

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>

# Penalized spline estimation in varying coefficient models with censored data

## Supplementary Material

K. Hendrickx<sup>1</sup> · P. Janssen<sup>1</sup> · A. Verhasselt<sup>1</sup>

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  $\|\mathbf{A}\|_2 = \sup_{\mathbf{x} \neq 0} \frac{\|\mathbf{Ax}\|_2}{\|\mathbf{x}\|_2}$ , with  $\mathbf{x} \in \mathbb{R}^{n_A}$  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  $\mathbf{g} = (g_1, \dots, g_d)$  on  $\mathcal{U}^d$ , the  $L_\infty$ -norm is given by:

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

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

<sup>1</sup>Hasselt University

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

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

*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  $g \in G(q_p, \boldsymbol{\xi}_p)$ , where  $G(q_p, \boldsymbol{\xi}_p)$  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

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

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

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

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

*Proof (Proof of Lemma 1)*

Since  $|\hat{Y}_{1i}^* - Y_{1i}^*| =$

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

we consider two cases and prove the following results,

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

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

For (1) we start by the triangle inequality,

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

We derive the order bound for  $|\hat{\varphi}_1(\mathbf{U}_i, \mathbf{X}_i, Z_i) - \varphi_1(\mathbf{U}_i, \mathbf{X}_i, Z_i)|$ , similar result holds if we replace  $\varphi_1$  and  $\hat{\varphi}_1$  by  $\psi_1$  and  $\hat{\psi}_1$  respectively.

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

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

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

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

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

For (2) we have

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

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

*Proof (Proof of Theorem 1, Part 1)*

Since

$$\|\hat{\beta}_1 - \beta_1\|_{L_2} \leq \|\hat{\beta}_1 - \beta_1^*\|_{L_2} + \|\beta_1^* - \tilde{\beta}_1\|_{L_2} + \|\tilde{\beta}_1 - \beta_1\|_{L_2},$$

the result follows by showing that

$$\|\hat{\beta}_1 - \beta_1^*\|_{L_2} \tag{3}$$

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

$$\|\beta_1^* - \tilde{\beta}_1\|_{L_2} = O_p\left(n^{-1/2} m_{\max}^{1/2}\right), \tag{4}$$

$$\|\tilde{\beta}_1 - \beta_1\|_{L_2} = O_p\left(n^{-1} m_{\max}^{3/2} \lambda_{\max} + \rho_n\right). \tag{5}$$

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

$$\begin{aligned} \|\hat{\alpha}_1 - \alpha_1^*\|_2 = \\ O_p\left(m_{\max}^{1/2} \left(\sup_{\mathbf{u}, \mathbf{x}} \left\{\tau_1(\mathbf{u}, \mathbf{x}) \sup_{t \leq \tau_1(\mathbf{u}, \mathbf{x})} |\hat{G}(t|\mathbf{u}, \mathbf{x}) - G(t|\mathbf{u}, \mathbf{x})| + \kappa(\mathbf{u}, \mathbf{x})\right\}\right)\right). \end{aligned}$$

From [1] we have

$$\begin{aligned} \hat{\alpha}_1 - \alpha_1^* &= \{(\mathbf{R}'\mathbf{R})^{-1} - (\mathbf{R}'\mathbf{R})^{-1}\mathbf{Q}_{\lambda}(\mathbf{R}'\mathbf{R})^{-1} + o_p(n^{-1} m_{\max}^{3/2} \lambda_{\max})(\mathbf{R}'\mathbf{R})^{-1}\} \\ &\quad \times \sum_{i=1}^n \mathbf{R}_i(\hat{Y}_{1i}^* - Y_{1i}^*) \\ &= \hat{\alpha}_{1,reg} - \alpha_{reg}^* - \{(\mathbf{R}'\mathbf{R})^{-1}\mathbf{Q}_{\lambda}(\mathbf{R}'\mathbf{R})^{-1} + o_p(n^{-1} m_{\max}^{3/2} \lambda_{\max})(\mathbf{R}'\mathbf{R})^{-1}\} \\ &\quad \times \sum_{i=1}^n \mathbf{R}_i(\hat{Y}_{1i}^* - Y_{1i}^*) \\ &= \left\{1 - (\mathbf{R}'\mathbf{R})^{-1}\mathbf{Q}_{\lambda} + o_p(n^{-1} m_{\max}^{3/2} \lambda_{\max})\right\} (\hat{\alpha}_{1,reg} - \alpha_{reg}^*), \end{aligned}$$

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

$$\begin{aligned} & \|\hat{\alpha}_1 - \alpha_1^*\|_2 \\ & \leq \left\{ 1 + \|(\mathbf{R}'\mathbf{R})^{-1}\|_2 \|\mathbf{Q}_\lambda\|_2 + o_p(n^{-1}m_{\max}^{3/2}\lambda_{\max}) \right\} \|\hat{\alpha}_{1,reg} - \alpha_{1,reg}^*\|_2. \end{aligned}$$

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

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

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

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

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

$$\begin{aligned} & \|\alpha_1^* - \tilde{\alpha}_1\|_2 \\ & \leq \left\{ 1 + \|(\mathbf{R}'\mathbf{R})^{-1}\|_2 \|\mathbf{Q}_\lambda\|_2 + o_p(n^{-1}m_{\max}^{3/2}\lambda_{\max}) \right\} \|\alpha_{1,reg}^* - \tilde{\alpha}_{1,reg}\|_2, \quad (6) \end{aligned}$$

and

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

By Assumption A.3,

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

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,

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

Therefore,

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

such that

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

Combining (6) and (7) gives,

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

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  $\sigma$ ; 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,*

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

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

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*

$$\begin{aligned} (a) \sup_{\mathbf{u}, \mathbf{x}} |\hat{m}_1(\mathbf{u}, \mathbf{x}) - m(\mathbf{u}, \mathbf{x})| &= O_p \left( n^{-1/2} + n^{-1}m_{max}^{3/2}\lambda_{max} + \rho_n \right. \\ &\quad \left. + m_{max}^{-1/2} \left( \sup_{\mathbf{u}, \mathbf{x}} \left\{ \tau_1(\mathbf{u}, \mathbf{x}) \sup_{t \leq \tau_1(\mathbf{u}, \mathbf{x})} |\hat{G}(t|\mathbf{u}, \mathbf{x}) - G(t|\mathbf{u}, \mathbf{x})| + \kappa(\mathbf{u}, \mathbf{x}) \right\} \right) \right). \\ (b) \max_{1 \leq i \leq n} |\hat{Y}_{1i, \sigma^2}^* - Y_{1i, \sigma^2}^*| &= O_p \left( n^{-1/2} + n^{-1}m_{max}^{3/2}\lambda_{max} + \rho_n + \right. \\ &\quad \left. \sup_{\mathbf{u}, \mathbf{x}} \left\{ \tau_1(\mathbf{u}, \mathbf{x}) \sup_{t \leq \tau_1(\mathbf{u}, \mathbf{x})} |\hat{G}(t|\mathbf{u}, \mathbf{x}) - G(t|\mathbf{u}, \mathbf{x})| + m_{max}^{-1/2}\kappa(\mathbf{u}, \mathbf{x}) + \kappa_\sigma(\mathbf{u}, \mathbf{x}) \right\} \right), \\ \text{where } Y_{1i, \sigma^2}^* &= \frac{\Delta_i(Z_i - m(\mathbf{U}_i, \mathbf{X}_i))^2}{G(Z_i|\mathbf{U}_i, \mathbf{X}_i)}. \\ (c) \sup_{\mathbf{u}, \mathbf{x}} |\hat{\sigma}_1(\mathbf{u}, \mathbf{x}) - \sigma(\mathbf{u}, \mathbf{x})| &= O_p \left( n^{-1/2} + n^{-1}m_{max}^{3/2}\lambda_{max} + \rho_n \right. \\ &\quad \left. + m_{max}^{-1/2} \left( \sup_{\mathbf{u}, \mathbf{x}} \left\{ \tau_1(\mathbf{u}, \mathbf{x}) \sup_{t \leq \tau_1(\mathbf{u}, \mathbf{x})} |\hat{G}(t|\mathbf{u}, \mathbf{x}) - G(t|\mathbf{u}, \mathbf{x})| \right. \right. \right. \\ &\quad \left. \left. \left. + m_{max}^{-1/2}\kappa(\mathbf{u}, \mathbf{x}) + \kappa_\sigma(\mathbf{u}, \mathbf{x}) \right\} \right) \right). \end{aligned}$$



*Proof (Proof of Lemma 3(a))*

Since the  $X_p$  are bounded (see Assumption A.3), we have,

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

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

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

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

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

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

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

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  $\|\tilde{\beta}_{1,reg}\|_{L_\infty} = O_P(1)$ . We therefore conclude that

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

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

$$\begin{aligned} &|\hat{Y}_{1i,\sigma^2}^* - Y_{1i,\sigma^2}^*| \\ &\leq |\hat{m}_1^2(\mathbf{U}_i, \mathbf{X}_i) - m^2(\mathbf{U}_i, \mathbf{X}_i)| + 2Z_i |\hat{m}_1(\mathbf{U}_i, \mathbf{X}_i) - m(\mathbf{U}_i, \mathbf{X}_i)| \\ &\quad + (Z_i - m(\mathbf{U}_i, \mathbf{X}_i))^2 |\hat{G}(Z_i|\mathbf{U}_i, \mathbf{X}_i) - G(Z_i|\mathbf{U}_i, \mathbf{X}_i)|. \end{aligned}$$

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  $\sup_{t \leq \tau_1(\mathbf{U}_i, \mathbf{X}_i)} |\hat{G}(t|\mathbf{U}_i, \mathbf{X}_i) - G(t|\mathbf{U}_i, \mathbf{X}_i)|$ .

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

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

where  $\tilde{Y}_{1i, \sigma^2}^* = Y_{1i, \sigma^2}^* 1_{\{Z_i \leq \tau_1(\mathbf{U}_i, \mathbf{X}_i)\}} + (Z_i - m^2(\mathbf{U}_i, \mathbf{X}_i))^2 1_{\{Z_i > \tau_1(\mathbf{U}_i, \mathbf{X}_i)\}}$ . Analogue to the second part of the proof of Lemma 1, we use  $\kappa_\sigma$  to bound the difference between  $\hat{Y}_{1i, \sigma^2}^*$  and  $Y_{1i, \sigma^2}^*$  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  $Y_{1i, \sigma^2}^*$ . We get

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

and

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

*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  $\sigma^2$ . Analogous to Lemma 3(a), the  $L_\infty$ -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*

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

**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*

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

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

*Proof (Proof of Lemma 5)*

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

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

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

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

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

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

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  $\xi_p$  and coefficient vector  $\mathbf{c}_p = (c_{1p}, \dots, c_{m_p p})'$  such that

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

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

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

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

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

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

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

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

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

and that

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

The result now follows from the rate obtained for  $\|\hat{\alpha}_1 - \tilde{\alpha}_1\|_2$  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

$$\begin{aligned} \hat{Y}_{2i}^* - Y_{2i}^* &= \hat{m}_1(\mathbf{U}_i, \mathbf{X}_i) - m(\mathbf{U}_i, \mathbf{X}_i) \\ &\quad + \frac{\hat{\sigma}_1(\mathbf{U}_i, \mathbf{X}_i)}{1 - \hat{F}(\hat{E}_i^T)} \int_{\hat{E}_i^T}^{\hat{S}_i} sd\hat{F}(s) - \frac{\sigma(\mathbf{U}_i, \mathbf{X}_i)}{1 - F(E_i^T)} \int_{E_i^T}^{S_i} sdF(s) \\ &= \{\hat{m}_1(\mathbf{U}_i, \mathbf{X}_i) - m(\mathbf{U}_i, \mathbf{X}_i)\} \end{aligned} \quad (15)$$

$$+ \frac{\hat{\sigma}_1(\mathbf{U}_i, \mathbf{X}_i) - \sigma(\mathbf{U}_i, \mathbf{X}_i)}{1 - \hat{F}(\hat{E}_i^T)} \int_{\hat{E}_i^T}^{\hat{S}_i} sd\hat{F}(s) \quad (16)$$

$$+ \frac{\sigma(\mathbf{U}_i, \mathbf{X}_i) \{\hat{F}(\hat{E}_i^T) - F(E_i^T)\}}{\{1 - \hat{F}(\hat{E}_i^T)\} \{1 - F(E_i^T)\}} \int_{\hat{E}_i^T}^{\hat{S}_i} sd\hat{F}(s) \quad (17)$$

$$+ \frac{\sigma(\mathbf{U}_i, \mathbf{X}_i)}{1 - F(E_i^T)} \left\{ \int_{\hat{E}_i^T}^{E_i^T} sd\hat{F}(s) + \int_{E_i^T}^{S_i} sd(\hat{F}(s) - F(s)) + \int_{S_i}^{\hat{S}_i} sd\hat{F}(s) \right\}. \quad (18)$$

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

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

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

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

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

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

with  $f_\varepsilon$  the density of  $\varepsilon$  and for some  $\theta$  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

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

We conclude that

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

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)

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

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

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

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

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

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

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

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

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

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

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

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  $\hat{\beta}_1$  for method 1 by proving that for  $p = 1, \dots, d$ ,

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

and

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

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  $\mathbf{B}(\mathbf{u})$ .

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

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

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

Then

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

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

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

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

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

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

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

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

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

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  $\infty$ , we have,

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

where we use the Cauchy-Schwarz inequality

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

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

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

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

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

We conclude

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



and

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

From Assumption D.1 it follows that these two terms converge to zero as  $n$  goes to  $\infty$ . 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.