\
&= \frac{1}{\sigma(\mathbf{U}_i,\mathbf{X}_i)\hat{\sigma}(\mathbf{U}_i,\mathbf{X}_i)}\Big[\min(\tau_2(\mathbf{U}_i,\mathbf{X}_i),Z_i)\big\{\sigma(\mathbf{U}_i,\mathbf{X}_i) - \hat{\sigma}_1(\mathbf{U}_i,\mathbf{X}_i)\big\} \\
&\qquad - \sigma(\mathbf{U}_i,\mathbf{X}_i)\big\{\hat{m}_1(\mathbf{U}_i,\mathbf{X}_i) - m(\mathbf{U}_i,\mathbf{X}_i))\big\} \\
&\qquad + m(\mathbf{U}_i,\mathbf{X}_i)\big\{\hat{\sigma}_1(\mathbf{U}_i,\mathbf{X}_i) - \sigma(\mathbf{U}_i,\mathbf{X}_i)\big\}\Big].
\end{aligned}
$$

It then follows from Lemma 3 and the convergence of $\hat{\sigma}_1(\mathbf{u},\mathbf{x})$ to $\sigma(\mathbf{u},\mathbf{x}) > 0$ that

$$|\hat{E}_i^T - E_i^T| = O_p(a_n),$$

which gives the rate for the second and the fourth term of (19). For the third term of (19), we have that

$$\hat{F}(\hat{E}_i^T) - F(E_i^T) = \{\hat{F}(\hat{E}_i^T) - F(\hat{E}_i^T)\} + \{F(\hat{E}_i^T) - F(E_i^T)\}.$$

Lemma 4 can be used for the first summand. For the second summand, we use a first order Taylor approximation and write

$$
\begin{aligned}
F(\hat{E}_i^T) - F(E_i^T) = &\Bigg(-\frac{\hat{m}_1(\mathbf{U}_i,\mathbf{X}_i) - m(\mathbf{U}_i,\mathbf{X}_i)}{\hat{\sigma}_1(\mathbf{U}_i,\mathbf{X}_i)} \\
&- \frac{\{\hat{\sigma}_1(\mathbf{U}_i,\mathbf{X}_i) - \sigma(\mathbf{U}_i,\mathbf{X}_i)\}\{\min(\tau_2(\mathbf{U}_i,\mathbf{X}_i),Z_i) - m(\mathbf{U}_i,\mathbf{X}_i)\}}{\hat{\sigma}_1(\mathbf{U}_i,\mathbf{X}_i)\sigma_1(\mathbf{U}_i,\mathbf{X}_i)}\Bigg) f_\varepsilon(\theta),
\end{aligned}
$$

with f_ε the density of ε and for some θ between $\frac{\min(\tau_2(\mathbf{U}_i,\mathbf{X}_i),Z_i) - \hat{m}_1(\mathbf{U}_i,\mathbf{X}_i)}{\hat{\sigma}_1(\mathbf{U}_i,\mathbf{X}_i)}$ and $\frac{\min(\tau_2(\mathbf{U}_i,\mathbf{X}_i),Z_i) - m(\mathbf{U}_i,\mathbf{X}_i)}{\sigma(\mathbf{U}_i,\mathbf{X}_i)}$. By the convergence of $\hat{\sigma}_1(\mathbf{u},\mathbf{x})$ to $\sigma(\mathbf{u},\mathbf{x}) > 0$ and the fact that $\sup_e |ef_\varepsilon(e)| < \infty$, we get

$$F(\hat{E}_i^T) - F(E_i^T) = O_p(a_n). \tag{20}$$

We conclude that

$$\Big|\hat{E}_i^T\{F(E_i^T) - \hat{F}(\hat{E}_i^T)\}\Big| = O_p(a_n),$$

where we use that by Lemma 3, $|\hat{E}_i^T| = |E_i^T| + O_p(a_n) < \infty$. Based on the analysis of (19) we obtain for the first term of (18)

$$\frac{\sigma(\mathbf{U}_i,\mathbf{X}_i)}{1 - F(E_i^T)}\int_{\hat{E}_i^T}^{E_i^T} s\,d\hat{F}(s) = O_p(a_n). \tag{21}$$

In a similar way, we obtain for the third term of (18)

$$\frac{\sigma(\mathbf{U}_i, \mathbf{X}_i)}{1 - F(E_i^T)} \int_{\hat{S}_i^T}^{S_i^T} sd\hat{F}(s) = O_p(a_n). \tag{22}$$

For the second integral in (18), we use partial integration and Lemma 4 to obtain

$$\begin{aligned}\int_{E_i^T}^{S_i^T} sd(\hat{F}(s) - F(s)) = \quad & S_i^T\{\hat{F}(S_i^T) - F(S_i^T)\} - E_i^T\{\hat{F}(E_i^T) - F(E_i^T)\} \\ & - \int_{E_i^T}^{S_i^T} \{\hat{F}(s) - F(s)\}ds = O_p(a_n).\end{aligned}$$

The terms (15)-(17) are more easy to handle. For (15) we use Lemma 3(a). For (16) and (17) we need that

$$\int_{\hat{E}_i^T}^{\hat{S}_i} sd\hat{F}(s) = O_p(1). \tag{23}$$

To show (23), note that, using similar reasoning as in [4], we can prove that

$$\int_{E_i^T}^{S_i} sd\hat{F}(s) = O_p(1).$$

Combining this result with the rates obtained in (21) and (22) yields

$$\int_{\hat{E}_i^T}^{\hat{S}_i} sd\hat{F}(s) = O_p(1).$$

By the convergence of $\hat{F}(\hat{E}_i^T)$ to $F(E_i^T) < 1$ (20), we get that (16) and (17) are both $O_p(a_n)$.

1.3 Proof of Theorem 2

Proof (Proof of Theorem 2)
We prove the asymptotic normality of the P-spline estimator $\hat{\boldsymbol{\beta}}_1$ for method 1 by proving that for $p = 1, \ldots, d$,

$$\left\{s.e.\left(\beta_{jp}^*(u_p) \mid \mathcal{X}_n\right)\right\}^{-1}\left\{\beta_{jp}^*(u_p) - \tilde{\beta}_{jp}(u_p)\right\} \xrightarrow{d} \mathrm{N}(0,1), \tag{24}$$

and

$$\left\{s.e.\left(\beta_{jp}^*(u_p) \mid \mathcal{X}_n\right)\right\}^{-1}\left\{(\hat{\beta}_{1p}(u_p) - \beta_{1p}^*(u_p)) + (\tilde{\beta}_{1p}(u_p) - \beta_p(u_p))\right\} \xrightarrow{p} 0. \tag{25}$$

The proof of (24) is based on the proof given in [1] where some steps can be simplified due to the independence of the observations.

Let $\mathbf{B}_p(\mathbf{u})$ be the column vector representing the p-th row of $\mathbf{B}(\mathbf{u})$.

$$\mathbf{B}_p'(\mathbf{u})(\boldsymbol{\alpha}^* - \tilde{\boldsymbol{\alpha}}) = \sum_{i=1}^n \mathbf{B}_p'(\mathbf{u})(\mathbf{R}'\mathbf{R} + \mathbf{Q}_{\boldsymbol{\lambda}})^{-1}\mathbf{R}_i(Y_{1i}^* - M_{1i}) = \sum_{i=1}^n d_i\xi_i,$$

where $d_i^2 = \sigma_{1,i}^2\{\mathbf{B}_p'(\mathbf{u})(\mathbf{R}'\mathbf{R} + \mathbf{Q}_{\boldsymbol{\lambda}})^{-1}\mathbf{R}_i\}^2$ and $\xi_i = \sigma_{1,i}^{-2}(Y_{1i}^* - M_{1i})$. Conditioning on $\mathcal{X}_n$ the ξ_i are independent with mean 0 and variance 1. To prove the asymptotic normality of the P-spline estimator we verify the Lindeberg condition

$$\frac{\max d_i^2}{\sum_{i=1}^n d_i^2} \xrightarrow{p} 0.$$

Then

$$\frac{\sum_{i=1}^n d_i\xi_i}{\sqrt{\sum_{i=1}^n d_i^2}} \xrightarrow{d} \mathrm{N}(0,1).$$

For any $\boldsymbol{\omega} = (\boldsymbol{\omega}_0', \ldots, \boldsymbol{\omega}_d')'$ with $\boldsymbol{\omega}_p = (\omega_{p1}, \ldots, \omega_{pm_p})'$, and especially for $\boldsymbol{\omega} = \{\mathbf{R}'\mathbf{R} + \mathbf{Q}_{\boldsymbol{\lambda}})^{-1}\mathbf{B}_p(\mathbf{u})\}$, we have by the Cauchy-Schwarz inequality

$$\begin{aligned}\boldsymbol{\omega}'\mathbf{R}_i\mathbf{R}_i'\boldsymbol{\omega} &= \left\{\sum_{p=0}^d X_{ip}\sum_{l=1}^{m_p}\omega_{pl}B_{pl}(U_{ip};q_p)\right\}^2 \\ &\le \left(\sum_{p=0}^d X_{ip}^2\right)\left[\sum_{p=0}^d\left\{\sum_{l=1}^{m_p}\omega_{pl}B_{pl}(U_{ip};q_p)\right\}^2\right].\end{aligned}$$

Set $g_{\boldsymbol{\omega},p}(u;q_p) = \sum_{l=1}^{m_p}\omega_{pl}B_{pl}(u_p;q_p)$ for $p = 0, \ldots, d$. By Assumption (B3) and Properties 2 and 4, we have

$$\boldsymbol{\omega}'\mathbf{R}_i\mathbf{R}_i'\boldsymbol{\omega} \lesssim \sum_{p=0}^d \|g_{\boldsymbol{\omega},p}\|_\infty^2 \lesssim m_{\max}\sum_{p=0}^d \|g_{\boldsymbol{\omega},p}\|_{L_2}^2 \asymp \|\boldsymbol{\omega}\|_2^2. \tag{26}$$

From Lemmas A.1 and A.2 in [5], we know that, except on an event with probability tending to zero, $n^{-1}\sum_{i=1}^n(\sum_{p=0}^d X_{ip}g_{\boldsymbol{\omega},p}(U_{ip};q_p))^2 \asymp m_{\max}^{-1}\|\boldsymbol{\omega}\|_2^2$. Thus

$$\begin{aligned}\boldsymbol{\omega}'\sum_{i=1}^n\left\{\mathbf{R}_i\mathbf{R}_i'\sigma_{1,i}^2\right\}\boldsymbol{\omega} &\ge n\min_{1\le i\le n}\sigma_{1,i}^2 n^{-1}\sum_{i=1}^n\Big(\sum_{p=0}^d X_{ip}g_{\boldsymbol{\omega},p}(U_{ip};q_p)\Big)^2 \\ &\gtrsim m_{\max}^{-1}n\|\boldsymbol{\omega}\|_2^2. \end{aligned}\tag{27}$$

Combining (26) and (27), we find that, except on an event whose probability tends to zero, we have

$$\frac{\max_i(\sigma_{1,i}^2\omega'\mathbf{R}_i\mathbf{R}_i'\omega)}{\boldsymbol{\omega}'(\sum_{i=1}^n\sigma_{1,i}^2\mathbf{R}_i\mathbf{R}_i')\boldsymbol{\omega}} \lesssim n^{-1}m_{\max}.$$

By Assumption (B6), it follows that the Lindeberg condition is fulfilled and hence the normality result in (24) follows.

We continue with the proof of (25). Since we assume that $\sigma_{1,i}^2$ is bounded away from zero and ∞, we have,

$$\begin{aligned}
\mathrm{Var}(\boldsymbol{\beta}_{1p}^*(\mathbf{u}) \mid \mathcal{X}_n) &= \mathrm{Cov}\left(\mathbf{B}_p'(\mathbf{u})\boldsymbol{\alpha}^* \mid \mathcal{X}_n\right) \\
&= \mathbf{B}(\mathbf{u})\left(\mathbf{R}'\mathbf{R}+\mathbf{Q}_{\boldsymbol{\lambda}}\right)^{-1}\left(\sum_{i=1}^n \mathbf{R}_i\mathbf{R}_i'\sigma_{1,i}^2\right)\left(\mathbf{R}'\mathbf{R}+\mathbf{Q}_{\boldsymbol{\lambda}}\right)^{-1}\mathbf{B}_p(\mathbf{u}) \\
&\gtrsim \mathbf{B}_p'(\mathbf{u})\left(\mathbf{R}'\mathbf{R}+\mathbf{Q}_{\boldsymbol{\lambda}}\right)^{-1}\mathbf{R}'\mathbf{R}\left(\mathbf{R}'\mathbf{R}+\mathbf{Q}_{\boldsymbol{\lambda}}\right)^{-1}\mathbf{B}_p(\mathbf{u}) \\
&\asymp \frac{n}{m_{\max}}\mathbf{B}_p'(\mathbf{u})\left(\mathbf{R}'\mathbf{R}+\mathbf{Q}_{\boldsymbol{\lambda}}\right)^{-1}\left(\mathbf{R}'\mathbf{R}+\mathbf{Q}_{\boldsymbol{\lambda}}\right)^{-1}\mathbf{B}_p(\mathbf{u}) \\
&\gtrsim \frac{n}{m_{\max}}\left(\frac{1}{\lambda_{\max}(\mathbf{R}'\mathbf{R}+\mathbf{Q}_{\boldsymbol{\lambda}})}\right)^2\sum_{l=1}^{m_p} B_{pl}^2(\mathbf{u}) \\
&\gtrsim \frac{n}{m_{\max}}\left(\frac{1}{\frac{n}{m_{\max}}\left(1+\frac{m_{\max}^{3/2}\lambda_{\max}}{n}\right)}\right)^2\frac{1}{m_p} \\
&\asymp \frac{1}{n}\left(1+\frac{m_{\max}^{3/2}\lambda_{\max}}{n}\right)^{-2},
\end{aligned}$$

where we use the Cauchy-Schwarz inequality

$$1 = \left(\sum_{l=1}^{m_p} B_{pl}(\mathbf{u})\right)^2 \le \sum_{l=1}^{m_p} B_{pl}^2(\mathbf{u})\sum_{l=1}^{m_p} 1 = m_p\sum_{l=1}^{m_p} B_{pl}^2(\mathbf{u}),$$

and the following upper bound for the largest eigenvalue $\lambda_{\max}(\mathbf{R}'\mathbf{R}+\mathbf{Q}_{\boldsymbol{\lambda}})$:

$$\begin{aligned}
\lambda_{\max}(\mathbf{R}'\mathbf{R}+\mathbf{Q}_{\boldsymbol{\lambda}}) &= \|\mathbf{R}'\mathbf{R}+\mathbf{Q}_{\boldsymbol{\lambda}}\|_2 \le \|\mathbf{R}'\mathbf{R}\|_2 + \|\mathbf{Q}_{\boldsymbol{\lambda}}\|_2 \\
&\lesssim \frac{n}{m_{\max}} + \sqrt{\sum_{p=1}^d \|\mathbf{Q}_{\boldsymbol{\lambda}}\|_\infty} \lesssim \frac{n}{m_{\max}} + \sqrt{d}\lambda_{\max}m_{\max}^{1/2}\max_{1\le p\le d} 4^{k_p} \\
&\lesssim \frac{n}{m_{\max}}\left(1+\frac{m_{\max}^{3/2}\lambda_{\max}}{n}\right).
\end{aligned}$$

By Property 4 of B-splines and Assumption (A5),

$$\begin{aligned}
\hat{\beta}_{1p}(u_p)-\beta_{1p}^*(u_p) &\le \sup_{u\in\mathcal{U}}|\hat{\beta}_{1p}(u_p)-\beta_{1p}^*(u_p)| = \|\hat{\beta}_{1p}-\beta_{1p}^*\|_\infty \\
&\lesssim \left(\frac{1}{m_p}\right)^{1/2}\|\hat{\beta}_{1p}-\beta_{1p}^*\|_{L_2} \asymp \left(\frac{1}{m_{\max}}\right)^{1/2}\|\hat{\beta}_{1p}-\beta_{1p}^*\|_{L_2}.
\end{aligned}$$

We conclude

$$\frac{\hat{\beta}_{1p}(u_p)-\beta_{1p}^*(u_p)}{s.e.\left(\beta_{1p}^*(u_p)\mid\mathcal{X}_n\right)} \lesssim \left(\frac{n}{m_{\max}}\right)^{1/2}\left(1+\frac{m_{\max}^{3/2}\lambda_{\max}}{n}\right)\|\hat{\beta}_{1p}-\beta_{1p}^*\|_{L_2},$$

and

$$\frac{\tilde{\beta}_{1p}(u_p) - \beta_p(u_p)}{s.e.\left(\beta^*_{1p}(u_p) \mid \mathcal{X}_n\right)} \lesssim n^{1/2}\left(1 + \frac{m_{\max}^{3/2}\lambda_{\max}}{n}\right)\|\tilde{\beta}_{1p} - \beta_p\|_{L_\infty}.$$

From Assumption D.1 it follows that these two terms converge to zero as n goes to ∞. The proof for method 2 is similar.

References

1. Antoniadis, A., Gijbels, I., and Verhasselt, A. (2012). Variable selection in varying-coefficient models using P-splines. *Journal of Computational and Graphical Statistics*, 21:638–661.
2. De Boor, C. (1978). A practical guide to splines. *Springer, New York.*
3. Hendrickx, K., Janssen, P., and Verhasselt, A. (2017). Penalized spline estimation in varying coefficient models with censored data. *submitted to TEST.*
4. Heuchenne, C. and Van Keilegom, I. (2007). Polynomial regression with censored data based on preliminary nonparametric estimation. *Annals of the Institute of Statistical Mathematics*, 59:273–297.
5. Huang, J., Wu, C., and Zhou, L. (2004). Polynomial spline estimation and inference for varying coefficient models with longitudinal data. *Statistica Sinica*, 14:763–788.
6. Schumaker, L. (2007). Spline functions: Basic theory, 3th edition. *Cambridge University Press, New York.*
7. Van Keilegom, I. and Akritas, M. (1999). Transfer of tail information in censored regression models. *The Annals of Statistics*, 27:1745–1784.