Made available by Hasselt University Library in https://documentserver.uhasselt.be

Penalized spline estimation in varying coefficient models with censored data

Peer-reviewed author version

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

K. Hendrickx[1], P. Janssen[1], A. Verhasselt[1]

1. Center for Statistics, I-Biostat, University of Hasselt, Belgium

November 29, 2017

Abstract

We consider P-spline smoothing in a varying coefficient regression model when the response is subject to random right censoring. We introduce two data-transformation approaches to construct a synthetic response vector that is used in a penalized least-squares optimization problem. We prove the consistency and asymptotic normality of the P-spline estimators for a diverging number of knots and show by simulation studies and real data examples that the combination of a data-transformation for censored observations with P-spline smoothing leads to good estimators of the varying coefficient functions.

keywords: Censoring, Non-parametric statistics, P-splines, Regularization, Varying coefficient model

1 Introduction

Parametric regression models are commonly used for exploring relationships between a response variable and a set of explanatory variables. Linear models are often a good first approximation of the underlying association patterns but sometimes not able to capture complex dynamic structures. An extension of the classical linear regression model is the varying coefficient model (VCM, [17]). These models are still linear in the regressors but with regression coefficients that are smooth functions in one or more other variables, considered as effect modifiers. VCMs have been used is a successful way in many applications, among which are longitudinal models ([19]; [16]), survival models ([5]; [26]), generalized regression models ([6]; [24]) and non-linear time series [7]. The most commonly used estimation methods for VCMs are kernel regression [36], polynomial splines [20] and smoothing splines [17]. In this paper, we concentrate on the penalized spline (P-spline) smoothing technique proposed by [11]. P-spline regression is an extension of B-spline regression with a penalty in terms of finite differences of the coefficients of adjacent B-splines to protect against overfitting. P-splines are determined by the degree and the number and location of the knot points of the B-splines, the order of the difference penalty and a smoothing parameter. The consistency and asymptotic normality of the P-spline estimators for the regression coefficients in a VCM with longitudinal data was proved by [1].

Often encountered in the statistical analysis are situations where the response is not fully observed due to random right censoring, for example in medical and health care studies where patients leave the study for numerous reasons before the event of interest occurs (23, [28]). Another example of censoring arises in reliability studies, where the failure time of a device might be censored if the device is still functional at the end of the experiment ([27]). The popular proportional hazard model for right censored data ([9]) models the instantaneous risk as a product of a baseline hazard an an exponential factor. It models the relation between the response and covariates in an indirect way and is less simple to interpret than classical mean-regression models, where interest is in direct modeling of the mean event time as a function of covariates. The accelerated failure time model ([35]) on the other hand does propose a direct linear relationship between the logarithm of the survival time and covariates, but unlike the Cox proportional hazard model, accelerated failure time models are often parametric and hence require additional assumptions on the underlying survival distribution. Ordinary least squares regression, which avoid specifying the distribution of the response variable for estimating the parameters in a linear regression model, needs however modification when some of the responses are not observed. Extensions of ordinary least squares to censored data settings were first considered by [4]. The estimation technique relies on constructing a synthetic response based on a transformation formula that is (conditional) mean preserving. The new response then replaces the original response in the ordinary least squares regression problem with complete data. The transformation studied by [4] uses the underlying regression model and therefore needs an iterative estimation algorithm (see [21] for the implementation of the iterative procedure). When transformed responses deal with transformations not depending on the unknown regression model but only on the censoring distribution, an iterative procedure is no longer needed at the cost however of increased variability in the transformed data. Transformations of this type were proposed by [15], [22], [25] and [38]. The combination of non-linear mean regression models with synthetic data approaches for right censored data has mainly been studied for univariate covariates, see e.g. [15] and [18] among others. Recently more attention to multivariate regression models with right censored data transformation techniques is given by [37] for the VCM and by [3] for the varying coefficient partially linear model.

The paper is organized as follows. In Section 2, we introduce the VCM for randomly right censored data. The P-spline estimator in case censoring is absent is described in Section 3. Data transformation approaches for right censored data are explained in Section 4. Motivated by the research of [15], we first introduce model-independent transformations in Section 4.1 and later discuss, inspired by the approach of [4], transformations that take the underlying regression model into account in Section 4.2. The consistency and asymptotic normality of the proposed estimators is given in Section 5. Section 6 contains the details on how to choose the parameters involved with the estimation in practical settings. The finite sample behavior of the proposed method is analyzed using a simulation study in Section 7 and compared with the estimates for a VCM with right censoring proposed by [37]. The proposed estimation procedure is applied to the 'addict dataset' ([8]) in Section 8. The paper ends with a discussion, given in Section 9. An Appendix is included that contains the Assumptions of our main results. The technical details of our results are included in the Supplementary Materials.

2 Model description

Consider the varying coefficient model:

$$\begin{aligned} Y &= m(\mathbf{U},\mathbf{X}) + \sigma(\mathbf{U},\mathbf{X})\varepsilon \\ &= \beta_1(U_1)X_1 + \ldots + \beta_d(U_d)X_d + \sigma(U_1, X_1, \ldots, U_d, X_d)\varepsilon, \end{aligned} \tag{2.1}$$

where, Y is the response variable, $\mathbf{U} = (U_1, \ldots, U_d)' \in \mathcal{U}^d$ and $\mathbf{X} = (X_1, \ldots, X_d)' \in I\!R^d$ are associated covariate vectors, where $\mathcal{U}^d$ denotes a d-dimensional interval on which the measurements are taken; ε is a mean-zero error term with variance one and (unknown) distribution function F, assumed to be independent of $\mathbf{U}, \mathbf{X}$. The functions $\beta_1(u_1), \ldots, \beta_d(u_d)$ are the unknown regression coefficient functions at $\mathbf{U} = \mathbf{u} \equiv (u_1, \ldots, u_d)'$ and $\sigma(\mathbf{u}, \mathbf{x})$ is the variance of Y conditional on $\mathbf{U} = \mathbf{u}$ and $\mathbf{X} = \mathbf{x} = (x_1, \ldots, x_d)'$. When $X_1 \equiv 1$, the function β_1 is a non-zero intercept function representing the baseline effect.

We consider the case that the response Y of interest is subject to random right censoring. Let C be the censoring variable with survival function $G(\cdot|\mathbf{u}, \mathbf{x})$ conditional on $(\mathbf{U}, \mathbf{X}) = (\mathbf{u}, \mathbf{x})$ and Δ be the censoring indicator $1_{\{Y \leq C\}}$. We observe a sample $(Z_i, \Delta_i, \mathbf{U}_i, \mathbf{X}_i), i = 1, \ldots, n$, from $(Z, \Delta, \mathbf{U}, \mathbf{X})$. We assume throughout that Y and C are independent, non-negative continuous random variables.

In this paper we focus on estimating the regression curve $m(\mathbf{u}, \mathbf{x})$. The estimation procedure for $\boldsymbol{\beta}(u) = (\beta_1(u_1), \ldots, \beta_d(u_d))'$ consists of two steps: a mean-preserving data-transformation followed by P-spline smoothing using the transformed data. We describe the P-spline smoothing procedure with fully observed responses Y_i in Section 3 and describe in Section 4 two data-transformation approaches that allow a separation between the P-spline technique and the censored nature of the data.

3 P-spline estimator

Suppose that we have uncensored observations $(Y_i, \mathbf{U}_i, \mathbf{X}_i)$ for $i = 1, \ldots, n$. We use P-spline smoothing to estimate the varying coefficients in model (2.1). P-splines are an extension of regression splines with a penalty on the coefficients of adjacent B-splines. Each coefficient function β_p is approximated by a normalized B-spline basis expansion $\beta_p(u_p) \approx \sum_{l=1}^{m_p} B_{pl}(u_p; q_p)\alpha_{pl}$, where $\{B_{pl}(\cdot; q) : l = 1, \ldots, K_p + q_p = m_p\}$ is the q_p-th degree B-spline basis using normalized B-splines such that $\sum_{l=1}^{m_p} B_{pl}(u_p; q_p) = 1$, with $K_p + 1$ equidistant knots $\boldsymbol{\xi}_p = (\xi_{p0}, \ldots, \xi_{pK_p})$. We use the notation $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1', \ldots, \boldsymbol{\alpha}_d')'$ with $\boldsymbol{\alpha}_p = (\alpha_{p1}, \ldots, \alpha_{pm_p})'$ for $p = 1, \ldots, d$ to denote the unknown vector of B-spline regression coefficients and write $\mathcal{D} = \sum_{p=1}^{d} m_p$ for the dimension of $\boldsymbol{\alpha}$.

The P-spline optimization problem is given by

$$\min_{\boldsymbol{\alpha}} \left[\sum_{i=1}^{n} \left\{ Y_i - \sum_{p=1}^{d} X_{ip} \sum_{l=1}^{m_p} B_{pl}(U_i; q_p) \alpha_{pl} \right\}^2 + \sum_{p=1}^{d} \lambda_p \left(\sum_{l=k_p+1}^{m_p} (\Delta_p^k \alpha_{pl})^2 \right) \right]$$
$$= \min_{\boldsymbol{\alpha}} \left\{ (\mathbf{Y} - \mathbf{R}\boldsymbol{\alpha})'(\mathbf{Y} - \mathbf{R}\boldsymbol{\alpha}) + \alpha' \mathbf{Q}_{\boldsymbol{\lambda}} \boldsymbol{\alpha} \right\}, \tag{3.1}$$

where $\mathbf{Y} = (Y_1, \ldots, Y_n)'$, $\mathbf{R} = (\mathbf{R}_1 | \ldots | \mathbf{R}_n)' \in I\!R^{n \times \mathcal{D}}$ with $\mathbf{R}_i = \mathbf{B}'(\mathbf{U}_i)\mathbf{X}_i \in I\!R^{\mathcal{D} \times 1}$ and $\mathbf{B}(\mathbf{u}) \in I\!R^{d \times \mathcal{D}}$ given by,

$$\mathbf{B}(\mathbf{u}) = \begin{pmatrix} B_{11}(u_1; q_0) & \ldots & B_{1m_1}(u_1; q_1) & 0 \ldots 0 & 0 & \ldots & 0 \\ 0 & \ldots & 0 & \ddots & 0 & \ldots & 0 \\ 0 & \ldots & 0 & 0 \ldots 0 & B_{d1}(u_d; q_d) & \ldots & B_{dm_d}(u_d; q_d) \end{pmatrix},$$

$\mathbf{Q}_{\boldsymbol{\lambda}} = \text{diag}(\lambda_1 \mathbf{D}'_{k_1} \mathbf{D}_{k_1}, \ldots, \lambda_d \mathbf{D}'_{k_d} \mathbf{D}_{k_d}) \in I\!R^{\mathcal{D} \times \mathcal{D}}$, a block diagonal matrix with $\lambda_p \mathbf{D}'_{k_p} \mathbf{D}_{k_p}$ on the diagonal where $\mathbf{D}_{k_p}$ is the matrix representation of the k_p-th order difference operator Δ^{k_p}, i.e. $\Delta^{k_p}(\alpha_{pl}) = \sum_{h=0}^{k_p} (-1)^h \binom{k_p}{h} \alpha_{p(l-h)}$ (for $l \geq k_p$), with $k_p \in I\!N$; and $\lambda = (\lambda_1, \ldots, \lambda_d)$ is the vector of smoothing parameters satisfying $\lambda_p > 0, p = 1, \ldots, d$.

P-splines are computationally attractive since a closed form of the regression coefficient estimator exists. [1] showed that $\mathbf{R}'\mathbf{R} + \mathbf{Q}_\lambda$ is invertible except on a set with probability tending to zero if $\frac{m_{\max}^{3/2} \lambda_{\max}}{n} = o(1)$, where $m_{\max} = \max(m_1, \ldots, m_d)$ and $\lambda_{\max} = \max(\lambda_1, \ldots, \lambda_d)$. Therefore the unique minimizer of $S(\boldsymbol{\alpha})$ is

$$\hat{\boldsymbol{\alpha}} = \left(\mathbf{R}'\mathbf{R} + \mathbf{Q}_{\boldsymbol{\lambda}} \right)^{-1} \mathbf{R}'\mathbf{Y}. \tag{3.2}$$

The P-spline estimator of $\boldsymbol{\beta}(\mathbf{u})$ is

$$\hat{\boldsymbol{\beta}}(\mathbf{u}) = \mathbf{B}(\mathbf{u}) \hat{\boldsymbol{\alpha}} = (\hat{\beta}_1(u_1), \ldots, \hat{\beta}_d(u_d))', \text{ with } \hat{\beta}_p(u_p) = \sum_{l=1}^{m_p} B_{pl}(u_p; q_p) \hat{\alpha}_{pl}. \tag{3.3}$$

In Section 44, we construct, for randomly right censored data, a new response vector $\mathbf{Y}^*$ (the transformed response vector), that will replace $\mathbf{Y}$ in (3.2).

4 Data transformation approaches

We consider a data transformation approach and define the transformed response Y^* as

$$Y^* = \Delta\varphi(\mathbf{U},\mathbf{X},Z) + (1-\Delta)\psi(\mathbf{U},\mathbf{X},Z) = \begin{cases} \varphi(\mathbf{U},\mathbf{X},Z) & \text{if uncensored} \\ \psi(\mathbf{U},\mathbf{X},Z) & \text{if censored,} \end{cases}$$

with transformation functions φ and ψ so that

$$E(Y^*|\mathbf{U},\mathbf{X}) = E(Y|\mathbf{U},\mathbf{X}). \tag{4.1}$$

Condition (4.1) ensures that inference based on $(Y_i^*, \mathbf{U}_i, \mathbf{X}_i)$ preserves the conditional mean. In Section 4.1 we look at transformations that do not depend on the underlying regression model (2.1). Transformations that depend on model (2.1) are considered in Section 4.2. When a transformation depends on the unknown regression model, initial estimates for the regression curve and variance function are needed. In the second transformation method, we use as initial estimates for m and σ the estimates based on the model-independent transformation method of Section 4.1. We use the notation φ_1, ψ_1 and φ_2, ψ_2 to denote the transformation functions φ, ψ in methods one and two.

4.1 Transformation method 1: model independent transformations

From condition (4.1) we obtain the integral equation

$$\varphi_1(\mathbf{u},\mathbf{x},y)G(y|\mathbf{u},\mathbf{x}) - \int_0^y \psi_1(\mathbf{u},\mathbf{x},c)dG(c|\mathbf{u},\mathbf{x}) = y. \tag{4.2}$$

A specific class of solutions to (4.2), for all $y > 0, \mathbf{u} \in \mathcal{U}^d$ and $\mathbf{x} \in I\!R^d$ is given in [15], with $z > 0$ and $\gamma \in I\!R$,

$$\begin{aligned} \varphi_1(\mathbf{u},\mathbf{x},z) &= (1+\gamma)\int_0^z \frac{dt}{G(t|\mathbf{u},\mathbf{x})} - \gamma\frac{z}{G(z|\mathbf{u},\mathbf{x})}, \\ \psi_1(\mathbf{u},\mathbf{x},z) &= (1+\gamma)\int_0^z \frac{dt}{G(t|\mathbf{u},\mathbf{x})}. \end{aligned} \tag{4.3}$$

The transformations only depend on the censoring distribution $G(\cdot|\mathbf{u},\mathbf{x})$ of C conditional on $(\mathbf{U},\mathbf{X}) = (\mathbf{u},\mathbf{x})$. Special cases of (4.3) are the methods proposed by [22] and [25], taking $\gamma = -1$ and $\gamma = 0$ respectively. Since the functions φ_1 and ψ_1 depend on the unknown conditional survival function of C, an estimator $\hat{G}(\cdot|\mathbf{u},\mathbf{x})$ of $G(\cdot|\mathbf{u},\mathbf{x})$ is needed. A well-known problem with right censored data is however the estimation of a distribution function in the tail of the distribution. We therefore do not transform data points in the tail. As suggested in [15], we define

$$\hat{\varphi}_1(\mathbf{u},\mathbf{x},z) = \bar{\varphi}_1(u,\mathbf{x},z)1_{\{z\leq\tau_1(\mathbf{u},\mathbf{x})\}} + z1_{\{z>\tau_1(\mathbf{u},\mathbf{x})\}}$$
$$\hat{\psi}_1(\mathbf{u},\mathbf{x},z) = \bar{\psi}_1(u,\mathbf{x},z)1_{\{z\leq\tau_1(\mathbf{u},\mathbf{x})\}} + z1_{\{z>\tau_1(\mathbf{u},\mathbf{x})\}}$$

for some $0 < \tau_1(\mathbf{u},\mathbf{x}) < \mathcal{T}(\mathbf{u},\mathbf{x}) = \sup\{t|H(z|\mathbf{u},\mathbf{x}) < 1\}$ with $H(z|\mathbf{u},\mathbf{x}) = P(Z \leq z|\mathbf{U} = \mathbf{u},\mathbf{X} = \mathbf{x})$ representing the distribution function of Z conditional on $(\mathbf{U},\mathbf{X}) = (\mathbf{u},\mathbf{x})$; where $\bar{\varphi}_1$ and $\bar{\psi}_1$ are given by (4.3) with G replaced by the estimator $\hat{G}$.

The synthetic response vector is defined as $\hat{\mathbf{Y}}_1^* = (\hat{Y}_{1i}^*,\ldots,\hat{Y}_{1n}^*)'$ with, for $i = 1,\ldots,n$,

$$\hat{Y}_{1i}^* = \Delta_i\hat{\varphi}_1(\mathbf{U}_i,\mathbf{X}_i,Z_i) + (1-\Delta_i)\hat{\psi}_1(\mathbf{U}_i,\mathbf{X}_i,Z_i),$$

and the P-spline estimator of $m(\mathbf{u},\mathbf{x})$ in method 1 is

$$\hat{m}_1(\mathbf{u},\mathbf{x}) = \mathbf{x}'\hat{\boldsymbol{\beta}}_1(\mathbf{u}) \quad \text{with} \quad \hat{\boldsymbol{\beta}}_1(\mathbf{u}) = \mathbf{B}(\mathbf{u})\big(\mathbf{R}'\mathbf{R} + \mathbf{Q}_{\boldsymbol{\lambda}}\big)^{-1}\mathbf{R}'\hat{\mathbf{Y}}_1^* \tag{4.4}$$

Remark 1. *In regression analysis, one is often interested in modeling* $E(f(Y)|\mathbf{U},\mathbf{X}) = m_f(\mathbf{U},\mathbf{X})$. *For example, taking* $f(y) = y$ *gives model (2.1), and* $f(y) = 1_{\{y\leq t\}}$ *corresponds to estimating the conditional distribution function of* Y. *It is possible to modify transformation functions* φ_1 *and* ψ_1 *such that we are estimating the conditional mean* m_f, *where* f *is a bounded non-decreasing function on* $[0,\tau_1(\mathbf{u},\mathbf{x})]$, *by defining the functions*

$$\hat{\varphi}_{1,f}(\mathbf{u},\mathbf{x},z) = \left\{(1+\gamma)\int_0^z \frac{df(t)}{\hat{G}(t|\mathbf{u},\mathbf{x})} - \gamma\frac{f(z)}{\hat{G}(z|\mathbf{u},\mathbf{x})}\right\}1_{\{z\leq\tau_1(\mathbf{u},\mathbf{x})\}} + f(z)1_{\{z>\tau_1(\mathbf{u},\mathbf{x})\}}$$

and,

$$\hat{\psi}_{1,f}(\mathbf{u},\mathbf{x},z) = \left\{(1+\gamma)\int_0^z \frac{df(t)}{\hat{G}(t|\mathbf{u},\mathbf{x})}\right\} 1_{\{z\leq\tau_1(\mathbf{u},\mathbf{x})\}} + f(z)1_{\{z>\tau_1(\mathbf{u},\mathbf{x})\}}$$

and transformed responses

$$\hat{Y}^*_{1,f} = \Delta\hat{\varphi}_{1,f}(\mathbf{U},\mathbf{X},Z) + (1-\Delta)\hat{\psi}_{1,f}(\mathbf{U},\mathbf{X},Z). \tag{4.5}$$

The modified transformation formula is also suited for estimating the conditional variance of Y, i.e. $f(t) = (t - m(\mathbf{u},\mathbf{x}))^2$, when $\gamma = -1$, since for $\gamma = -1$, the non-decreasing condition for f is not necessary (see e.g. [12]). As a consequence, when a varying coefficient model for $\sigma^2(\mathbf{U},\mathbf{X})$ is assumed, we can obtain a consistent estimate for $\sigma^2(\mathbf{u},\mathbf{x})$ by constructing

$$\hat{Y}^*_{1,\sigma^2} = \frac{\Delta(Z-\hat{m}_1(\mathbf{U},\mathbf{X}))^2}{\hat{G}(Z)} 1_{\{Z\leq\tau_1(\mathbf{U},\mathbf{X})\}} + (Z-\hat{m}_1(\mathbf{U},\mathbf{X}))^2 1_{\{Z>\tau_1(\mathbf{U},\mathbf{X})\}}.$$

An estimate of $\sigma^2(\mathbf{u},\mathbf{x})$ is given by

$$\hat{\sigma}^2_1(\mathbf{u},\mathbf{x}) = \mathbf{x}'\mathbf{B}_{\sigma^2}(\mathbf{u})\big(\mathbf{R}'_{\sigma^2}\mathbf{R}_{\sigma^2} + \mathbf{Q}_{\boldsymbol{\lambda},\sigma^2}\big)^{-1}\mathbf{R}'_{\sigma^2}\hat{\mathbf{Y}}^*_{1,\sigma^2} \tag{4.6}$$

where the matrices $\mathbf{B}_{\sigma^2}, \mathbf{R}_{\sigma^2}$ and $\mathbf{Q}_{\boldsymbol{\lambda},\sigma^2}$ are the matrices $\mathbf{B}, \mathbf{R}$ and $\mathbf{Q}_{\boldsymbol{\lambda}}$ (introduced in Section 3) according to the model for σ^2. Another approach could be to estimate $E(Y^2|\mathbf{U},\mathbf{X})$ and considering the difference $E(Y^2|\mathbf{U},\mathbf{X}) - (E(Y|\mathbf{U},\mathbf{X}))^2$. Note that we are not restricted to transformations with $\gamma = -1$ when we are estimating the conditional expectation of Y^2 since the function $f(t) = t^2$ in increasing on $I\!R^+$. Although the latter approach gives a consistent estimator for the variance function, in practice, numerical difficulties arise by taking the differences, since the difference is not guaranteed to be positive in finite samples.

4.2 Transformation method 2: model dependent transformations

Based on the expression for the conditional expectation,

$$E(Y|Z,\Delta,\mathbf{U},\mathbf{X}) = \Delta Z$$

$$+ (1-\Delta)\left\{ m(\mathbf{U},\mathbf{X}) + \frac{\sigma(\mathbf{U},\mathbf{X})}{1-F\left(\frac{Z-m(\mathbf{U},\mathbf{X})}{\sigma(\mathbf{U},\mathbf{X})}\right)} \int_{(Z-m(\mathbf{U},\mathbf{X}))/\sigma(\mathbf{U},\mathbf{X})}^{\infty} t dF(t) \right\},$$

it follows that $E\big(Y_{2[0]}^{*}|\mathbf{U},\mathbf{X}\big) = E\big(Y|\mathbf{U},\mathbf{X}\big)$, for

$$Y_{2[0]}^{*} = \Delta\varphi_{2[0]}^{*}(\mathbf{U},\mathbf{X},Z) + (1-\Delta)\psi_{2[0]}^{*}(\mathbf{U},\mathbf{X},Z),$$

where $\varphi_{2[0]}^{*}(\mathbf{U},\mathbf{X},Z) = Z$ and

$$\psi_{2[0]}^{*}(\mathbf{U},\mathbf{X},Z) = m(\mathbf{U},\mathbf{X}) + \frac{\sigma(\mathbf{U},\mathbf{X})}{1-F\left(\frac{Z-m(\mathbf{U},\mathbf{X})}{\sigma(\mathbf{U},\mathbf{X})}\right)} \int_{(Z-m(\mathbf{U},\mathbf{X}))/\sigma(\mathbf{U},\mathbf{X})}^{\infty} t dF(t).$$

In order to construct an estimator $\hat{\psi}_2$ of $\psi_{2[0]}^{*}$, we again consider a truncation device that avoids problems associated with the instability of an estimator for F. We follow the idea of [18] and define ψ_2 and Y_2^{*} as follows:

$$\begin{aligned} \psi_2(\mathbf{U},\mathbf{X},Z) &= m(\mathbf{U},\mathbf{X}) + \frac{\sigma(\mathbf{U},\mathbf{X})}{1-F(E^T)} \int_{E^T}^{S} t dF(t), \\ Y_2^{*} &= \Delta\varphi_2(\mathbf{U},\mathbf{X},Z) + (1-\Delta)\psi_2(\mathbf{U},\mathbf{X},Z), \end{aligned} \tag{4.7}$$

where the truncated residual $E^T = \min(E,S)$ with

$$E = \frac{Z-m(\mathbf{U},\mathbf{X})}{\sigma(\mathbf{U},\mathbf{X})} \quad \text{and} \quad S = \frac{\tau_2(\mathbf{U},\mathbf{X})-m(\mathbf{U},\mathbf{X})}{\sigma(\mathbf{U},\mathbf{X})},$$

for some $\tau_2(\mathbf{u},\mathbf{x}) < \mathcal{T}(\mathbf{u},\mathbf{x})$. Let

$$\hat{E} = \frac{Z-\hat{m}_1(\mathbf{U},\mathbf{X})}{\hat{\sigma}_1(\mathbf{U},\mathbf{X})}, \quad \hat{S} = \frac{\tau_2(\mathbf{U},\mathbf{X})-\hat{m}_1(\mathbf{U},\mathbf{X})}{\hat{\sigma}_1(\mathbf{U},\mathbf{X})} \quad \text{and} \quad \hat{E}^T = \min(\hat{E},\hat{S}).$$

We obtain the estimator $\hat{\psi}_2$ by replacing, in (4.7), m and σ by $\hat{m}_1$ and $\hat{\sigma}_1$, defined in (4.4) and (4.6), E^T and S by $\hat{E}^T$ and $\hat{S}$ and by replacing F by the Kaplan-Meier type estimator

$\hat{F}$, constructed with residual observations $\hat{E}_i$, i.e.

$$\hat{F}(t) = 1 - \prod_{i:\hat{E}_i \leq t} \left(1 - \frac{1}{\sum_{j=1}^n 1_{\{\hat{E}_j \geq \hat{E}_i\}}} \right)^{\Delta_i},$$

The transformed response vector $\hat{\mathbf{Y}}_2^* = (\hat{Y}_{21}^*, \ldots, \hat{Y}_{2n}^*)'$ is defined by,

$$\hat{Y}_{2i}^* = \Delta_i Z_i + (1 - \Delta_i)\hat{\psi}_2(\mathbf{U}_i, \mathbf{X}_i, Z_i). \tag{4.8}$$

The P-spline estimator $\hat{\boldsymbol{\beta}}_2(\mathbf{u})$ of $\boldsymbol{\beta}(\mathbf{u})$ in method 2 is obtained by replacing $\mathbf{Y}$ in (3.3) by $\hat{\mathbf{Y}}_2^*$.

Remark 2. *Note that, for method 1, $E(Y_1^*|\mathbf{U}, \mathbf{X}) = E(Y|\mathbf{U}, \mathbf{X})$ if $Z \leq \tau_1(\mathbf{U}, \mathbf{X})$ but for method 2 (as in [18]), $E(Y_2^*|\mathbf{U}, \mathbf{X}) \neq E(Y|\mathbf{U}, \mathbf{X})$, since we truncate the integral in (4.7) and as a consequence we estimate a truncated mean $E(Y 1_{\{\varepsilon \leq S\}}|Z, \Delta, \mathbf{U}, \mathbf{X})$. The conditional expectation of Y_2^* will, however, be arbitrarily close to the conditional expectation of Y if S can be chosen arbitrarily close to $\tau_F = \sup\{t|F(t) < 1\}$, which is possible when $\tau_F \leq \tau_J$, where J is the distribution function of $\{C - m(\mathbf{U}, \mathbf{X})\}/\sigma(\mathbf{U}, \mathbf{X})$ and $\tau_J = \sup\{t|J(t) < 1\}$.*

5 Asymptotic behavior

In Theorem 1, we show the consistency of the P-spline estimators obtained under transformation methods 1 and 2. The asymptotic normality of the estimators is considered in Theorem 2. Before stating the main results, we first give the following definition.

Definition 1. *Let $\mathbb{G}(q_p, \boldsymbol{\xi}_p)$ be the space of spline functions on $\mathcal{U}_p$ with fixed degree q_p and knot sequence $\boldsymbol{\xi}_p$. Let dist$(\beta_p, \mathbb{G}(q_p, \boldsymbol{\xi}_p)) = \inf_{g \in \mathbb{G}(q_p, \boldsymbol{\xi}_p)} \sup_{u \in \mathcal{U}} \mid \beta_p(u) - g(u) \mid$ be the L_∞-distance between β_p and $\mathbb{G}(q_p, \boldsymbol{\xi}_p)$. The approximation error due to spline approximation is given by*

$$\rho_n = \max_{1 \leq p \leq d} dist(\beta_p, \mathbb{G}(q_p, \boldsymbol{\xi}_p)).$$

We use the notations $\hat{\boldsymbol{\beta}}_j = (\hat{\beta}_{j1}, \ldots, \hat{\beta}_{jd})'$, $\boldsymbol{\beta}_j^* = (\beta_{j1}^*, \ldots, \beta_{jd}^*)'$ and $\tilde{\boldsymbol{\beta}}_j = (\tilde{\beta}_{j1}, \ldots, \tilde{\beta}_{jd})'$ for methods $j = 1, 2$, when we replace $\mathbf{Y}$ in expression (3.3) by $\hat{\mathbf{Y}}_j^* = (\hat{Y}_{j1}^*, \ldots, \hat{Y}_{jn}^*)'$,

$\mathbf{Y}_j^* = (Y_{j1}^*, \ldots, Y_{jn}^*)'$, and $\mathbf{M} = (M_{j1}, \ldots, M_{jn})'$ with $M_{ji} = E(Y_{ji}^*|\mathbf{U}_i, \mathbf{X}_i)$ for $i = 1, \ldots, n$ respectively. Note that $E(\boldsymbol{\beta}_j^*|\mathcal{X}_n) = \tilde{\boldsymbol{\beta}}_j$ for $j = 1, 2$ where $\mathcal{X}_n = \{(\mathbf{U}_i', \mathbf{X}_i')', i = 1, \ldots, n\}$. See the Appendix for the definition of the L_2-distance and for Assumptions A-D in Theorems 1 and 2.

Theorem 1. *Suppose Assumptions A, B.1 and B.2 hold, then,*

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

where $\kappa(\mathbf{u},\mathbf{x})$ is given by,

$$\max_{\phi=\varphi_1,\psi_1} \left[E\left\{ 1_{\{Z > \tau_1(\mathbf{U},\mathbf{X})\}} |Z - \phi(\mathbf{U},\mathbf{X},Z)| \middle| \mathbf{U} = \mathbf{u}, \mathbf{X} = \mathbf{x} \right\} \right].$$

If, further Assumptions B.3 and C hold, then,

$$\begin{aligned}\|\hat{\boldsymbol{\beta}}_2 - \tilde{\boldsymbol{\beta}}_2\|_{L_2} = O_p\Bigg(& n^{-1/2} m_{max}^{1/2} + n^{-1/2}\log n + n^{-1} m_{max}^{3/2} \lambda_{max} + \rho_n + \\ m_{max}^{-1/2} \Bigg[& \sup_{\mathbf{u},\mathbf{x}} \left\{ \tau_1(\mathbf{u},\mathbf{x}) \sup_{t \le \tau_1(\mathbf{u},\mathbf{x})} |\hat{G}(t|\mathbf{u},\mathbf{x}) - G(t|\mathbf{u},\mathbf{x})| + \kappa(\mathbf{u},\mathbf{x}) + \kappa_\sigma(\mathbf{u},\mathbf{x}) \right\} \Bigg] \Bigg).\end{aligned}$$

where $\kappa_\sigma(\mathbf{u},\mathbf{x})$ is given by,

$$E\left\{ 1_{\{Z > \tau_1(\mathbf{U},\mathbf{X})\}} \left(Z - m(\mathbf{U},\mathbf{X},Z)\right)^2 |1 - \Delta / G(Z|\mathbf{U},\mathbf{X})| \middle| \mathbf{U} = \mathbf{u}, \mathbf{X} = \mathbf{x} \right\}.$$

Remark 3. *If $\sup_{\mathbf{u},\mathbf{x}} \kappa(\mathbf{u},\mathbf{x}) \to 0$, the tail-contribution is negligible and the truncation device is justified. This condition was first introduced by [15] and suggests taking $\tau_1(\mathbf{u},\mathbf{x})$ as a sequence converging to $\mathcal{T}(\mathbf{u},\mathbf{x})$. If, e.g., conditional on $(\mathbf{U},\mathbf{X}) = (\mathbf{u},\mathbf{x})$, $Y \sim Exp(\theta_{\mathbf{u},\mathbf{x}})$ and $C \sim Exp(\nu)$ are independent exponentially distributed random variables, then $\kappa(\mathbf{u},\mathbf{x}) = O\left(n^{-\theta_{u,\mathbf{x}}} \log n\right)$ by taking $\tau_1(\mathbf{u},\mathbf{x}) = \log n$ for all $\mathbf{u},\mathbf{x}$. As another illustration, suppose $Y \sim U[0, \theta_{\mathbf{u},\mathbf{x}}]$ conditional on $(\mathbf{U},\mathbf{X}) = (\mathbf{u},\mathbf{x})$ and $C \sim U[0, \nu]$ are independent uniform ran-*

dom variables. After some tedious calculations we can show that $\kappa(\mathbf{u},\mathbf{x}) \to 0$ *for* $\tau_1(\mathbf{u},\mathbf{x}) = n^{-1}(n-1)\theta_{\mathbf{u},\mathbf{x}}$ *and* $\theta_{\mathbf{u},\mathbf{x}} \leq \nu$. κ_σ *arises similarly when method 1 is used to estimate* σ *using the transformation with* $\gamma = -1$.

Remark 4. *Suppose that each* β_p *is an* r *times continuously differentiable function (*$p = 1,\ldots,d$*), if* $q = q_p \geq r-1$, $m_{max} \asymp n^{1/(2r+1)}$ *and* $\lambda_{max} \asymp n^{\iota}$ *with* $\iota \leq (r-1/2)/(2r+1)$*, then* $\|\beta_p^* - \beta_p\|_{L_2} = O_p(n^{-r/(2r+1)})$ *reaches the optimal rate of convergence for non-parametric regression estimators. ([32]). The convergence rate of our P-spline estimator* $\hat{\beta}_p^*$ *is further influenced by the censored nature of the data.*

Theorem 2 gives the asymptotic normality results of the P-spline estimator. The variance-covariance matrix of $\boldsymbol{\beta}_j^*(\mathbf{u})$, conditional on $\mathcal{X}_n = \{(\mathbf{U}_i', \mathbf{X}_i')', i = 1,\ldots,n\}$, is given by,

$$\mathbf{B}(\mathbf{u})\big(\mathbf{R}'\mathbf{R} + \mathbf{Q}_{\boldsymbol{\lambda}}\big)^{-1}\Big(\sum_{i=1}^{n} \sigma_{j,i}^2 \mathbf{R}_i \mathbf{R}_i'\Big)\big(\mathbf{R}'\mathbf{R} + \mathbf{Q}_{\boldsymbol{\lambda}}\big)^{-1}\mathbf{B}'(\mathbf{u}), \tag{5.1}$$

where $\sigma_{j,i}^2 = \mathrm{Var}(Y_{ji}^*|\mathbf{U}_i, \mathbf{X}_i)$.

Theorem 2. *If Assumptions A, B.1, B.2 and D.1 hold, then, for all* $u_p \in \mathcal{U}_p$, $p = 1,\ldots,d$,

$$\big(s.e.\,\big(\beta_{1,p}^*(u_p) \mid \mathcal{X}_n\big)\big)^{-1}\Big(\hat{\beta}_{1,p}(u_p) - \beta_p(u_p)\Big) \xrightarrow{d} N(0,1).$$

If Assumptions A, B, C and D.2 hold, then, for all $u_p \in \mathcal{U}_p$, $p = 1,\ldots,d$

$$\big(s.e.\,\big(\beta_{2,p}^*(u_p) \mid \mathcal{X}_n\big)\big)^{-1}\Big(\hat{\beta}_{2,p}(u_p) - \tilde{\beta}_{2,p}(u_p)\Big) \xrightarrow{d} N(0,1).$$

6 Practical technicalities

6.1 Choice of the truncation points

We estimate the functional regression coefficients in VCM (2.1) by a combination of a data transformation for censored data and the P-spline estimator for complete case data. The proposed data transformations involve an estimator of a distribution function. In the presence of censoring, non-parametric estimators of a distribution function are often inaccurate in the tail. To control this instability we use a truncation device that avoids the generation of transformed data in the tail.

In a clinical trial, censoring is often due to the termination of the study and hence not influenced by patient specific characteristics. In such situations the conditional survival function of C does not depend on the covariates, i.e. $G(\cdot|\mathbf{u},\mathbf{x}) \equiv G(\cdot)$, and the Kaplan-Meier product-limit estimator can be used to estimate the survival distribution of the censoring variable C. Note that, when estimating the censoring distribution G, the independent but non-identically distributed event times $Y_i, i = 1, \ldots n$ now play the role of censoring variables. For such situation the uniform strong consistency of the Kaplan-Meier estimator is still valid (see e.g. [40] and [3]). If censoring is informative, but Y and C are conditionally independent given $\mathbf{U}, \mathbf{X}$, the conditional (on $\mathbf{U}, \mathbf{X}$) distribution of C, should be estimated in method 1, using, for example, the [2] estimator. However this may cause problems with the curse of dimensionality and one may want to consider a parametric or semi-parametric model for the censoring distribution instead.

In method 1, we do not transform data points when the observed response Z falls within the truncation area (τ_1, ∞). Choosing τ_1 too small implies that a lot of observations will not be transformed. On the other hand when τ_1 is chosen too large, large transformed responses are possible. In our numerical results we consider a censoring variable C independent of $(\mathbf{U}, \mathbf{X})$. We take $\tau_1 = \inf\{t|\hat{G}(t) < 0.01\}$ for method 1 and suggest to consider all jumps of the Kaplan-Meier estimator $\hat{F}$ in method 2 by taking $\hat{S} = \max(\hat{E}_1, \ldots, \hat{E}_n)$.

6.2 Smoothing parameter selection

Smoothing parameters are needed to control the amount of smoothing in the estimation process and imply a compromise between bias and variance. Undersmoothing arises by choosing too small values for the smoothing parameters, as a result, the bias will decrease at the price of an increased variance. When the smoothing parameters are too large, oversmoothing leads to a small variance but large bias (see [14, p. 187]). Cross-validation (CV) is a popular parameter selection technique with complete case data based on minimizing the prediction error. With censored data, the prediction error cannot be calculated directly. We suggest to consider the transformed responses and choose the smoothing parameter $\boldsymbol{\lambda}$ that minimizes,

$$CV(\boldsymbol{\lambda}) = \sum_{i=1}^{n} \left\{ \frac{\hat{Y}_{ji}^{*} - \mathbf{X}_i' \hat{\boldsymbol{\beta}}_j(\mathbf{U}_i)}{1 - h_{ii}} \right\}^2 ,$$

where h_{ii} is the i-th diagonal element of the hat-matrix $H = \mathbf{R}(\mathbf{R}'\mathbf{R} + \mathbf{Q}_{\boldsymbol{\lambda}})^{-1}\mathbf{R}'$. The idea of using transformed responses in the prediction error calculation was also considered in [15] and [34] among others. In practice, $CV(\boldsymbol{\lambda})$ is minimized over a d-dimensional grid of λ-values. With P-spline smoothing it is advisable to first consider a grid of the smoothing parameters on a logarithmic scale, which can later be fine-tuned when a more accurate smoothing parameter is desirable. Note that the P-spline estimator of β_p depends on the degree of the B-spline basis q_p, the number of knots $K_p + 1$, the order of the difference penalty k_p and the smoothing parameter λ_p. Cross-validation can be used to select several parameters, however, a good chosen smoothing parameter for fixed values of q_p, K_p and k_p will ensure a good fit. Cubic splines and a second order difference penalty are frequently used. A change in one of the parameters influences the choice of the other parameters, as a consequence, it is sufficient to select the smoothing parameters and keep the other parameters fixed.

6.3 Transformation parameter selection in method 1

The transformation parameter γ in method 1 determines the synthetic responses. We suggest to choose γ in a data-driven way. A cross-validation procedure can simultaneously select the smoothing parameter $\boldsymbol{\lambda}$ and transformation parameter γ when we search over a $(d+1)$-dimensional grid.

A second selection technique for the transformation parameter γ is based on the following observation. For $\gamma = -1$, all censored observations less than τ_1 are set equal to zero ($\psi_1 \equiv 0$), the uncensored observations are enlarged in order to compensate. If γ increases, we see that the variance of censored observations increases and that the enlargement of the uncensored observations is less pronounced (see Table 7.5). Therefore, we propose to select the transformation parameter γ that minimizes the sample variance of the transformed responses, denoted by the minimal-variance (MV) parameter γ_{MV}. Compared to CV-selection, the MV-selection procedure is appealing for being not computational intensive.

7 Finite sample behavior

In this section, we illustrate the finite sample behavior of our proposed P-spline estimates for VCMs when the observations are subject to random right censoring. Simulation studies are used to address the following objectives:

1. Compare our P-spline method with the smooth-backfitting (SBF) approach of [37].
2. Investigate the quality of the data-transformation methods given in Section 4.
3. Evaluate the cross-validation selection criterion for the P-spline smoothing parameters.

We consider three different simulation scenarios. The first model is also used in [24] and in [37] and contrasts the performance between a spline smoothing and kernel approach for model-independent data transformation techniques. The second and third simulation model illustrate how model-dependent transformations increase the performance of model-independent approaches. The main difference between the two latter models is the nature of the random error terms which is homoscedastic in Model 2 and hetereoscedastic in Model 3. Therefore, Model 3 also gives insight in the quality of the variance estimation discussed in Remark 1. The simulation scenarios are as follows:

Model 1: $Y = m(\mathbf{U}, \mathbf{X}) + \sigma(\mathbf{U}, \mathbf{X})\varepsilon = \beta_0(U_0) + \beta_1(U_1)X_1 + \beta_2(U_2)X_2 + \sigma(\mathbf{U}, \mathbf{X})\varepsilon$, where $\beta_0(u) = 1 + \exp(2u - 1), \beta_1(u) = 0.5\cos(2\pi u), \beta_2(u) = u^2$ and $\sigma(\mathbf{U}, \mathbf{X}) = 0.5 + (x_1^2 + x_2^2)/(1 + x_1^2 + x_2^2)\exp(-2 + (u_0 + u_1)/2)$. The variables U_0, U_1, and U_2 are sampled from a Uniform[0, 1]-distribution, the vector (X_1, X_2) is generated from a bivariate normal distribution with mean $(0, 0)'$ and variance-covariance matrix $\begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$, and the random error has a normal distribution centered around 0 with standard deviation $\zeta = 1$ respectively $\zeta = 1.5$. The censoring variables are generated samples from a $\mathrm{N}(\mu_c, 1.5)$-distribution.

Model 2: $Y = m(\mathbf{U},\mathbf{X}) + \varepsilon = \beta_1(U_1)X_1 + \beta_2(U_2)X_2 + \varepsilon$, where $\beta_1(u) = 2 + \sin(2\pi u)$, $\beta_2(u) = 1 + 0.1\exp(4x-1)$ with $U_1, U_2 \sim \mathrm{U}[0,1]$ and $(X_1, X_2)' \sim \mathrm{N}_2((3,3)', \left(\begin{smallmatrix} 0.25 & -0.125 \\ -0.125 & 0.25 \end{smallmatrix}\right))$; ε has a standard normal distribution and the censoring variable has a uniform distribution on $[6.5, R_c]$.

Model 3: $Y = m(U,X) + \sigma(U)\varepsilon = \beta_0(U) + \beta_1(U)X + \sigma(U)\varepsilon$, where $\beta_0(u) = 2\exp(-2u - u^2))$, $\beta_1(u) = 1 + 5(u-0.5)^2$ and $\sigma^2(u) = \alpha\exp(-2u-0.4)/4$ where $\alpha = 1, 2$. We generate U from a Uniform$[0,1]$-distribution and X from a normal distribution with mean 1 and standard deviation 0.25; ε has a standard normal distribution and C is sampled from a $\mathrm{N}(\mu_c, 1)$-distribution

The parameters μ_c (in Model 1 and 3) and R_c (in Model 2) are chosen to control the level of censoring to $PC = 10\%, 30\%$ and 50%, respectively. No negative responses are observed in these simulation set-ups in correspondence to our model assumptions. We simulate 200 times a random sample of size $n = 250, 500$ from Models 2 and 3. For Model 1, we consider the exact same simulation settings as in [37] and generate 500 samples of sizes $n = 200, 400$.

To evaluate the performance of the coefficient function estimates, we generate a uniform test sample $u_1, \ldots, u_{101}$ in $[0,1]$ for the random variables U_j and calculate the values for β_j and $\hat{\beta}_j$ in each simulation run. We then compute the relative error (RE) defined as (for $\hat{\beta}_j$),

$$\mathrm{RE}(\hat{\beta}_j) = \|\hat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j\|_2 / \|\boldsymbol{\beta}_j\|_2,$$

with $\boldsymbol{\beta}_j = (\boldsymbol{\beta}_j(u_1), \ldots, \boldsymbol{\beta}_j(u_{101}))'; \hat{\boldsymbol{\beta}}_j = (\hat{\boldsymbol{\beta}}_j(u_1), \ldots, \hat{\boldsymbol{\beta}}_j(u_{101}))'$ and where $\|\cdot\|_2$ is the Euclidean distance. For the performance of the regression function estimate $\hat{m}$, we generate a test sample $\mathbf{x}_j = (x_{1j}, x_{2j})$, $j = 1, \ldots, 101$, calculate $m_j = m(\mathbf{u}_j, \mathbf{x}_j)$ and $\hat{m}_j = \hat{m}(\mathbf{u}_j, \mathbf{x}_j)$, and compute the relative estimation error $\mathrm{RE}(\hat{m}) = \|\hat{\mathbf{m}} - \mathbf{m}\|_2 / \|\mathbf{m}\|_2$, where $\mathbf{m} = (m_1, \ldots, m_{101})$ and $\hat{\mathbf{m}} = (\hat{m}_1, \ldots, \hat{m}_{101})$. Tables 7.1- 7.2, Table 7.3 and Table 7.6 report the RE for the three simulation models introduced above.

We smooth each of the coefficient functions β_j with B-splines of degree 3 on 10 equidistant knots and use a penalty term with second order finite differences. The smoothing parameters λ_j are selected in a grid of size 8^d, where d equals the number of coefficient functions in the different simulation models. The CV-smoothing parameters (Section 6.2) are compared with optimal smoothing parameters that minimize the relative estimation error of the regression function m, referred to as the optimal selection criterion. Moreover, we present results for the smooth-backfitting estimates, where the optimal selection criterion is used to choose the bandwidths in a grid of equal size 8^d.

The simulation results, reported in Tables 7.1-7.6 and Figures 7.1-7.2, are discussed in the subsections below. The first objective is considered in Section 7.1. The importance of the transformation parameter selection in method 1 and the difference between model-dependent and model-independent transformations (objective 2) are outlined in Section 7.2. Section 7.3 addresses objective 3 and deals with the quality of the cross-validation smoothing parameter.

7.1 Comparison between P-spline and SBF-estimates

[37] proposed a smoothing estimation approach for the VCMs with right censored responses. Their technique is a kernel analogue of the model-independent transformation method of Section 4.1 that combines an SBF-estimator with the transformation method proposed by [22] using $\gamma = -1$. It is reasonable to compare our P-spline estimates using transformation method 1 with $\gamma = -1$ with the method proposed by [37] since in both approaches the transformed response variable and covariates are the same. Table 7.1 and Table 7.2 therefore contrast the RE of a P-spline and kernel smoothing approach for the simulation scenario considered in [37]. The SBF-estimates of [37] (SBF, $M1_K$) perform often slightly better than the P-spline estimates with $\gamma = -1$ (P-SPLINE, $M1_K$) in Model 1. In Model 2, the P-spline estimates, however, outperform the SBF-estimates for PC = 10%, 30% (see Table 7.3).

In addition, we investigate the combination of an SBF-estimate with a data-driven MV - transformation parameter, instead of with the transformation proposed by [22]. The relative errors for both the P-spline and SBF-estimates decrease considerably if $\gamma = -1$ is changed to $\gamma = \gamma_{MV}$ (see Tables 7.1-7.3). We conclude from this decrease that the method proposed in [37] can be improved if a different transformation parameter is considered. Interestingly, the choice between a P-spline smoothing or kernel smoothing approach has much less influence on the behavior of the estimates than the transformation parameter that is selected for the construction of the synthetic response. For the model-independent transformation methods, both combinations of a P-spline or SBF approach with a data-driven transformation parameter represent good choices for estimating the coefficient functions in the VCM under right censored observations. As expected, the relative errors in Tables 7.1-7.3 decrease with increasing sample size. On the contrary, an increase of the relative errors occurs if the percentages of censoring or the error variability increase.

From a theoretical point of view, both our P-spline and the SBF-estimates of [37] converge at rate $n^{2/5}$ to a normal limiting distribution for suitably chosen smoothing parameters in case the censoring distribution is known and in case the coefficient functions are twice continuously differentiable (see Remark 4 and Lemma 1 in [37]). The difference between the true and estimated coefficient functions depends further on the approximation error of the censoring distribution for both P-spline and SBF-estimates. Hence, the choice between our P-spline approach and the SBF method of [37] is hardly decided by the theoretical properties of the estimators. From a practical point of view, we note that P-spline estimates are obtained using simple matrix algebra whereas SBF-estimates require an iterative estimation procedure. The computations for the model-independent data transformation approaches took only a few seconds for the P-spline estimates and was, on average, 22 times larger for the SBF method than for the P-spline method in Simulation Model 2 (results not shown).

7.2 Findings on the transformation method

For the model-independent transformation method 1 of Section 4.1, Tables 7.1-7.3 show that a data-driven choice for the transformation parameter γ decreases the RE of the estimates compared to the choice $\gamma = -1$. Moreover, Table 7.3 shows how the estimates for transformation method 1 with the transformation by [22] ($\gamma = -1$) behave worse than the estimates that are obtained when censoring is ignored (i.e. Z is considered as the true response). Consequently, we do not advise to use the transformation approach by [22]. Similar relative errors are obtained with the proposed data-driven transformations (MV and CV), with a slightly better result for the CV-method when the percentage of censoring is large. The computation cost for CV-selection is, however, considerably larger than for MV-selection. Therefore, we recommend to use the MV-transformation parameter when method 1 is used to obtain the synthetic response variable.

Table 7.3 and Table 7.6 report the performance of the model-dependent transformation method of Section 4.2 in case the initial starting estimates are obtained from the model-independent transformation method using $\gamma = \gamma_{MV}$. Transformation method 2 outperforms transformation method 1 for both the homoscedastic Model 2 and the heteroscedastic Model 3. Pointwise confidence bands of the P-spline estimates in Model 2 are illustrated in Figure 7.1. The curves show the 5% and 95% empirical quantiles at each grid point u_j and expose that the estimates obtained with method 2 are close to the true coefficient functions, even though in theory, method 2 is estimating a slightly different model. The results of method 2 are insensitive towards changes of γ in the initial transformation (results not shown). Additionally, Figure 7.1 shows once more the poor performance of the model-independent estimates using $\gamma = -1$.

7.3 Behavior of the smoothing parameter selection techniques

Table 7.4 presents the ratio of the relative error for m obtained with CV-selected smoothing parameters and optimal smoothing parameters in simulation Model 2 and illustrates that the CV-procedure works reasonably well (the ratio is close to one). Figure 7.2 presents scaled values of $\mathrm{CV}(\lambda_1, \lambda_2)$ and relative error of m for λ_1 and λ_2 (in Model 2) varying in $10^{\{0.5,0.6,\ldots,2.6\}}$ and demonstrates that the size of the CV-selected and optimal smoothing parameters are comparable. The behavior of both curves is similar. As a consequence the CV-method tends to select smoothing parameters that minimize the relative regression error for m. A data-driven bandwidth choice for the bandwidths of the SBF-estimates was proposed in [37] and based on their results in Table 4 on p. 243, their comparison between the performance with optimal and data-driven bandwidth parameters is similar to our comparison in Table 7.4.

8 Real data example: Addict data

In a study by [8] data were collected on a cohort of 238 heroin addicts, who entered maintenance programs between February 1986 and August 1987, to study retention of patients in methadone treatment. All patients had been referred to one of two methadone treatment

Table 7.1: Simulation Model 1: average relative error for the estimates of the functions (F) $\beta_0, \beta_1, \beta_2$ and m obtained with the P-spline estimator and the smooth-backfitting estimator (SBF) with optimal smoothing parameters; using transformation method 1 (M1) (M1$_{MV}$: M1 with minimal-variability transformation, M$_K$: M1 with transformation by [22] using $\gamma = -1$). n is the sample size, $\zeta = s.d.(\varepsilon)$ and PC is the percentage of censoring.

				P-SPLINE		SBF	
n	ζ	PC	F	M1$_{MV}$	M1$_K$	M1$_{MV}$	M1$_K$
200	1	10	β_0	0.0374	0.0728	0.0424	0.0721
			β_1	0.3303	0.7483	0.3555	0.7711
			β_2	0.2147	0.5440	0.2074	0.4799
			m	0.0742	0.1715	0.0781	0.1631
		30	β_0	0.0539	0.1517	0.0574	0.1334
			β_1	0.4443	1.3628	0.4856	1.3090
			β_2	0.3082	0.9922	0.2910	0.8652
			m	0.1039	0.3189	0.1072	0.2853
		50	β_0	0.0812	0.2594	0.0812	0.2232
			β_1	0.6312	2.0543	0.6832	1.8961
			β_2	0.4640	1.4777	0.4276	1.2512
			m	0.1543	0.4898	0.1530	0.4208
	1.5	10	β_0	0.0539	0.1020	0.0574	0.0938
			β_1	0.4490	0.9118	0.4802	0.9152
			β_2	0.3028	0.7186	0.2791	0.6345
			m	0.1034	0.2223	0.1049	0.2045
		30	β_0	0.0707	0.1903	0.0714	0.1658
			β_1	0.5513	1.5512	0.6005	1.4506
			β_2	0.3918	1.1901	0.3586	1.0379
			m	0.1323	0.3792	0.1319	0.3330
		50	β_0	0.1010	0.3069	0.0938	0.2648
			β_1	0.7315	2.2477	0.7752	2.0514
			β_2	0.5446	1.7031	0.4926	1.4363
			m	0.1836	0.5545	0.1746	0.4740

Table 7.2: Simulation Model 1: average relative error for the estimates of the functions (F) $\beta_0, \beta_1, \beta_2$ and m obtained with the P-spline estimator and the smooth-backfitting estimator (SBF) with optimal smoothing parameters; using transformation method 1 (M1) (M1$_{MV}$: M1 with minimal-variability transformation, M$_K$: M1 with transformation by [22] using $\gamma = -1$). n is the sample size, $\zeta = s.d.(\varepsilon)$ and PC is the percentage of censoring.

				P-SPLINE		SBF	
n	ζ	PC	F	M1$_{MV}$	M1$_K$	M1$_{MV}$	M1$_K$
400	1	10	β_0	0.0265	0.0490	0.0316	0.0529
			β_1	0.2433	0.5631	0.2693	0.5855
			β_2	0.1581	0.3923	0.1559	0.3491
			m	0.0539	0.1249	0.0583	0.1217
		30	β_0	0.0374	0.1127	0.0424	0.0990
			β_1	0.3268	1.0173	0.3599	1.0072
			β_2	0.2238	0.7260	0.2159	0.6332
			m	0.0755	0.2385	0.0794	0.2152
		50	β_0	0.0592	0.1954	0.0608	0.1706
			β_1	0.4839	1.4747	0.5142	1.3873
			β_2	0.3342	1.2063	0.3127	1.0467
			m	0.1145	0.3803	0.1145	0.3332
	1.5	10	β_0	0.0387	0.0700	0.0436	0.0693
			β_1	0.3360	0.6946	0.3606	0.7205
			β_2	0.2234	0.5107	0.2066	0.4508
			m	0.0762	0.1612	0.0787	0.1530
		30	β_0	0.0500	0.1459	0.0539	0.1292
			β_1	0.4177	1.2250	0.4506	1.1785
			β_2	0.2851	0.9122	0.2627	0.7905
			m	0.0975	0.2958	0.0985	0.2627
		50	β_0	0.0735	0.2421	0.0721	0.2059
			β_1	0.5633	1.7462	0.5967	1.5620
			β_2	0.3960	1.4647	0.3561	1.2511
			m	0.1364	0.4618	0.1319	0.3909

Table 7.3: Simulation Model 2: average relative error for the estimates of the functions (F) β_1, β_2 and m obtained with the P-spline estimator and the smooth-backfitting estimator (SBF) with optimal smoothing parameters; using transformation methods 1 (M1) and 2 (M2). ($M1_{CV}$: M1 with cross-validation transformation, $M1_{MV}$: M1 with minimal-variability transformation, M_K: M1 with transformation by [22] using $\gamma = -1$. M_Z indicates the estimator when no transformation is applied to the observed response (Z, Δ). n is the sample size, PC is the percentage of censoring.

			P-SPLINE					SBF	
n	PC	F	$M1_{CV}$	$M1_{MV}$	$M1_K$	M2	M_Z	$M1_{MV}$	$M1_K$
250	10	β_1	0.0514	0.0519	0.1718	0.0424	0.0778	0.0806	0.2127
		β_2	0.0674	0.0680	0.2218	0.0564	0.0876	0.0983	0.2659
		m	0.0260	0.0262	0.0740	0.0229	0.0474	0.0355	0.0840
	30	β_1	0.0856	0.0868	0.3522	0.0546	0.1704	0.1164	0.4003
		β_2	0.1121	0.1131	0.4621	0.0730	0.1580	0.1430	0.4853
		m	0.0414	0.0419	0.1637	0.0303	0.1135	0.0521	0.1675
	50	β_1	0.1245	0.1322	0.7222	0.0862	0.2741	0.1828	0.7120
		β_2	0.1612	0.1716	0.9379	0.1137	0.2266	0.2029	0.8216
		m	0.0684	0.0742	0.3867	0.0571	0.1820	0.0830	0.3409
500	10	β_1	0.0367	0.0366	0.1230	0.0301	0.0641	0.0653	0.1549
		β_2	0.0482	0.0481	0.1604	0.0394	0.0724	0.0783	0.1961
		m	0.0188	0.0190	0.0557	0.0157	0.0433	0.0274	0.0679
	30	β_1	0.0608	0.0605	0.2713	0.0361	0.1612	0.0914	0.3172
		β_2	0.0800	0.0796	0.3598	0.0484	0.1493	0.1092	0.3925
		m	0.0301	0.0300	0.1276	0.0207	0.1106	0.0411	0.1412
	50	β_1	0.0997	0.1070	0.6291	0.0667	0.2682	0.1552	0.6224
		β_2	0.1365	0.1460	0.7931	0.0923	0.2240	0.1722	0.6943
		m	0.0539	0.0620	0.3652	0.0465	0.1812	0.0716	0.3146

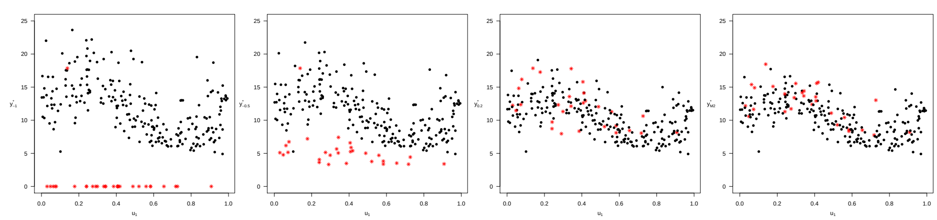
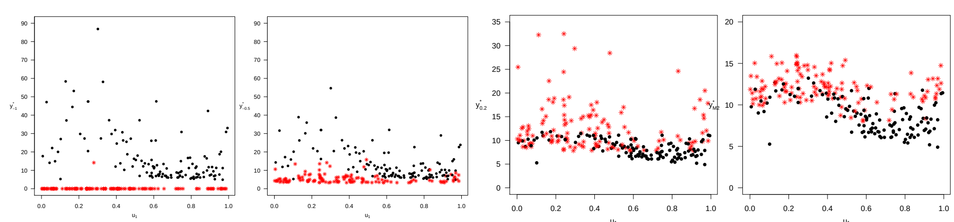
average relative error based on true (unobserved) responses with the P-spline estimate for: $n = 250$: β_1: 0.0400; β_2: 0.0534, m: 0.0217 and $n = 500$: β_1: 0.0375; β_2: 0.0287, m: 0.0145

clinics for maintenance. Methadone is a drug similar to heroin which prevents or reduces withdrawal symptoms when a patient stays off heroin. Patients detoxifying from methadone

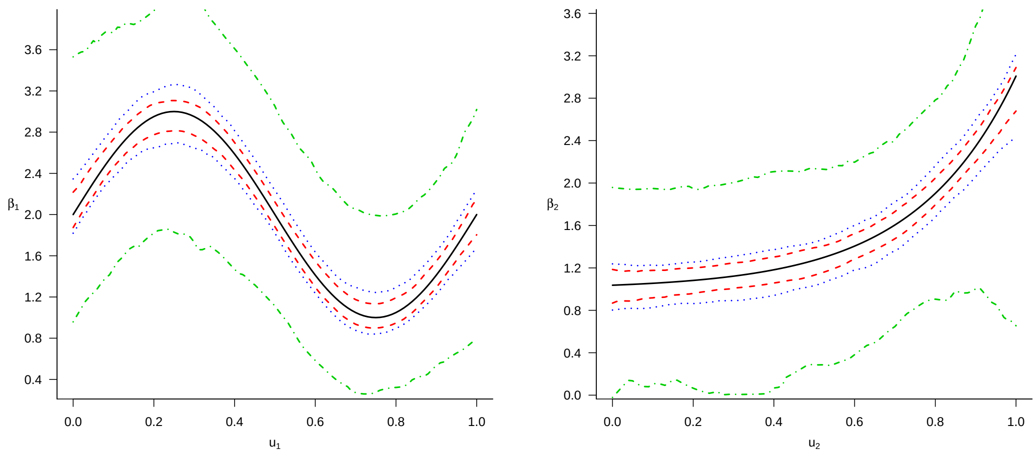
Table 7.4: Simulation Model 2: average ratio of RE($\hat{m}$) based on $\boldsymbol{\lambda}_{CV}$ and $\boldsymbol{\lambda}_{opt}$ for the P-spline estimates using transformation methods 1 (M1) and 2 (M2). (M1_{CV}: M1 with cross-validation transformation, M1_{MV}: M1 with minimal-variability transformation, M$_K$: M1 with transformation by [22] using $\gamma = -1$. n is the sample size, PC is the percentage of censoring.

	$n = 250$				$n = 500$			
PC	M1_{CV}	M1_{MV}	M1_K	M2	M1_{CV}	M1_{MV}	M1_K	M2
10	1.2126	1.2152	1.3705	1.1529	1.2028	1.1959	1.3234	1.1462
30	1.2349	1.2281	1.4376	1.1342	1.2563	1.2560	1.3614	1.1620
50	1.1688	1.1589	1.2220	1.0875	1.1686	1.1354	1.1328	1.0705

Table 7.5: Simulation Model 2: Responses transformed with method 1 (M1) for different choices of γ and method 2 (M2) for PC = 10% and PC = 50% for $n = 250$. Uncensored observations are indicated by black dots, censored observations are indicated by red asterisks.

PC	$M1(\gamma = -1)$	$M1(\gamma = -0.5)$	$M1(\gamma = 0.2)$	$M2$
10%				
50%				

maintenance soon return to illicit opiate abuse, and methadone is only beneficial to addicts in treatment. The main objective of the study was to investigate the effectiveness of treatment programs based on the time an addict spends in a clinic, the larger this duration time the more effective the therapy is. The response is the duration time T, in days) of heroin addicts from entry to a clinic until departure or end of study period; 150 out of the 238 patients left the clinic during the study period, the remaining 88 patients still in the clinic at the end of



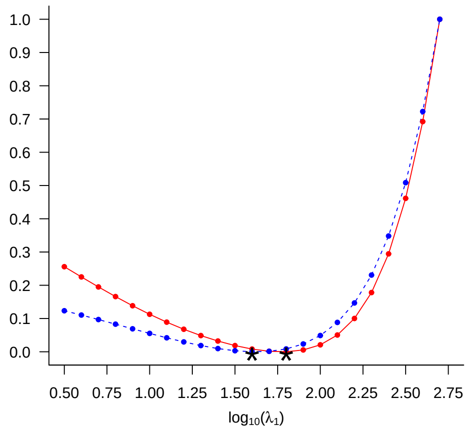
(a) (b)

Figure 7.1: Simulation Model 2: Pointwise confidence band for the P-spline estimates of (a) β_1 and (b) β_2 obtained with method 1 ($\gamma = -1$) (green, dashed dotted), method 1 and γ_{MV} (blue, dotted) and method 2 (red, dashed) for $n = 500$ and PC = 30%.

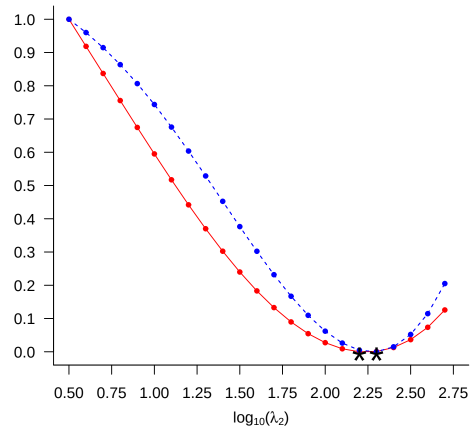
the study period are censored cases. We focus on the effect of clinic (C, 1= clinic 1, 0 = clinic 2) and a history of imprisonment (P, 1= yes, 0= no) on the time remaining on methadone treatment in a VCM where the coefficients vary with the maximum methadone dosage (M, in mg/day), i.e.,

$$E(T|M, C, P) = \beta_1(M) + \beta_2(M) \times C + \beta_3(M) \times P.$$

We present results for a homoscedastic model based on method 2 only, since method 2 outperformed method 1 in our simulation study and since similar results were obtained with a hetereoscedastic model. We smooth the coefficients by P-splines of degree 3 on 15 equidistant knots with a second order difference penalty. The initial estimate for the regression coefficients is obtained using the first method and an MV-transformation parameter ($\gamma_{MV} = -0.2$). The smoothing parameters ($\lambda_1 = 50, \lambda_2 = 250$ and $\lambda_3 = 100$) were selected by cross-validation on a logarithmic scale. Figure 8.3 presents the resulting estimated mean survival time obtained with transformation method 2 . Only in the second clinic doses above 80 mg/day were given



(a)



(b)

Figure 7.2: Simulation Model 2: (a) CV (red,solid) and relative regression error (blue, dashed) curves for $\lambda_1 \in 10^{\{0.5,0.6,\ldots,2.6\}}$ and for λ_2 minimizing CV resp. relative error. (b) CV (red, solid) and relative regression error (blue, dashed) curves for $\lambda_2 \in 10^{\{0.5,0.6,\ldots,2.6\}}$ and for λ_1 minimizing CV resp. relative error. The black asterisk indicates the minimal value. The curves are based on one simulated data set of size $n = 500$ and PC =30% using method 1 with MV-transformation parameter.

to the patients, however our model reveals that these doses no longer result in larger duration times. This finding could not be obtained if a linear term was considered for the methadone effect. For small methadone doses, the estimated mean survival time is similar for all patients but when the dosage increases, the second clinic tends to do a better job in retaining its patients under treatment. Figure 8.3 also shows that the length of time in treatment is shorter for patients with a history of imprisonment.

9 Summary and further research

We propose a P-spline smoothing technique for the estimation of the varying coefficients in a VCM with responses that are subject to right censoring. Using the mean-preserving principle we transform the original observations into 'synthetic' observations, which are then

Table 7.6: Simulation Model 3: average relative error for the estimates of the functions (F) β_0, β_1, m and σ^2 obtained with the P-spline estimator using method 1 with minimal-variability transformation (M1) and method 2 (M2). n is the sample size, PC is the percentage of censoring.

n :			250			500		
PC:			10%	30%	50%	10%	30%	50%
α	M	F						
1	M1	β_0	0.1984	0.3393	0.5431	0.1392	0.2731	0.3891
		β_1	0.1371	0.2502	0.4077	0.0995	0.2021	0.2868
		m	0.0411	0.0781	0.1328	0.0306	0.0614	0.0956
	M2	β_0	0.1494	0.2294	0.3778	0.1024	0.1780	0.2735
		β_1	0.0983	0.1769	0.3048	0.0694	0.1399	0.2178
		m	0.0309	0.0580	0.1063	0.0228	0.0456	0.0775
2	M1	β_0	0.2274	0.3647	0.5532	0.1624	0.2787	0.4053
		β_1	0.1538	0.2682	0.4139	0.1127	0.2045	0.3011
		m	0.0468	0.0847	0.1371	0.0355	0.0642	0.1012
	M2	β_0	0.1899	0.2591	0.3880	0.1336	0.1939	0.2811
		β_1	0.1224	0.1954	0.3096	0.0874	0.1463	0.2255
		m	0.0383	0.0655	0.1109	0.0290	0.0491	0.0819
1	M1	σ^2	0.2006	0.3398	0.7071	0.1594	0.2896	0.4480
2	M1	σ^2	0.2158	0.3186	0.4803	0.1674	0.2623	0.3658

used for the P-spline estimation. We emphasize the benefit of a data-driven data transformation when the transformation formula is independent of the underlying VCM. Better results are obtained with data transformations that take the true VCM into account. The latter transformation formulas require prior knowledge of the VCM which is obtained from the model-independent transformation methods. We give asymptotic support for the behavior of our proposed P-spline estimators and prove the consistency and asymptotic normality of our P-spline estimators. Simulation studies compare its finite sample behavior with that of the SBF-estimator proposed by [37] and illustrate good finite sample performance of our proposed P-spline estimates and moreover suggest improvements for the method proposed in [37]. We

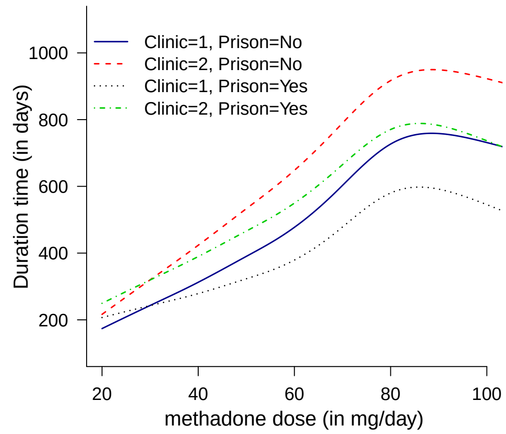


Figure 8.3: Addict data. Fitted P-spline regression function with method 2 using method 1 with $\gamma_{MV} = -0.2$ and $\lambda_{0,CV} = 50, \lambda_{1,CV} = 250, \lambda_{2,CV} = 100$.

conclude for the model-independent transformation methods that the combination of either P-spline smoothing or SBF-smoothing with a data-driven transformation parameter are both good approaches for estimating the coefficient functions in a VCM.

Our simulation results show that the complexity of the simulation scenario (number of explanatory variables, variance of the error terms, shape of the coefficient functions,...) influences the performance of and the comparison between the P-spline and SBF-estimates. Further research is needed to explore which estimation approach (P-spline or SBF) works best in particular scenarios and to investigate how the choice between the different methods presented in this paper depends on factors such as the number of covariates.

For data subject to right censoring, the synthetic data approach for heteroscedastic models is considered in e.g. [15], [18] and [37]. None of these authors consider variance-based weighting in the estimation of the mean regression curve. Although it is common practice to use weighted least squares when heterogeneity is present in the data (e.g. for non-censored data [31] use reweighting for heteroscedastic VCMs), [1] shows a good performance of P-spline estimators in VCMs for non-censored data even if the heteroscedasticity is ignored in the estimation process. How to bring in variance-based reweighting in the estimation process and studying the impact of reweighting on the quality of the P-spline estimators in heteroscedastic VCMs are challenging open problems.

Finally note that, for interval censored observations, the construction of synthetic data has been considered in a few papers, e.g. [39] proposed a mean preserving transformation and [29] studied a Buckley-James type estimator in the classical linear regression context. Finding appropriate transformations for interval censored data in VCMs is an unexplored interesting open question.

10 Appendix

10.1 Definitions and Properties

This section contains the Definition of the L_2-distance and the Assumptions needed for the main results, i.e., Theorem 1 and 2.

Assumption A.

1. For all $p = 1, \dots, d$, the random variable U_p has distribution function F_{U_p} on $\mathcal{U}_p = [a_p, b_p]$. The distribution function F_{U_p} has Lebesgue density f_{U_p} which is bounded away from zero and infinity, uniformly in $\mathcal{U}_p$, i.e. there exist positive constants N_1 and N_2 such that $N_1 \leq f_{U_p}(u) \leq N_2$ for $u \in \mathcal{U}_p$.
2. The eigenvalues $\eta_1(\mathbf{u}), \dots, \eta_d(\mathbf{u})$ of $\Sigma(\mathbf{u}) = E(\mathbf{X}\mathbf{X}'|\mathbf{U} = \mathbf{u})$ are bounded away from zero and infinity, uniformly over all $\mathbf{u} \in \mathcal{U}^d$, i.e. there exist positive constants N_3 and N_4 such that $N_3 \leq \eta_1(\mathbf{u}) \leq \cdots \leq \eta_d(\mathbf{u}) \leq N_4$ for $\mathbf{u} \in \mathcal{U}^d$.
3. There exists a positive constant N_5 such that $\mid X_p \mid \leq N_5$ for $p = 1, \dots, d$.
4. There exists a positive constant N_6 such that $\sigma_j^2(\mathbf{u}, \mathbf{x}) \leq N_6 < \infty$ for $j = 1, 2$ and for every $\mathbf{u} \in \mathcal{U}^d, \mathbf{x} \in \mathbb{R}^d$, where $\sigma_j^2(\mathbf{u}, \mathbf{x}) = \text{Var}(Y_j^* \mid \mathbf{U} = \mathbf{u}, \mathbf{X} = \mathbf{x})$.
5. $\limsup_n \left(\frac{\max_p m_p}{\min_p m_p}\right) < \infty$.
6. $n^{-1} m_{\max}^{3/2} \lambda_{\max} \to 0$ and $n^{-1} m_{\max} \to 0$ as $n \to \infty$.
7. $n^{-1} m_{\max} \log(m_{\max}) \to 0$ as $n \to \infty$.
8. $\rho_n \to 0$ as $n \to \infty$.

Assumption B.

1. $\sup_{\mathbf{u},\mathbf{x}} \tau_1(\mathbf{u},\mathbf{x}) \sup_{t\leq\tau_1(\mathbf{u},\mathbf{x})} |\hat{G}(t|\mathbf{u},\mathbf{x}) - G(t|\mathbf{u},\mathbf{x})| = o_p(1)$.
2. $\sup_{\mathbf{u},\mathbf{x}} \kappa(\mathbf{u},\mathbf{x}) \to 0$ as $n \to \infty$.
3. $\sup_{\mathbf{u},\mathbf{x}} \kappa_\sigma(\mathbf{u},\mathbf{x}) \to 0$ as $n \to \infty$.

Assumption C.

1. $\beta_p \in C^3([a_p, b_p])$, for each $p = 1, \ldots, d$, where $C^r([a, b])$ is the space of $r-$times continuously differentiable functions on $[a, b]$.
2. $m_{\max}^{3/2}\Big[\sup_{\mathbf{u},\mathbf{x}}\big\{\tau_1(\mathbf{u},\mathbf{x}) \sup_{t\leq\tau_1(\mathbf{u},\mathbf{x})} |\hat{G}(t|\mathbf{u},\mathbf{x}) - G(t|\mathbf{u},\mathbf{x})| + \kappa(\mathbf{u},\mathbf{x})$
   $+ \kappa_\sigma(\mathbf{u},\mathbf{x})\big\} + \rho_n\Big] \to 0$ and $n^{-1/2}m_{\max}^2 \to 0$; $n^{-1}m_{\max}^{3/2}\lambda_{\max} \to 0$
   as $n \to \infty$.

Assumption D.

1. $m_{\max}^{-1/2}n^{1/2}\Big(\sup_{\mathbf{u},\mathbf{x}}\big(\tau_1(\mathbf{u},\mathbf{x}) \sup_{t\leq\tau_1(\mathbf{u},\mathbf{x})} |\hat{G}(t|\mathbf{u},\mathbf{x}) - G(t|\mathbf{u},\mathbf{x})| +$
   $\kappa(\mathbf{u},\mathbf{x})\big)\Big) \to 0$ and $n^{-1/2}m_{\max}\lambda_{\max} + n^{1/2}\rho_n \to 0$ as $n \to \infty$.
2. $m_{\max}^{-1}n^{1/2} \sup_{\mathbf{u},\mathbf{x}}\big(\tau_1(\mathbf{u},\mathbf{x}) \sup_{t\leq\tau_1(\mathbf{u},\mathbf{x})} |\hat{G}(t|\mathbf{u},\mathbf{x}) - G(t|\mathbf{u},\mathbf{x})| + \kappa(\mathbf{u},\mathbf{x})$
   $+ \quad \kappa_\sigma(\mathbf{u},\mathbf{x})\big) \to 0$ and $m_{\max}^{-1/2} \log n + n^{-1/2}m_{\max}\lambda_{\max} + n^{1/2}\rho_n \to 0$
   as $n \to \infty$.

Assumption A.1 guarantees that the observation points are randomly scattered and is a natural assumption in non-parametric regression (see e.g. [13]). All A assumptions are common in P-spline theory (see e.g. [1]). In particular A.1-A.4 are common in mean regression in varying coefficient models. Also note that Assumptions A.5, A.6 andA.7 are satisfied with the choice of number of knots and smoothing parameter of Remark 4. When all β_p have bounded r-th derivatives $\rho_n = O_p(m_{\max}^{-r})$ ([30]). Assumption B ensures that the censored nature of the data is taken into account and is illustrated by an example in Remark 3. When the Kaplan-Meier estimator is used to estimate G, it follows from [40] that $\sup_{t\leq\tau_1(\mathbf{u},\mathbf{x})} |\hat{G}(t)-G(t)| = O_p(n^{-1/2})$. Assumption C guarantees that, uniformly over $\mathcal{U}_p$, the second order derivative of $\hat{\beta}_{1p}$ is a consistent estimator for β_{1p}, for $p = 1, \ldots, d$. It is a technical assumption needed in the proof of Theorem 1, Part 2 and guarantees that the Kaplan-Meier estimator based on residual observations constructed with method 1 converges to the true error distribution F. Assumption D is an assumption on the convergence rate of the P-spline estimators and guarantees that the squared L_∞-distance between the P-spline estimators $\hat{\boldsymbol{\beta}}_j$ and $\boldsymbol{\beta}_j^*$ converges to zero at a faster rate than the variance given by (5.1). For the examples considered in Remark 3, Assumptions C and D are also fulfilled when G is estimated using the Kaplan-Meier estimator, $m_{\max} \asymp n^{1/5}$ and $\lambda_{\max} \asymp n^{\iota}, \iota < 3/10$.

Definition 2. *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_2-norm is given by:*

$$\|f\|_{L_2} = \left\{ \int_{\mathcal{U}} f^2(t)dt \right\}^{1/2}, \quad \|\mathbf{g}\|_{L_2} = \left(\sum_{p=1}^{d} \|g_p\|_{L_2}^2 \right)^{1/2},$$

Definition 3. *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 ζ_{max} is the largest eigenvalue of $\mathbf{A}'\mathbf{A}$.*

Definition 4. *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 5. *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 [10] or [30].

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

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

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

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* $g \in \mathbb{G}_p, p = 1, \ldots, d$ *where* $\mathbb{G}_p$ *is the space of spline functions of degree* q_p *on* $\mathcal{U}_p$ *with knots* $\boldsymbol{\xi}_p$.

We use as notations $\hat{\boldsymbol{\alpha}}_j, \boldsymbol{\alpha}_j^*$ and $\tilde{\boldsymbol{\alpha}}_j$ for methods $j = 1, 2$, 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 $\mathbf{M} = (M_{j1}, \ldots, M_{jn})'$ 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 $\tilde{\boldsymbol{\beta}}_j = (\tilde{\beta}_{j1}, \ldots, \tilde{\beta}_{jd})'$.

10.2 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 $\hat{Y}_{1i}^*$ responses, derived in Lemma 1.

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

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

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

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

we consider two cases and prove the following results,

$$\max_{1 \le i \le n} \left\{ \mid \hat{Y}_{1i}^* - Y_{1i}^* \mid 1_{\{Z_i \le \tau_1(\mathbf{U}_i, \mathbf{X}_i)\}} \right\} \lesssim \sup_{\mathbf{u},\mathbf{x}} \left(\tau_1(\mathbf{u}, \mathbf{x}) \sup_{t \le \tau_1(\mathbf{u},\mathbf{x})} |\hat{G}(t|\mathbf{u}, \mathbf{x}) - G(t|\mathbf{u}, \mathbf{x})| \right). \quad (10.1)$$

$$\max_{1 \le i \le n} \left\{ \mid \hat{Y}_{1i}^* - Y_{1i}^* \mid 1_{\{Z_i > \tau_1(\mathbf{U}_i, \mathbf{X}_i)\}} \right\} \lesssim \sup_{\mathbf{u},\mathbf{x}} \kappa(\mathbf{u}, \mathbf{x}). \quad (10.2)$$

For (10.1) we start by the triangle inequality,

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

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 φ_1 and $\hat{\varphi}_1$ by ψ_1 and $\hat{\psi}_1$ respectively.

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

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

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,

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

For (10.2) we have,

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

By combining (10.1) and (10.2), the result of Lemma 1 follows. □

Proof of Theorem 1, Part 1. Since

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

the result follows by showing that,

$$\begin{aligned} &\|\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}^*_1\|_{L_2} \\ &\quad = O_p\left(\sup_{\mathbf{u},\mathbf{x}}\left\{\tau_1(\mathbf{u},\mathbf{x}) \sup_{t\leq\tau_1(\mathbf{u},\mathbf{x})} |\hat{G}(t|\mathbf{u},\mathbf{x}) - G(t|\mathbf{u},\mathbf{x})| + \kappa(\mathbf{u},\mathbf{x})\right\}\right) && (10.3) \\ &\|\boldsymbol{\beta}^*_1 - \tilde{\boldsymbol{\beta}}_1\|_{L_2} = O_p\left(n^{-1/2}m^{1/2}_{\max}\right) && (10.4) \\ &\|\tilde{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1\|_{L_2} = O_p\left(n^{-1}m^{3/2}_{\max}\lambda_{\max} + \rho_n\right). && (10.5) \end{aligned}$$

We start with the proof of (10.3). By Property 2 it suffices to show that

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

From [1] we have,

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

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,

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

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

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

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,

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

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

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

We continue with the proof of (10.4). Using similar arguments as is the proof of (10.3), we have

$$\begin{aligned}
&\|\boldsymbol{\alpha}^*_1 - \tilde{\boldsymbol{\alpha}}_1\|_2 \\
&\le \left\{1 + \|(\mathbf{R}'\mathbf{R})^{-1}\|_2\|\mathbf{Q}_{\boldsymbol{\lambda}}\|_2 + o_p(n^{-1}m_{\max}^{3/2}\lambda_{\max})\right\}\|\boldsymbol{\alpha}^*_{1,reg} - \tilde{\boldsymbol{\alpha}}_{1,reg}\|_2, \qquad (10.6)
\end{aligned}$$

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,

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

$$+\sum_{i\neq j} E\left\{B_{pl}(U_{ip};q_p)B_{pl}(U_{jp};q_p)(Y_{1i}^*-M_{1i})(Y_{1j}^*-M_{1j})\right\}\Bigg].$$

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,

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

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,

$$\|\boldsymbol{\alpha}_{1,reg}^*-\tilde{\boldsymbol{\alpha}}_{1,reg}\|_2^2 = O_p\left(n^{-1}m_{\max}^2\right). \tag{10.7}$$

Combining (10.6) and (10.7) 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 (10.5) follows immediately. □

10.3 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 $\hat{Y}_2^*$ responses. 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 [33]. 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 (4.4) and (4.6) 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,*

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

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

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

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

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

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

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

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,

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

By Lemma A.10 of [20], 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,

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

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 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,

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

and,

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

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,*

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

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

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

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 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 [10]),

$$\hat{\boldsymbol{\beta}}_1^{(v)} = K_p^v \, \mathbf{b}(u_p, q-v)' \mathbf{D}_v \hat{\boldsymbol{\alpha}}_{1p}, \tag{10.1}$$

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 (10.1) 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,

$$\|\hat{\boldsymbol{\beta}}_1^{(v)} - \tilde{\boldsymbol{\beta}}_1^{(v)}\|_{L_2} = O_p(m_{\max}^{v-1/2} \|\hat{\boldsymbol{\alpha}}_1 - \tilde{\boldsymbol{\alpha}}_1\|_2). \tag{10.2}$$

We now use the fact that there exists a spline function (see Corollary 6.21 and (2.120) of Theorem 2.59 in [30]) $\zeta_p(u_p) = \sum_{l=1}^{m_p} c_{pl} B_{pl}(u_p, q_p)$ 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,

$$\|\tilde{\boldsymbol{\beta}}_1^{(v)} - \boldsymbol{\zeta}^{(v)}\|_{L_2} = O_p(m_{\max}^v \rho_n + n^{-1} m_{\max}^{3/2} \lambda_{\max}). \tag{10.3}$$

To show the validity of (10.3), we proceed as follows. By Lemma A.7 of [20], 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 (10.3) follows. Also note ([30]) that ζ_p satisfies

$$\|\beta_p^{(v)} - \zeta_p^{(v)}\|_{L_\infty} = O(m_p^{v-r}). \tag{10.4}$$

The rates in (10.2)-(10.4) provide the key for the proof. Indeed

$$\|\hat{\boldsymbol{\beta}}_1^{(v)} - \boldsymbol{\beta}^{(v)}\|_{L_\infty} \leq \|\hat{\boldsymbol{\beta}}_1^{(v)} - \boldsymbol{\zeta}^{(v)}\|_{L_\infty} + \|\boldsymbol{\zeta}^{(v)} - \boldsymbol{\beta}^{(v)}\|_{L_\infty}. \tag{10.5}$$

For the second term in (10.5) we use (10.4). For the first term, note that,

$$\|\hat{\boldsymbol{\beta}}_1^{(v)} - \boldsymbol{\zeta}^{(v)}\|_{L_\infty} \lesssim m_{\max}^{-1/2} \|\hat{\boldsymbol{\beta}}_1^{(v)} - \boldsymbol{\zeta}^{(v)}\|_{L_2} \tag{10.6}$$

and that,

$$\begin{aligned} \|\hat{\boldsymbol{\beta}}_1^{(v)} - \boldsymbol{\zeta}^{(v)}\|_{L_2} &\leq \|\hat{\boldsymbol{\beta}}_1^{(v)} - \tilde{\boldsymbol{\beta}}^{(v)}\|_{L_2} + \|\tilde{\boldsymbol{\beta}}_1^{(v)} - \boldsymbol{\zeta}^{(v)}\|_{L_2} \\ &= O_p(m_{\max}^{v-1/2} \|\hat{\boldsymbol{\alpha}}_1 - \tilde{\boldsymbol{\alpha}}_1\|_2 + m_{\max}^v \rho_n + n^{-1} m_{\max}^{3/2} \lambda_{\max}). \end{aligned} \tag{10.7}$$

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 (10.2)-(10.7).

□

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,

$$\begin{aligned} \hat{Y}_{2i}^* - Y_{2i}^* &= \hat{m}_1(\mathbf{U}_i, \mathbf{X}_i) - m(\mathbf{U}_i, \mathbf{X}_i) \\ &\quad + \frac{\hat{\sigma}_1(\mathbf{U}_i, \mathbf{X}_i)}{1 - \hat{F}(\hat{E}_i^T)} \int_{\hat{E}_i^T}^{\hat{S}_i} s d\hat{F}(s) - \frac{\sigma(\mathbf{U}_i, \mathbf{X}_i)}{1 - F(E_i^T)} \int_{E_i^T}^{S_i} s dF(s) \\ &= \{\hat{m}_1(\mathbf{U}_i, \mathbf{X}_i) - m(\mathbf{U}_i, \mathbf{X}_i)\} && (10.8) \\ &\quad + \frac{\hat{\sigma}_1(\mathbf{U}_i, \mathbf{X}_i) - \sigma(\mathbf{U}_i, \mathbf{X}_i)}{1 - \hat{F}(\hat{E}_i^T)} \int_{\hat{E}_i^T}^{\hat{S}_i} s d\hat{F}(s) && (10.9) \end{aligned}$$

$$+\frac{\sigma(\mathbf{U}_i,\mathbf{X}_i)\{\hat{F}(\hat{E}_i^T)-F(E_i^T)\}}{\{1-\hat{F}(\hat{E}_i^T)\}\{1-F(E_i^T)\}}\int_{\hat{E}_i^T}^{\hat{S}_i} sd\hat{F}(s) \tag{10.10}$$

$$+\frac{\sigma(\mathbf{U}_i,\mathbf{X}_i)}{1-F(E_i^T)}\left\{\int_{\hat{E}_i^T}^{E_i^T} sd\hat{F}(s)+\int_{E_i^T}^{S_i} sd(\hat{F}(s)-F(s))+\int_{S_i}^{\hat{S}_i} sd\hat{F}(s)\right\}. \tag{10.11}$$

We first consider the three integrals in (10.11). Using integration by part, we have,

$$\begin{aligned}
&\int_{\hat{E}_i^T}^{E_i^T} sd\hat{F}(s) = E_i^T\hat{F}(E_i^T)-\hat{E}_i^T\hat{F}(\hat{E}_i^T)-\int_{\hat{E}_i^T}^{E_i^T}\hat{F}(s)ds \\
&= E_i^T\{\hat{F}(E_i^T)-F(E_i^T)\}+\{E_i^TF(E_i^T)-\hat{E}_i^TF(E_i^T)\}+\hat{E}_i^T\{F(E_i^T)-\hat{F}(\hat{E}_i^T)\} \\
&\qquad -\int_{\hat{E}_i^T}^{E_i^T}\hat{F}(s)ds.
\end{aligned} \tag{10.12}$$

For the first term of (10.12), 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 $|E_i^T| \le \{\sigma(\mathbf{U}_i,\mathbf{X}_i)\}^{-1}\{|\min(Z_i,\tau_2(\mathbf{U}_i,\mathbf{X}_i))|+|m(\mathbf{U}_i,\mathbf{X}_i)|\}<\infty$. To get a consistency rate for the second and the fourth term of (10.12), 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 (10.12). For the third term of

(10.12), 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,

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

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

$$F(\hat{E}_i^T) - F(E_i^T) = O_p(a_n). \tag{10.13}$$

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 (10.12) we conclude for the first term of (10.11),

$$\frac{\sigma(\mathbf{U}_i, \mathbf{X}_i)}{1 - F(E_i^T)} \int_{\hat{E}_i^T}^{E_i^T} s d\hat{F}(s) = O_p(a_n). \tag{10.14}$$

In a similar way, we obtain for the third term of (10.11)

$$\frac{\sigma(\mathbf{U}_i, \mathbf{X}_i)}{1 - F(E_i^T)} \int_{\hat{S}_i^T}^{S_i^T} s d\hat{F}(s) = O_p(a_n). \tag{10.15}$$

For the second integral in (10.11), we use partial integration and Lemma 4 to obtain

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

The terms (10.8)-(10.10) are more easy to handle. For (10.8) we use Lemma 3(a). For (10.9) and (10.10) we need that

$$\int_{\hat{E}_i^T}^{\hat{S}_i} sd\hat{F}(s) = O_p(1). \tag{10.16}$$

To show (10.16), note that, using similar reasoning as in [18], 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 (10.14) and (10.15) 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 $F(E_i^T) < 1$ (10.13), we get that (10.9) and (10.10) are both $O_p(a_n)$. □

10.4 Proof of Theorem 2

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 $p = 1, \ldots, d$,

$$\left\{s.e.\left(\beta_{jp}^*(u_p) \mid \mathcal{X}_n\right)\right\}^{-1} \left\{\beta_{jp}^*(u_p) - \tilde{\beta}_{jp}(u_p)\right\} \xrightarrow{d} \mathrm{N}(0,1) \tag{10.1}$$

$$\left\{s.e.\left(\beta^*_{jp}(u_p) \mid \mathcal{X}_n\right)\right\}^{-1}\left\{(\hat{\beta}_{1p}(u_p) - \beta^*_{1p}(u_p)) + (\tilde{\beta}_{1p}(u_p) - \beta_p(u_p))\right\} \xrightarrow{p} 0. \tag{10.2}$$

The proof of (10.1) 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 $d_i^2 = \sigma^2_{1,i}\{\mathbf{B}'_p(\mathbf{u})(\mathbf{R}'\mathbf{R} + \mathbf{Q}_{\boldsymbol{\lambda}})^{-1}\mathbf{R}_i\}^2$ and $\xi_i = \sigma^{-2}_{1,i}(Y^*_{1i} - M_{1i})$. 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 that the Lindeberg condition,

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

is satisfied, 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,

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

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

$$\boldsymbol{\omega}'\mathbf{R}_i\mathbf{R}'_i\boldsymbol{\omega} \lesssim \sum_{p=0}^{d}\|g_{\boldsymbol{\omega},p}\|^2_\infty \lesssim m_{\max}\sum_{p=0}^{d}\|g_{\boldsymbol{\omega},p}\|^2_{L_2} \asymp \|\boldsymbol{\omega}\|_2^2. \tag{10.3}$$

From Lemmas A.1 and A.2 in [20], 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,

$$\begin{aligned}\boldsymbol{\omega}'\sum_{i=1}^{n}\left\{\mathbf{R}_i\mathbf{R}_i'\sigma_{1,i}^2\right\}\boldsymbol{\omega} &\geq n\min_{1\leq i\leq n}\sigma_{1,i}^2 n^{-1}\sum_{i=1}^{n}\Big(\sum_{p=0}^{d} X_{ip} g_{\boldsymbol{\omega},p}(U_{ip};q_p)\Big)^2 \\ &\gtrsim m_{\max}^{-1} n\|\boldsymbol{\omega}\|_2^2. \end{aligned} \tag{10.4}$$

Combining (10.3) and (10.4), 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 assumption is fulfilled and hence the normality result in (10.1) follows.

We continue with the proof of (10.2). Since we assume that $\sigma_{1,i}^2$ is bounded away from zero and ∞, we have,

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

where we use the Cauchy-Schwarz inequality,

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

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

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

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,

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

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 exactly the same but we do not look at the difference $\tilde{\beta}_{2p} - \beta_p$. □

Acknowledgements

The research of the first author was supported by the Research Foundation Flanders (FWO) [grant number 11W7315N]. Support from the IAP Research Network P7/06 of the Belgian State (Belgian Science Policy) is gratefully acknowledged. For the simulations we used the infrastructure of the VSC - Flemish Supercomputer Center, funded by the Hercules Foundation and the Flemish Government - department EWI.

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(3):638–661.

[2] Beran, R. (1981). Nonparametric regression with randomly censored survival data. *Technical report, University of California, Berkeley.*

[3] Bravo, F. (2014). Varying coefficients partially linear models with randomly censored data. *Annals of the Institute of Statistical Mathematics*, 66(2):383–412.

[4] Buckley, J. and James, I. (1979). Linear regression with censored data. *Biometrika*, 66:429–436.

[5] Cai, J., Fan, J., Zhou, H., and Zhou, Y. (2007). Hazard models with varying coefficients for multivariate failure time data. *The Annals of Statistics*, pages 324–354.

[6] Cai, Z., Fan, J., and Li, R. (1999). Generalized varying-coefficient models. *Department of Statistics, UCLA.*

[7] Cai, Z., Fan, J., and Yao, Q. (2000). Functional-coefficient regression models for nonlinear time series. *Journal of the American Statistical Association*, 95(451):941–956.

[8] Caplehorn, J. and Bell, J. (1991). Methadone dosage and retention of patients in maintenance treatment. *Medical Journal of Australia*, 154(3):195–199.

[9] Cox, D. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society, Series B*, 34(2):187–220.

[10] De Boor, C. (1978). A practical guide to splines. *Springer, New York*.

[11] Eilers, P. H. and Marx, B. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science*, 11(2):89–102.

[12] El Gouch, A. and Van Keilegom, I. (2008). Nonparametric regression with dependent censored data. *Scandinavian Journal of Statistics*, 35:228–247.

[13] Eubank, R. and Speckman, P. (1990). Curve fitting by polynomial trigonometric regression. *Biometrika*, 77:1–9.

[14] Fahrmeir, L. and Tutz, G. (2001). Multivariate statistical modelling based on generalized linear models, second edition. *Springer*.

[15] Fan, J. and Gijbels, I. (1994). Censored regression: local linear approximations and their applications. *Journal of the American Statistical Association*, 89(426):560–570.

[16] Fan, J. and Zhang, W. (2008). Statistical methods with varying coefficient models. *Statistics and its interface*, 1:179–195.

[17] Hastie, T. and Tibshirani, R. (1993). Varying coefficient models. *Journal of the Royal Statistical Society, series B*, 55:757–796.

[18] Heuchenne, C. and Van Keilegom, I. (2007). Polynomial regression with censored data based on preliminary nonparametric estimation. *Ann. Inst. Statist. Math.*, 59(2):273–297.

[19] Hoover, D., Rice, J., Wu, C., and Yang, L. (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika*, 85(4):809–822.

[20] 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.

[21] Jin, Z., Lin, Y., and Ying, Z. (2006). On least-squares regression with censored data. *Biometrika*, 93(1):147–161.

[22] Koul, H., Susarla, V., and Van Ryzin, J. (1981). Regression analysis with randomly right-censored data. *The Annals of Statistics*, 9:1276–1288.

[23] Lagakos, S., Barraj, L. M., and Degruttola, V. (1988). Nonparametric analysis of truncated survival data, with application to aids. *Biometrika*, 75(3):515–523.

[24] Lee, Y., Mammen, E., and Park, B. (2012). Flexible generalized varying coefficient regression models. *The Annals of Statistics*, 40(3):1906–1933.

[25] Leurgans, S. (1995). Linear models, random censoring and synthetic data. *Biometrika*, 74:301–309.

[26] Ma, Y. and Wei, Y. (2012). Analysis on censored quantile residual life model via spline smoothing. *Statistica Sinica*, 22(1):47.

[27] Meeker, W. (1987). Limited failure population life tests: application to integrated circuit reliability. *Technometrics*, 29(1):51–65.

[28] Nahman, N. S., Middendorf, D. F., Bay, W. H., McElligott, R., Powell, S., and Anderson, J. (1992). Modification of the percutaneous approach to peritoneal dialysis catheter placement under peritoneoscopic visualization: clinical results in 78 patients. *Journal of the American Society of Nephrology*, 3(1):103–107.

[29] Rabinowitz, D., Tsiatis, A., and Aragon, J. (1995). Regression with interval-censored data. *Biometrika*, 82(3):501–513.

[30] Schumaker, L. (2007). Spline functions: Basic theory, 3th edition. *Cambridge University Press, New York*.

[31] Shen, S., Cui, J., and Wang, C. (2015). Re-weighting estimation of the coefficients in the varying coefficient model with heteroscedastic errors. *Journal of Statistical Computation and Simulation*, pages 1–15.

[32] Stone, C. (1992). Optimal global rates of convergence for nonparametric regression. *The Annals of Statistics*, 10(4):1040–1053.

[33] Van Keilegom, I. and Akritas, M. (1999). Transfer of tail information in censored regression models. *THe Annals of Statistics*, pages 1745–1784.

[34] Wang, S., Nan, B., Zhu, J., and Beer, D. (2008). Doubly penalized Buckley-James method for survival data with high-dimension cavariates. *Biometrics*, 64:132–140.

[35] Wei, L. J. (1992). The accelerated failure time model: a useful alternative to the cox regression model in survival analysis. *Statistics in medicine*, 11 14-15:1871–9.

[36] Wu, C. O., Chiang, C., and Hoover, D. (1998). Asymptotic confidence regions for kernel smoothing of a varying-coefficient model with longitudinal data. *Journal of the American statistical Association*, 93(444):1388–1402.

[37] Yang, S., El Gouch, A., and Van Keilegom, I. (2014). Varying coefficient models having different smoothing variables with randomly censored data. *Electronic Journal of Statistics*, 8:226–252.

[38] Zheng, Z. (1987). A class of estimators of the parameters in linear regression with censored data. *Acta mathematicae applicatae sinica*, 3(3):231–241.

[39] Zheng, Z. (2008). A class of estimators of the mean survival time from interval censored data with application to linear regression. *Applied Mathematics-A Journal of Chinese Universities*, 23(4):377–390.

[40] Zhou, M. (1991). Some properties of the kaplan-meier estimator for independent non-identically distributed random variables. *The Annals of Statistics*, 19(4):2266–2274.