I N S T I T U T D E S T A T I S T I Q U E

B I O S T A T I S T I Q U E E T

S C I E N C E S A C T U A R I E L L E S

(I S B A)

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

D I S C U S S I O N
P A P E R

2013/34

Varying coefficient models having different smoothing variables with randomly censored data

YANG, S., EL GHOUCH, A. and I. VAN KEILEGOM

Varying coefficient models having different smoothing variables with randomly censored data

Seong J. Yang, Anouar El Ghouch and Ingrid Van Keilegom[1]

Université catholique de Louvain

Abstract

The varying coefficient model is a useful alternative to the classical linear model, since the former model is much richer and more flexible than the latter. We propose estimators of the coefficient functions for the varying coefficient model in the case where different coefficient functions depend on different covariates and the response is subject to random right censoring. Since our model has an additive structure and requires multivariate smoothing we employ a smooth backfitting technique, that is known to be an effective way to avoid "the curse of dimensionality" in structured nonparametric models. The estimators are based on synthetic data obtained by an unbiased transformation. The asymptotic normality of the estimators is established and a simulation study illustrates the reliability of our estimators.

Keywords

Smooth backfitting, unbiased transformation, random right censoring, local polynomial smoothing, bandwidth parameter, curse of dimensionality

AMS 2000 subject classifications:

62G08; 62N01

[1]Institute of Statistics, Biostatistics and Actuarial Science, Université catholique de Louvain, Voie du Roman Pays 20, 1348 Louvain-la-Neuve, Belgium. Email addresses: `seong.j.yang@gmail.com`, `anouar.elghough@uclouvain.be`, `ingrid.vankeilegom@uclouvain.be`

1 Introduction and model

Investigating a relation between a response and a set of covariates is a key issue in many statistical problems. Among others, mean regression models extract central trends of data by specifying the conditional mean function of a response variable given values of the covariates. A number of regression models and estimation methods have been proposed in the literature. The most traditional and simplest way to model the relation is to employ the classical linear regression model. However, this model is often too restrictive and unable to capture complicated characteristics which might exist in the data. From this point of view, the varying coefficient model is a very useful alternative. It was first proposed by Hastie and Tibshirani [1993] and takes the following form:

$$m(\mathbf{X}, \mathbf{Z}) = Z_1\alpha_1(X_1) + \cdots + Z_d\alpha_d(X_d), \tag{1}$$

where $\mathbf{X} = (X_1, \ldots, X_d)^\top$, $\mathbf{Z} = (Z_1, \ldots, Z_d)^\top$ and $m(\mathbf{x}, \mathbf{z})$ is a conditional mean function of some response given $\mathbf{X} = \mathbf{x}$ and $\mathbf{Z} = \mathbf{z}$. The α_j's are unknown coefficient functions. This model allows each coefficient function to depend on different covariates, which is not the case for many other models available in the literature. It makes the model much more flexible compared to the classical linear model, since each coefficient function is modelled nonparametrically. Moreover, by considering this model, one can incorporate nonlinear interaction effects into the model. The structure of the model is simple, since the conditional mean function is still linear in the Z_j variables. If all coefficient functions are constant functions, the model reduces to the classical linear model.

On the other hand, situations in which a response is not fully observed due to random right censoring are often encountered, for example, in medical research where patients may leave a study for various reasons. In this case, well-known regression techniques are not directly applicable since the response is only partially observed. To deal with this random right censoring, the synthetic data approach based on unbiased transformations has been studied by many authors. Koul et al. [1981] and Leurgans [1987] first proposed estimators based on different types of transformations in the classical linear model, and they were further studied by Zheng [1987], Zhou [1992], Srinivasan and Zhou [1994], Lai et al. [1995] and many others. Fan and Gijbels [1994] extended these results to nonparametric regression models and they considered a more general transformation including the transformations given in Koul et al. [1981] and Leurgans [1987] as special cases. El Ghouch and Van Keilegom [2008] further generalized the transformation and adapted the method to dependent censored data. By using a synthetic data method, one first transforms data preserving the conditional mean, and one then applies existing regression techniques as if the responses were not censored.

In this paper, for a response variable Y which is subject to random right censoring, we consider the problem of estimating the conditional mean regression function of $\phi(Y)$ given covariates for some known function ϕ. We assume that the regression function has the varying coefficient structure, that is,

$$E(\phi(Y)|\mathbf{X}, \mathbf{Z}) = \sum_{j=1}^{d} Z_j \alpha_j(X_j).$$

Note that we are estimating the conditional mean of $\phi(Y)$ rather than that of Y. In accordance with one's interest, various choices are possible for ϕ. For example, the choice $\phi(y) = I(y \leq t)$ (for fixed t) corresponds to the estimation of the conditional probability function and letting ϕ be the identity function leads to the estimation of the conditional mean of Y. For the estimation of our model, we employ a smooth backfitting (SBF) technique that is known to be an effective estimation method for structured nonparametric models. Note here that model (1) has an additive structure similar to the additive model. The SBF method was originally introduced by Mammen et al. [1999] for the additive model, and Lee et al. [2012a] studied it under the varying coefficient model. Unlike marginal integration methods, see for example Yang et al. [2006], it is known that the SBF method is free of the curse of dimensionality which usually arises when multivariate smoothing is required, since it requires only one and two dimensional smoothing. It is worthwhile to mention that model (1) has some advantages over the additive model. As pointed out before, nonlinear interaction effects can be dealt with in the former model but not in the latter model. The additive model assumes that each covariate affects the response separately. Another advantage is that the former model allows discrete variables, whereas all covariates of the latter model have to be continuous. The major hurdle of model (1) is that covariates need to pair up, which sometimes appears to be artificial. Nevertheless, even if this model is not true, it may still be used to approximate the true regression function. Recently, Lee et al. [2012b] introduced a more flexible varying coefficient model. They allow the cases where the model can contain all possible interaction effects between Z_j variables and X_j variables. In this paper, we restrict our attention to model (1) in the censored data context. We believe that our results can be extended to the model given in Lee et al. [2012b].

As mentioned before, regression models with censored data have been extensively studied, but most attention has been given to the case of univariate covariates. Recently, there have been several papers in the context of more than one covariate. Among others, Lopez [2009] and Lopez et al. [2013] considered single index modelling, which is known to be a useful dimension reduction technique. They proposed an estimator of the parameter vector in this model when random right censoring is present, and they derived their asymptotic properties. Bravo [2012] studied the partially linear varying coefficient model with random right censoring, which is an extension of Fan and Huang [2005] to the censored data context. In fact, the model studied in Bravo [2012] becomes a particular case of our model by letting $X_1 = \cdots = X_d$ in model (1) if we ignore the parametric part. Its estimation is substantially simpler than ours, since each coefficient function depends on the same univariate covariate so that only univariate smoothing techniques are required. Additive regression modelling with censored data was studied in De Uña Álvarez and Roca Pardiñas [2009] based on the backfitting algorithm proposed by Opsomer [2000]. However, theoretical properties have not been established. The purpose of this paper is to offer a very flexible model and to study its estimation with censored data when there are several covariates.

The rest of the paper is organized as follows. In Section 2 we introduce a well-known unbiased data transformation technique. Section 3 presents our main theoretical results. The proposed method based on local linear fitting is described and its asymptotic properties are established. In Section 4 we briefly show the extension of the results to local polynomial fitting. Section 5 and 6 are devoted to numerical studies. We conclude by giving some discussion in Section 7. The proofs of the theorems and lemmas are given in the Appendix at the end of the paper.

2 Transformation of data

Let $\mathbf{U} = (\mathbf{X}^\top, \mathbf{Z}^\top)^\top$, $\mathbf{X} \in [0,1]^d$, be the vector of covariates and let Y and C be a response and censoring variable, respectively. For randomly right censored data, we observe $(T_i, \delta_i, \mathbf{U}_i)$ $i = 1, \dots, n$, a random sample of $(T, \delta, \mathbf{U})$, where

$$T = Y \wedge C \text{ and } \delta = I(Y \le C),$$

and where $a \wedge b$ denotes the minimum value of a and b. Here, the problem is that the Y_i's are not fully observed due to censoring so that $E(\phi(Y)|\mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z})$ cannot be estimated in a direct way. For the unbiased transformation of the data, the following assumptions are needed:

(A1) Y and C are independent.

(A2) $P(Y \leq C|\mathbf{U}, Y) = P(Y \leq C|Y)$.

These are common assumptions made when one uses the Kaplan-Meier estimator for the censoring distribution. We consider the transformation given by Koul et al. [1981]:

$$Y^G = \frac{\delta\phi(T)}{1 - G(T-)},$$

where G is the distribution function of the censoring variable C. Under the above assumptions, we have

$$E(\phi(Y)|\mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z}) = E(Y^G|\mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z}),$$

so that the conditional mean is preserved under this transformation. The variable Y^G is observable as long as G is known. So, with Y_i^G instead of $\phi(Y_i)$ one can apply existing regression techniques for uncensored data.

We impose another assumption on the function ϕ:

(A3) Let τ be the right endpoint of the support of T, and let $I = (-\infty, \tau_0]$ for some $\tau_0 < \tau$. We assume that ϕ is bounded on I, and equals 0 outside the interval I.

This kind of truncation is common in the context of censored regression. It is necessary to deal with the lack of information in the right tail of the distribution of Y. See, for example, Lopez et al. [2013] and El Ghouch and Van Keilegom [2008].

We assume that (A1)$\sim$(A3) hold throughout the paper.

3 Estimation method with local linear fitting

We start with the (unrealistic) case where the distribution G is known. In a second step we will verify what changes when G needs to be estimated.

3.1 Smooth backfitting with censored data when G is known

In kernel regression, it is widely known that procedures based on local linear fitting have better theoretical properties than those based on local constant fitting, which suffer from boundary problems. The local linear method corrects the boundary problem. Moreover, the local constant SBF estimator does not have the oracle property, but the local linear SBF estimator does. Here, the oracle property means that the estimator of each component function has the same asymptotic distribution as if we knew all other remaining coefficient functions. This is demonstrated in Mammen et al. [1999] for the additive model and in Lee et al. [2012a] for the varying coefficient model. In this section, we introduce the local linear SBF method based on the unbiased transformation introduced in the previous section when G is known. For this, we present the estimation method along the lines of Lee et al. [2012a] and explain how one can apply the existing method to our case.

The local linear estimation technique can be applied to estimate the coefficient functions via the approximation $\alpha_j(X_{i,j}) \approx \alpha_j(x_j) + (X_{i,j} - x_j)\alpha'_j(x_j)$. Next, we consider the following least squares criterion weighted by a kernel function:

$$\int \frac{1}{n}\sum_{i=1}^{n}\left[Y_i^G - \sum_{j=1}^{d}(\alpha_j(x_j) + (X_{i,j} - x_j)\alpha'_j(x_j))Z_{i,j}\right]^2 K_{\mathbf{h}}(\mathbf{x}, \mathbf{X}_i)d\mathbf{x}, \tag{2}$$

where $Z_{i,j}$ and $X_{i,j}$ denote the jth component of $\mathbf{Z}_i$ and $\mathbf{X}_i$, $K_{\mathbf{h}}(\mathbf{u}, \mathbf{v}) = \prod_{j=1}^{d} K_{h_j}(u_j, v_j)$, $\mathbf{x} = (x_1, \ldots, x_d)^\top$ and $\mathbf{h} = (h_1, \ldots, h_d)^\top$ is a bandwidth vector. Observe that the criterion is a smoothed version of the kernel weighted local least squares criterion obtained by doing integration. This is why this method is called the "smooth" backfitting. A boundary corrected kernel is used for this estimation as in Mammen et al. [1999] and Lee et al. [2012a]. It is given by

$$K_{h_j}(u, v) = \frac{K((u-v)/h_j)}{\int K((w-v)/h_j)dw} I(u, v \in [0, 1]), \tag{3}$$

for some base kernel function K. If we let

$$\begin{aligned}\mathbf{f}(\mathbf{x}) &= (\alpha_1(x_1), \alpha'_1(x_1)h_1, \ldots, \alpha_d(x_d), \alpha'_d(x_d)h_d)^\top, \text{ and}\\ \mathbf{v}(\mathbf{X}_i, \mathbf{Z}_i; \mathbf{x}) &= (Z_{i,1}, Z_{i,1}(X_{i,1} - x_1)/h_1, \ldots, Z_{i,d}, Z_{i,d}(X_{i,d} - x_d)/h_d)^\top,\end{aligned}$$

then, (2) can be rewritten as

$$SL^G(\mathbf{f}) = \int \frac{1}{n}\sum_{i=1}^{n}\left[Y_i^G - \mathbf{v}(\mathbf{X}_i, \mathbf{Z}_i; \mathbf{x})^\top \mathbf{f}(\mathbf{x})\right]^2 K_{\mathbf{h}}(\mathbf{x}, \mathbf{X}_i)d\mathbf{x}. \tag{4}$$

When G is known, our estimator, let's say $\hat{\boldsymbol{\alpha}}^G$, is defined as the minimizer of (4) over $\mathbf{f}$, when this least squares criterion has finite values.

The SBF method can be better understood by the projection theory. So we represent our estimator in the context of the projection theory. We define some spaces of tuples of functions as follows:

$$
\begin{aligned}
L_2(\hat{\mathbf{M}}) &= \{\mathbf{f} : \mathbf{f}(\mathbf{x}) = (\mathbf{f}_1(\mathbf{x})^\top, \ldots, \mathbf{f}_d(\mathbf{x})^\top)^\top, \mathbf{f}_j(\mathbf{x}) = (f_{j0}(\mathbf{x}), f_{j1}(\mathbf{x}))^\top, \\
&\qquad f_{jk} : R^d \to R, k = 0, 1,\ \|\mathbf{f}\|^2_{\hat{\mathbf{M}}} < \infty\}, \\
\mathcal{H}(\hat{\mathbf{M}}) &= \{\mathbf{f} \in L_2(\hat{\mathbf{M}}) : f_{jk}(\mathbf{x}) = g_{jk}(x_j) \text{ for some function } g_{jk} : R \to R, \\
&\qquad j = 1, \ldots, d,\ k = 0, 1\},
\end{aligned}
$$

where

$$
\begin{aligned}
\|\mathbf{f}\|^2_{\hat{\mathbf{M}}} &= \int \mathbf{f}(\mathbf{x})^\top \hat{\mathbf{M}}(\mathbf{x}) \mathbf{f}(\mathbf{x}) d\mathbf{x}, \text{ and} \\
\hat{\mathbf{M}}(\mathbf{x}) &= \frac{1}{n} \sum_{i=1}^n \mathbf{v}(\mathbf{X}_i, \mathbf{Z}_i; \mathbf{x}) \mathbf{v}(\mathbf{X}_i, \mathbf{Z}_i; \mathbf{x})^\top K_{\mathbf{h}}(\mathbf{x}, \mathbf{X}_i).
\end{aligned}
$$

Note that (4) has finite values if and only if $\|\mathbf{f}\|^2_{\hat{\mathbf{M}}} < \infty$. Therefore, our minimization problem takes place in the space $\mathcal{H}(\hat{\mathbf{M}})$. Note further that (4) can be decomposed into two parts as (see Lee et al. [2012a])

$$
\int \frac{1}{n} \sum_{i=1}^n \left[Y_i^G - \mathbf{v}(\mathbf{X}_i, \mathbf{Z}_i; \mathbf{x})^\top \tilde{\boldsymbol{\alpha}}^G(\mathbf{x}) \right]^2 K_{\mathbf{h}}(\mathbf{x}, \mathbf{X}_i) d\mathbf{x} + \|\tilde{\boldsymbol{\alpha}}^G - \mathbf{f}\|^2_{\hat{\mathbf{M}}},
$$

by introducing

$$
\tilde{\boldsymbol{\alpha}}^G(\mathbf{x}) = \hat{\mathbf{M}}(\mathbf{x})^{-1} \frac{1}{n} \sum_{i=1}^n \mathbf{v}(\mathbf{X}_i, \mathbf{Z}_i; \mathbf{x}) Y_i^G K_{\mathbf{h}}(\mathbf{x}, \mathbf{X}_i),
$$

which is the minimizer of (4) in the space $L_2(\hat{\mathbf{M}})$. Equipped with the norm $\|\cdot\|_{\hat{\mathbf{M}}}$, the spaces defined above are Hilbert spaces and our estimator $\hat{\boldsymbol{\alpha}}^G$ can be expressed as follows:

$$
\hat{\boldsymbol{\alpha}}^G = \underset{\mathbf{f} \in \mathcal{H}(\hat{\mathbf{M}})}{\arg\min} \|\tilde{\boldsymbol{\alpha}}^G - \mathbf{f}\|_{\hat{\mathbf{M}}} = \Pi(\tilde{\boldsymbol{\alpha}}^G | \mathcal{H}(\hat{\mathbf{M}})),
$$

where the operator $\Pi(\cdot|S)$ stands for a projection onto S. Note that $\hat{\boldsymbol{\alpha}}^G$ is unique since it is defined as a projection onto the Hilbert space $\mathcal{H}(\hat{\mathbf{M}})$. Moreover, by considering Gâteaux derivatives, one can show that $\hat{\boldsymbol{\alpha}}^G = (\hat{\boldsymbol{\alpha}}_1^G, \ldots, \hat{\boldsymbol{\alpha}}_d^G)^\top$ satisfies the following SBF equation:

$$
\hat{\boldsymbol{\alpha}}_j^G(x_j) = \tilde{\boldsymbol{\alpha}}_j^G(x_j) - \sum_{k \neq j} \int \hat{\mathbf{Q}}_j(x_j)^{-1} \hat{\mathbf{Q}}_{jk}(x_j, x_k) \hat{\boldsymbol{\alpha}}_k^G(x_k) dx_k, \quad \forall j = 1, \ldots, d, \tag{5}
$$

where

$$\hat{\mathbf{Q}}_j(x_j) = \frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix} 1 & \frac{X_{i,j}-x_j}{h_j} \\ \frac{X_{i,j}-x_j}{h_j} & \left(\frac{X_{i,j}-x_j}{h_j}\right)^2 \end{pmatrix} K_{h_j}(x_j, X_{i,j}) Z_{i,j}^2,$$

$$\hat{\mathbf{Q}}_{jk}(x_j, x_k) = \frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix} 1 & \frac{X_{i,k}-x_k}{h_k} \\ \frac{X_{i,j}-x_j}{h_j} & \left(\frac{X_{i,j}-x_j}{h_j}\right)\left(\frac{X_{i,k}-x_k}{h_k}\right) \end{pmatrix} K_{h_j}(x_j, X_{i,j}) K_{h_k}(x_k, X_{i,k}) Z_{i,j} Z_{i,k}, \text{ and}$$

$$\tilde{\boldsymbol{\alpha}}_j^G(x_j) = \begin{pmatrix} \tilde{\alpha}_{j0}^G(x_j) \\ \tilde{\alpha}_{j1}^G(x_j) \end{pmatrix} = \hat{\mathbf{Q}}_j(x_j)^{-1}\frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix} 1 \\ \frac{X_{i,j}-x_j}{h_j} \end{pmatrix} K_{h_j}(x_j, X_{i,j}) Z_{i,j} Y_i^G. \tag{6}$$

Note that in general $\tilde{\boldsymbol{\alpha}}^G(\mathbf{x})$ is not equal to $(\tilde{\boldsymbol{\alpha}}_1^G(x_1)^\top, \ldots, \tilde{\boldsymbol{\alpha}}_d^G(x_d)^\top)^\top$, since $\tilde{\boldsymbol{\alpha}}^G$ does not belong to $\mathcal{H}(\hat{\mathbf{M}})$. The solution of (5) is given by the following SBF algorithm:

$$\begin{aligned} \hat{\boldsymbol{\alpha}}_j^{G,[r]}(x_j) &= \tilde{\boldsymbol{\alpha}}_j^G(x_j) - \sum_{k=1}^{j-1}\int \hat{\mathbf{Q}}_j(x_j)^{-1}\hat{\mathbf{Q}}_{jk}(x_j, x_k)\hat{\boldsymbol{\alpha}}_k^{G,[r-1]}(x_k)dx_k \qquad (7) \\ &\quad - \sum_{k=j+1}^{d}\int \hat{\mathbf{Q}}_j(x_j)^{-1}\hat{\mathbf{Q}}_{jk}(x_j, x_k)\hat{\boldsymbol{\alpha}}_k^{G,[r]}(x_k)dx_k, \quad \forall j = 1, \ldots, d. \end{aligned}$$

One can iterate the above algorithm for $r = 1, 2, \ldots$, with some initial values $\hat{\boldsymbol{\alpha}}_j^{G,[0]}(x_j)$ ($j = 1, \ldots, d$), until it converges. Then, the limit of the algorithm is the estimate of the coefficient function. Note that, here the first component of $\hat{\boldsymbol{\alpha}}_j^G(x_j)$ estimates $\alpha_j(x_j)$, and the second one estimates $h_j\alpha_j'(x_j)$.

Remark 1. Let

$$\mathbf{M}(\mathbf{x}) = p(\mathbf{x})\left[E(\mathbf{Z}\mathbf{Z}^\top|\mathbf{X} = \mathbf{x}) \otimes \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + diag(E(Z_j^2|\mathbf{X} = \mathbf{x})) \otimes \begin{pmatrix} 0 & 0 \\ 0 & \int u^2 K(u)du \end{pmatrix}\right],$$

where $\otimes$ denotes the Kronecker product and p is the density function of $\mathbf{X}$. Then, $\|\mathbf{f}\|_{\mathbf{M}}$, $L_2(\mathbf{M})$ and $\mathcal{H}(\mathbf{M})$ can be defined similarly as $\|\mathbf{f}\|_{\hat{\mathbf{M}}}$, $L_2(\hat{\mathbf{M}})$ and $\mathcal{H}(\hat{\mathbf{M}})$, respectively, by replacing $\hat{\mathbf{M}}$ by $\mathbf{M}$. Note that $\hat{\mathbf{M}}(\mathbf{x})$ converges to $\mathbf{M}(\mathbf{x})$ in a certain sense under the assumptions given below.

Lee et al. [2012a] introduced the SBF algorithm to solve the SBF equation for non-censored data. Using the same arguments as therein, one can show that, under Assumption (B) given below, as $r \to \infty$, $\hat{\boldsymbol{\alpha}}^{G,[r]}$ converges to

$$\sum_{l=0}^{\infty}\hat{U}^l\hat{\mathbf{r}}^G, \tag{8}$$

where

$$\hat{U} = \hat{P}_d \cdots \hat{P}_1, \quad \hat{P}_j = \Pi(\cdot|\mathcal{H}_j(\hat{\mathbf{M}})^{\perp}),$$
$$\hat{\mathbf{r}}^G = (I - \hat{U})\tilde{\boldsymbol{\alpha}}^G,$$

where $S^{\perp}$ stands for the orthogonal complement of S and $\mathcal{H}_j(\hat{\mathbf{M}})$ $(j = 1, \ldots, d)$ are subspaces of $\mathcal{H}(\hat{\mathbf{M}})$ defined as

$$\mathcal{H}_j(\hat{\mathbf{M}}) = \{\mathbf{f} \in L_2(\hat{\mathbf{M}}) : f_{jk}(\mathbf{x}) = g_{jk}(x_j), \text{ for some function } g_{jk} : R \to R, \\ f_{lk}(\mathbf{x}) = 0, \ l \neq j, \ k = 0, 1\}.$$

Formula (8) is very useful to derive asymptotic results since it gives an explicit formula for the limit of the SBF algorithm. In the following, we collect the assumptions needed for the convergence of the SBF algorithm and the asymptotic results.

Assumption B

(B1) $E(\mathbf{Z}\mathbf{Z}^{\top}|\mathbf{X} = \mathbf{x})$ is continuous and its smallest eigenvalue is bounded away from zero on $\mathbf{x} \in [0, 1]^d$.

(B2) $\sup_{\mathbf{x}\in[0,1]^d} E(Z_j^4|\mathbf{X} = \mathbf{x}) < \infty$ for $j = 1, \ldots, d$.

(B3) The density p of $\mathbf{X}$ is bounded away from zero and is continuous on $[0, 1]^d$.

(B4) K is a bounded and symmetric density function supported on [-1,1] and is Lipschitz continuous.

(B5) The bandwidth h_j satisfies $h_j \to 0$ and $\log n/nh_j \to 0$ as $n \to \infty$ for $j = 1, \ldots, d$.

Assumption C

(C1) $E(\mathbf{Z}\mathbf{Z}^{\top}\sigma_G^2(\mathbf{X}, \mathbf{Z})|\mathbf{X} = \mathbf{x})$ is continuous in $\mathbf{x} \in [0, 1]^d$, where $\sigma_G^2(\mathbf{X}, \mathbf{Z}) = Var(Y^G|\mathbf{X}, \mathbf{Z})$.

(C2) The function α_j, $j = 1, \ldots, d$, is twice continuously differentiable on (0,1) and $E(Z_j Z_k|\mathbf{X} = \mathbf{x})$ is continuously partially differentiable in $\mathbf{x} \in (0, 1)^d$ for $j, k = 1, \ldots, d$.

Under Assumption (B), one can show that $\|\hat{U}\|_{op} < 1$ and $\|\hat{\mathbf{r}}^G\|_{\mathbf{M}} < \infty$ with probability tending to one, where $\| \cdot \|_{op}$ denotes the operator norm defined in the space $\mathcal{H}(\mathbf{M})$. If we choose a starting point satisfying $\|\hat{\boldsymbol{\alpha}}^{G,[0]}\|_{\mathbf{M}} < \infty$, then it can be shown that (8) is indeed the unique solution of the SBF equation (5). The following Lemma is a direct application of Theorems 3 and 4 in Lee et al. [2012a].

Lemma 1. *Under Assumption (B), $\hat{\boldsymbol{\alpha}}^{G,[r]}$ converges to the unique solution $\hat{\boldsymbol{\alpha}}^{G}$ of (5) with probability tending to one provided that the initial point satisfies $\|\hat{\boldsymbol{\alpha}}^{G,[0]}\|_{\mathbf{M}} < \infty$. Moreover, under Assumptions (B) and (C), if h_j and $n^{-1/5}$ are of the same order, then for any $\mathbf{x} \in (0,1)^d$ and for $j = 1, \dots, d$, $\hat{\boldsymbol{\alpha}}_j^G(x_j)$ are asymptotically independent, and*

$$n^{2/5}(\hat{\boldsymbol{\alpha}}_j^G(x_j) - \boldsymbol{\alpha}_j(x_j)) \to N(\boldsymbol{\beta}_j(x_j), \mathbf{V}_j(x_j)),$$

where

$$\begin{aligned} \boldsymbol{\beta}_j(x_j) &= \frac{b_j^2}{2}\alpha_j{}''(x_j)\begin{pmatrix} \mu_2(K) \\ 0 \end{pmatrix}, \text{ and} \\ \mathbf{V}_j(x_j) &= \frac{E(Z_j^2\sigma_G^2(\mathbf{X},\mathbf{Z})|X_j = x_j)}{b_j p_j(x_j)(E(Z_j^2|X_j = x_j))^2}\begin{pmatrix} \mu_0(K^2) & \frac{\mu_1(K^2)}{\mu_2(K)} \\ \frac{\mu_1(K^2)}{\mu_2(K)} & \frac{\mu_2(K^2)}{\mu_2(K)^2} \end{pmatrix}, \end{aligned}$$

with $b_j = \lim_{n\to\infty} n^{1/5}h_j$, $\mu_l(K^m) = \int u^l K^m(u)du$ and p_j is the marginal density of X_j.

Note that, since $\sigma_G^2(\mathbf{x},\mathbf{z}) \geq Var(\phi(Y)|\mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z})$, the asymptotic variance $\mathbf{V}_j(x_j)$ is larger than the corresponding asymptotic variance for the uncensored case. This is a common situation that arises in censored data since the synthetic data method inflates uncensored observations.

3.2 Smooth backfitting with censored data when G is unknown

We defined our estimator and derived its asymptotic distribution in the previous section as if we knew the censoring distribution G. However in practice G is, unfortunately, unknown, but it can be consistently estimated by the following Kaplan-Meier estimator $\hat{G}$ given by

$$1 - \hat{G}(t) = \prod_{i=1}^{n}\left(1 - \frac{(1-\delta_i)I(T_i \leq t)}{\sum_{j=1}^{n} I(T_j \geq T_i)}\right).$$

Replacing G by $\hat{G}$ in (4) gives the following redefined loss function $SL^{\hat{G}}(\mathbf{f})$:

$$SL^{\hat{G}}(\mathbf{f}) = \int \frac{1}{n}\sum_{i=1}^{n}\left[Y_i^{\hat{G}} - \mathbf{v}(\mathbf{X}_i, \mathbf{Z}_i; \mathbf{x})^\top \mathbf{f}(\mathbf{x})\right]^2 K_{\mathbf{h}}(\mathbf{x}, \mathbf{X}_i)d\mathbf{x}, \tag{9}$$

where $Y_i^{\hat{G}} = \delta_i \phi(T_i)/(1 - \hat{G}(T_i-))$. Then we define our estimator $\hat{\boldsymbol{\alpha}}^{\hat{G}}$ based on the estimated transformed data $Y_i^{\hat{G}}$ as follows:

$$\hat{\boldsymbol{\alpha}}^{\hat{G}} = \arg\min_{\mathbf{f} \in \mathcal{H}(\hat{\mathbf{M}})} SL^{\hat{G}}(\mathbf{f}) = \Pi(\tilde{\boldsymbol{\alpha}}^{\hat{G}} | \mathcal{H}(\hat{\mathbf{M}})),$$

where

$$\tilde{\boldsymbol{\alpha}}^{\hat{G}} = \arg\min_{\mathbf{f} \in L_2(\hat{\mathbf{M}})} SL^{\hat{G}}(\mathbf{f}) = \hat{\mathbf{M}}(\mathbf{x})^{-1} \frac{1}{n} \sum_{i=1}^{n} \mathbf{v}(\mathbf{X}_i, \mathbf{Z}_i; \mathbf{x}) Y_i^{\hat{G}} K_{\mathbf{h}}(\mathbf{x}, \mathbf{X}_i).$$

The estimator $\hat{\boldsymbol{\alpha}}^{\hat{G}}$ satisfies the SBF equation (5) with G being replaced by $\hat{G}$. Unlike the case where G is known, the direct application of the theorems in Lee et al. [2012a] is not valid when G is estimated by the Kaplan-Meier estimator since $Y_1^{\hat{G}}, \ldots, Y_n^{\hat{G}}$ are not independent. Below is a useful lemma for investigating the properties of the SBF algorithm. Recall that the definition of $\tilde{\boldsymbol{\alpha}}_j^G(x_j)$ is given in (6). The estimator $\tilde{\boldsymbol{\alpha}}_j^{\hat{G}}(x_j)$ is defined by replacing G by $\hat{G}$.

Lemma 2. *Under Assumption (B) and for $j = 1, \ldots, d$,*

$$\tilde{\boldsymbol{\alpha}}_j^{\hat{G}}(x_j) - \tilde{\boldsymbol{\alpha}}_j^G(x_j) = O_p\left(\sup_{t \le \tau_0} |\hat{G}(t) - G(t)|\right),$$

uniformly in $x_j \in [0, 1]$.

This lemma tells us that the difference between $\tilde{\boldsymbol{\alpha}}_j^{\hat{G}}(x_j)$ and $\tilde{\boldsymbol{\alpha}}_j^G(x_j)$ is uniformly bounded by the approximation error of the censoring distribution.

3.2.1 Convergence of the smooth backfitting algorithm

As in Section 2.1, one can find the solution of the SBF equation by the application of the SBF algorithm with G being replaced by $\hat{G}$. Since $\|\hat{U}\|_{op} < 1$ is already established in Lee et al. [2012a], to show the convergence of the SBF algorithm, it suffices to show that $\|\hat{\mathbf{r}}^{\hat{G}}\|_{\mathbf{M}} < \infty$, where $\hat{\mathbf{r}}^{\hat{G}} = (I - \hat{U})\tilde{\boldsymbol{\alpha}}^{\hat{G}}$, with an initial value satisfying $\|\hat{\boldsymbol{\alpha}}^{\hat{G},[0]}\|_{\mathbf{M}} < \infty$.

Theorem 1. *Under Assumption (B), with probability tending to one, the SBF algorithm converges to the unique solution $\hat{\boldsymbol{\alpha}}^{\hat{G}} = \sum_{l=0}^{\infty} \hat{U}^l \hat{\mathbf{r}}^{\hat{G}}$ provided that $\sup_{t \le \tau_0} |\hat{G}(t) - G(t)| = o_p(1)$, and that the initial point satisfies $\|\hat{\boldsymbol{\alpha}}^{\hat{G},[0]}\|_{\mathbf{M}} < \infty$.*

Remark 2. There may exist many possible choices for the initial point. Among them, $\hat{\boldsymbol{\alpha}}^{\hat{G},[0]} = (\tilde{\boldsymbol{\alpha}}_j^{\hat{G}}(x_1)^\top, \ldots, \tilde{\boldsymbol{\alpha}}_j^{\hat{G}}(x_d)^\top)^\top$ can be a good suggestion since with this choice,

$$\|\hat{\boldsymbol{\alpha}}^{\hat{G},[0]}\|_{\mathbf{M}} \leq C_1 \sum_{j=1}^{d} \left[\int \tilde{\alpha}_{j0}^{\hat{G}}(x_j)^2 q_j(x_j) dx_j + \mu_2(K) \cdot \int \tilde{\alpha}_{j1}^{\hat{G}}(x_j)^2 q_j(x_j) dx_j \right]^{\frac{1}{2}},$$

for some positive constant C_1, where $q_j(x_j) = E(Z_j^2|X_j = x_j)p_j(x_j)$.

Remark 3. The consistency of $\hat{G}$ is not necessary for the convergence of the algorithm. The convergence would be guaranteed if the estimated transformed response $Y^{\hat{G}}$ satisfies some moment condition. However, the limit of the algorithm may not estimate the true coefficient functions consistently unless the estimator of G is consistent.

3.2.2 Asymptotic distribution of the smooth backfitting estimator

In this subsection, the asymptotic distribution of the SBF estimator $\hat{\boldsymbol{\alpha}}^{\hat{G}}(\mathbf{x})$ will be presented. Recall that the asymptotic distribution of $\hat{\boldsymbol{\alpha}}^{G}(\mathbf{x})$ was already given in Lemma 1. If we show that the difference between $\hat{\boldsymbol{\alpha}}^{\hat{G}}(\mathbf{x})$ and $\hat{\boldsymbol{\alpha}}^{G}(\mathbf{x})$ is negligible at a certain rate, the desired result will follow. This can be done by using the fact that $\hat{\boldsymbol{\alpha}}^{\hat{G}}$ and $\hat{\boldsymbol{\alpha}}^{G}$ are the further projections of $\tilde{\boldsymbol{\alpha}}^{\hat{G}}$ and $\tilde{\boldsymbol{\alpha}}^{G}$ onto $\mathcal{H}(\hat{\mathbf{M}})$. We demonstrated in Lemma 2 that the difference between $\tilde{\boldsymbol{\alpha}}_j^{\hat{G}}(x_j)$ and $\tilde{\boldsymbol{\alpha}}_j^{G}(x_j)$ is bounded by the approximation error of G in probability. Note here that $(\tilde{\boldsymbol{\alpha}}_1^{G}(x_1)^\top, \ldots, \tilde{\boldsymbol{\alpha}}_d^{G}(x_d)^\top)^\top$ differs in general from $\tilde{\boldsymbol{\alpha}}^{G}(\mathbf{x})$, which means that $\hat{\boldsymbol{\alpha}}^{G}(\mathbf{x})$ is not the projection of $(\tilde{\boldsymbol{\alpha}}_1^{G}(x_1)^\top, \ldots, \tilde{\boldsymbol{\alpha}}_d^{G}(x_d)^\top)^\top$. The same is true when G is replaced by $\hat{G}$. However, since $\hat{\boldsymbol{\alpha}}_j^{G}(x_j)$ (i.e., the jth component function of the projection of $\tilde{\boldsymbol{\alpha}}^{G}$) has the same asymptotic variance as $\tilde{\boldsymbol{\alpha}}_j^{G}(x_j)$, we can expect it is also true for $\hat{\boldsymbol{\alpha}}_j^{\hat{G}}(x_j)$ and $\tilde{\boldsymbol{\alpha}}_j^{\hat{G}}(x_j)$. We will prove the next lemma using this idea.

Lemma 3. *Under Assumption (B), if h_j and $n^{-1/5}$ are of the same order, then for any $\mathbf{x} \in [0,1]^d$,*

$$\hat{\boldsymbol{\alpha}}^{\hat{G}}(\mathbf{x}) - \hat{\boldsymbol{\alpha}}^{G}(\mathbf{x}) = O_p\left(\sup_{t \leq \tau_0} |\hat{G}(t) - G(t)|\right) + o_p(n^{-2/5}).$$

From Lemma 3, we conclude that $\hat{\boldsymbol{\alpha}}^{\hat{G}}(\mathbf{x}) - \hat{\boldsymbol{\alpha}}^{G}(\mathbf{x}) = o_p(n^{-2/5})$ for any $\mathbf{x} \in [0,1]^d$ if G is continuous, since G is approximated at the rate $O_p((\log n/n)^{1/2})$ by the Kaplan-Meier estimator (see e.g. Lo and Singh [1986]). We already know the asymptotic distribution of $\hat{\boldsymbol{\alpha}}^{G}(\mathbf{x})$ and its rate of convergence. So, a direct application of Lemma 3 together with Lemma 1 gives the following theorem.

Theorem 2. *Under Assumptions (B) and (C), if h_j and $n^{-1/5}$ are of the same order and if G is continuous, then for any $\mathbf{x} \in (0,1)^d$ and for $j = 1, \ldots, d$, $\hat{\boldsymbol{\alpha}}_j^{\hat{G}}(x_j)$ are asymptotically independent, and*

$$n^{2/5}(\hat{\boldsymbol{\alpha}}_j^{\hat{G}}(x_j) - \boldsymbol{\alpha}_j(x_j)) \to N(\boldsymbol{\beta}_j(x_j), \mathbf{V}_j(x_j)),$$

where $\boldsymbol{\beta}_j(x_j)$ and $\mathbf{V}_j(x_j)$ are defined in the statement of Lemma 1.

4 Extension to local polynomial fitting

In this section, we extend the results studied in the previous section to the local polynomial setting. We focus on the case of odd orders since they are known to be preferable to the even order cases; see Section 3.3.2 in Fan and Gijbels [1996]. We will briefly show the results without proofs. The following is the redefined loss function to be minimized for the estimation of the coefficient functions, where we approximate the coefficient functions by a pth order Taylor expansion:

$$SL_p^{\hat{G}}(\mathbf{f}) = \int \frac{1}{n} \sum_{i=1}^{n} \left[Y_i^{\hat{G}} - \sum_{j=1}^{d} \mathbf{w}_j(x_j, X_{i,j})^\top \mathbf{f}_j(x_j) Z_{i,j} \right]^2 K_{\mathbf{h}}(\mathbf{x}, \mathbf{X}_i) d\mathbf{x},$$

where $\mathbf{f} = (\mathbf{f}_1^\top, \ldots, \mathbf{f}_d^\top)^\top$, $\mathbf{f}_j = (f_{j0}, \ldots, f_{jp})^\top$ for univariate functions f_{jk},

$$\mathbf{w}_j(v_j, u_j) = \left(1, \left(\frac{u_j - v_j}{h_j}\right), \ldots, \left(\frac{u_j - v_j}{h_j}\right)^p\right)^\top,$$

and $\hat{G}$ is the Kaplan-Meier estimator of G. Let $\hat{\boldsymbol{\alpha}}_p^{\hat{G}}$ be the minimizer of $SL_p^{\hat{G}}(\mathbf{f})$ over $\mathbf{f}$ when $SL_p^{\hat{G}}(\mathbf{f}) < \infty$. Then, $\hat{\boldsymbol{\alpha}}_p^{\hat{G}}(\mathbf{x}) = (\hat{\boldsymbol{\alpha}}_{p,1}^{\hat{G}}(x_1)^\top, \ldots, \hat{\boldsymbol{\alpha}}_{p,d}^{\hat{G}}(x_d)^\top)^\top$ is the local polynomial SBF estimator and satisfies the SBF equation analogous to (5), which we will not present in detail. Moreover, the SBF algorithm to find the minimizer can be given in the same way as in (7) and its convergence is guaranteed with probability tending to one under Assumption (B). Note that the estimator of the kth derivative of $\alpha_j(x_j)$ is given by $k!\hat{\alpha}_{p,jk}^{\hat{G}}(x_j)/h_j^k$, where $\hat{\boldsymbol{\alpha}}_{p,j}^{\hat{G}}(x_j) = (\hat{\alpha}_{p,j0}^{\hat{G}}(x_j), \ldots, \hat{\alpha}_{p,jp}^{\hat{G}}(x_j))^\top$, since $\hat{\alpha}_{p,jk}^{\hat{G}}(x_j)$ is an estimator of $h_j^k \alpha_j^{(k)}(x_j)/k!$.

We need an additional smoothness condition for the asymptotic distribution of the local polynomial SBF estimator:

(C2′) The function α_j, $j = 1, \ldots, d$, is $p+1$ times continuously differentiable on (0,1), and $E(Z_j Z_k | \mathbf{X} = \mathbf{x})$ is continuously partially differentiable in $\mathbf{x} \in (0,1)^d$ for $j, k = 1, \ldots, d$.

The next lemma is analogous to Lemma 3 and gives the approximation error between $\hat{\boldsymbol{\alpha}}_p^{\hat{G}}(\mathbf{x})$ and $\hat{\boldsymbol{\alpha}}_p^{G}(\mathbf{x})$. The proof is omitted.

Lemma 4. *Under Assumption (B), if h_j and $n^{-1/(2p+3)}$ are of the same order, then for any $\mathbf{x} \in [0,1]^d$,*

$$\hat{\boldsymbol{\alpha}}_p^{\hat{G}}(\mathbf{x}) - \hat{\boldsymbol{\alpha}}_p^{G}(\mathbf{x}) = O_p\left(\sup_{t \le \tau_0} |\hat{G}(t) - G(t)|\right) + o_p(n^{-(p+1)/(2p+3)}),$$

for odd p.

Now, the following theorem follows from Lemma 4.

Theorem 3. *Under Assumptions (B), (C1) and (C2′), if h_j and $n^{-1/(2p+3)}$ are of the same order and if G is continuous, then for any $\mathbf{x} \in (0,1)^d$ and for $j = 1, \ldots, d$, $\hat{\boldsymbol{\alpha}}_{p,j}^{\hat{G}}(x_j)$ are asymptotically independent, and*

$$n^{(p+1)/(2p+3)}(\hat{\boldsymbol{\alpha}}_{p,j}^{\hat{G}}(x_j) - \boldsymbol{\alpha}_j(x_j)) \to N(\boldsymbol{\beta}_{p,j}(x_j), \mathbf{V}_{p,j}(x_j)),$$

for odd p, where

$$\begin{aligned}
\boldsymbol{\beta}_{p,j}(x_j) &= \frac{b_j^{p+1}}{(p+1)!}\alpha_j^{(p+1)}(x_j)\boldsymbol{\Omega}_1^{-1}\boldsymbol{\eta} \\
\mathbf{V}_{p,j}(x_j) &= \frac{E(Z_j^2\sigma_G^2(\mathbf{X},\mathbf{Z})|X_j = x_j)}{b_j p_j(x_j)(E(Z_j^2|X_j = x_j))^2}\boldsymbol{\Omega}_1^{-1}\boldsymbol{\Omega}_2\boldsymbol{\Omega}_1^{-1} \\
(\boldsymbol{\Omega}_1)_{l,m} &= \mu_{l+m}(K),\ l,m = 0,\ldots,p \\
(\boldsymbol{\Omega}_2)_{l,m} &= \mu_{l+m}(K^2),\ l,m = 0,\ldots,p \\
(\boldsymbol{\eta})_l &= \mu_{p+1+l}(K),\ l = 0,\ldots,p,
\end{aligned}$$

with $b_j = \lim_{n\to\infty} n^{1/(2p+3)}h_j$ and p_j is the marginal density of X_j.

5 Simulation study

In this section, we will present the finite sample performance of the proposed estimator. We generate random samples from the following model:

$$Y = m(\mathbf{X},\mathbf{Z}) + \sigma(\mathbf{X},\mathbf{Z})\epsilon,$$

where $m(\mathbf{X},\mathbf{Z}) = Z_1\alpha_1(X_1) + Z_2\alpha_2(X_2) + Z_3\alpha_3(X_3)$. The variables X_1, X_2 and X_3 are generated from $U[0,1]$, and the vector $(Z_2, Z_3)^\top$ from a bivariate normal distribution with mean $(0,0)^\top$, and variance $\begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$, independently of $\mathbf{X} = (X_1, X_2, X_3)^\top$. We take $Z_1 \equiv 1$, $\alpha_1(x) = 1 + \exp(2x-1)$, $\alpha_2(x) = 0.5\cos(2\pi x)$ and $\alpha_3(x) = x^2$. The standard deviation function is set to $\sigma(\mathbf{x}, \mathbf{z}) = 0.5 + \frac{z_2^2+z_3^2}{1+z_2^2+z_3^2}\exp(-2 + \frac{x_1+x_2}{2})$. The error ϵ was generated from a normal distribution with mean 0 and standard deviation γ. A similar model was considered in Yang et al. [2006] and Lee et al. [2012a]. We also generate a normal censoring variable with mean μ and variance 1.5. Here, μ was selected to control the percentage of censoring (PC). We set $\phi(t) = tI(t \leq \tau_0)$ so that the objective of this study is to estimate the truncated conditional mean of Y given $\mathbf{X} = \mathbf{x}$ and $\mathbf{Z} = \mathbf{z}$. For truncation, $\tau_0 = 5$ was used, which means that only a small proportion of the observed T_i's are truncated. We examine the performance of our estimator for several choices of PC. We try three cases $\mu = 4.4197, 3.1083$ and 2.2, which yields approximately 10%, 30% and 50% of censoring, respectively. The noncensored case is also considered to see how random right censoring affects the estimation in our model.

The coefficient functions are estimated by the local linear SBF method. The trapezoidal rule with 51 equally spaced grid points on [0,1] is used for the numerical integration. We compute the estimated mean integrated squared error (MISE) of the regression function:

$$\begin{aligned}
\text{MISE} &= \frac{1}{T}\frac{1}{N}\sum_{k=1}^{T}\sum_{j=1}^{N}\Big(\hat{m}^{[j]}(\mathbf{X}_k, \mathbf{Z}_k) - m(\mathbf{X}_k, \mathbf{Z}_k)\Big)^2 \\
&= \frac{1}{T}\frac{1}{N}\sum_{k=1}^{T}\sum_{j=1}^{N}\left(\hat{m}^{[j]}(\mathbf{X}_k, \mathbf{Z}_k) - \frac{1}{N}\sum_{l=1}^{N}\hat{m}^{[l]}(\mathbf{X}_k, \mathbf{Z}_k)\right)^2 \\
&\quad + \frac{1}{T}\sum_{k=1}^{T}\left(\frac{1}{N}\sum_{l=1}^{N}\hat{m}^{[l]}(\mathbf{X}_k, \mathbf{Z}_k) - m(\mathbf{X}_k, \mathbf{Z}_k)\right)^2 \\
&= \text{IV} + \text{IB}^2,
\end{aligned}$$

where N stands for the number of replications, T for the size of a test sample and $\hat{m}^{[j]}$, $j = 1, \cdots, N$, is the local linear SBF estimator for each replication. We choose $N = 500$ and $T = 500$. We try 8^3 bandwidth choices $(h_1, h_2, h_3) \in \{0.05, 0.15, \ldots, 0.75\}^3$, and the Epanechnikov kernel is used for the kernel K.

We run simulations for different sample sizes, different noise levels and different censoring percentages. Tables 1 and 2 report the results for sample sizes $n = 200$ and 400, and for different values of γ and PC. Each time we report the result for the bandwidth vector which minimizes the MISE. With the optimal bandwidth, which yields the optimal result for each setting, the MISE

Table 1: Optimal results when estimating the regression function m for $n = 200$. Here, γ is the standard deviation of the error, PC is the percentage of censoring, FUN is the function of interest, MISE is the mean integrated squared error, IV is the integrated variance, and IB^2 is the integrated squared bias.

n	γ	PC(%)	FUN	MISE	IV	IB^2
200	1	0	α_1	0.0185	0.0098	0.0087
			α_2	0.0270	0.0189	0.0081
			α_3	0.0194	0.0136	0.0058
			m	0.0654	0.0429	0.0225
		10	α_1	0.0502	0.0236	0.0267
			α_2	0.0880	0.0697	0.0183
			α_3	0.0537	0.0447	0.0091
			m	0.1900	0.1406	0.0494
		30	α_1	0.1633	0.0702	0.0932
			α_2	0.2029	0.1339	0.0690
			α_3	0.1479	0.1278	0.0201
			m	0.5013	0.3350	0.1664
		50	α_1	0.4293	0.1568	0.2725
			α_2	0.3556	0.2703	0.0853
			α_3	0.2910	0.2368	0.0542
			m	1.0254	0.6312	0.3942
	1.5	0	α_1	0.0501	0.0163	0.0338
			α_2	0.0470	0.0336	0.0134
			α_3	0.0322	0.0210	0.0112
			m	0.1271	0.0701	0.0570
		10	α_1	0.0933	0.0345	0.0588
			α_2	0.1173	0.0796	0.0378
			α_3	0.0753	0.0586	0.0167
			m	0.2750	0.1718	0.1032
		30	α_1	0.2535	0.1029	0.1507
			α_2	0.2416	0.1632	0.0784
			α_3	0.1968	0.1665	0.0303
			m	0.6698	0.4252	0.2446
		50	α_1	0.5794	0.1985	0.3810
			α_2	0.3915	0.3014	0.0901
			α_3	0.3526	0.2926	0.0600
			m	1.2723	0.7593	0.5130

Table 2: Optimal results when estimating the regression function m for $n = 400$. Here, γ is the standard deviation of the error, PC is the percentage of censoring, FUN is the function of interest, MISE is the mean integrated squared error, IV is the integrated variance, and IB^2 is the integrated squared bias.

n	γ	PC(%)	FUN	MISE	IV	IB^2
400	1	0	α_1	0.0126	0.0056	0.0070
			α_2	0.0169	0.0134	0.0035
			α_3	0.0123	0.0069	0.0054
			m	0.0382	0.0235	0.0147
		10	α_1	0.0316	0.0148	0.0168
			α_2	0.0560	0.0379	0.0181
			α_3	0.0334	0.0243	0.0091
			m	0.1067	0.0688	0.0379
		30	α_1	0.0926	0.0394	0.0532
			α_2	0.1543	0.1114	0.0429
			α_3	0.0988	0.0792	0.0196
			m	0.3059	0.2064	0.0995
		50	α_1	0.2594	0.1008	0.1586
			α_2	0.2710	0.1850	0.0860
			α_3	0.2106	0.1668	0.0438
			m	0.6618	0.4096	0.2522
	1.5	0	α_1	0.0394	0.0079	0.0315
			α_2	0.0298	0.0170	0.0128
			α_3	0.0218	0.0104	0.0114
			m	0.0829	0.0322	0.0507
		10	α_1	0.0615	0.0163	0.0451
			α_2	0.0740	0.0492	0.0248
			α_3	0.0494	0.0349	0.0145
			m	0.1636	0.0890	0.0746
		30	α_1	0.1618	0.0589	0.1029
			α_2	0.1895	0.1254	0.0641
			α_3	0.1254	0.0973	0.0281
			m	0.4228	0.2525	0.1703
		50	α_1	0.3957	0.1301	0.2656
			α_2	0.2942	0.2067	0.0875
			α_3	0.2420	0.1906	0.0514
			m	0.8373	0.4760	0.3613

values for each coefficient function are also computed and presented in those Tables. As expected we find overall increasing patterns in MISE as PC and γ increase. In the censored cases, there is a tendency for the ratio IV/IB^2 to decline when PC=50%, which could be counterintuitive. One possible reason is that, with high PC, optimal bandwidths are selected to be very large to control the explosion of the variance, which results in relatively large biases. We also find that the MISE decreases as n doubles, and that the rate of decrease is close to $2^{-4/5} \approx 0.57$. Note here that the asymptotic MISE of our estimator is proportional to $n^{-4/5}$ with the optimal bandwidth rate $h_j \sim n^{-1/5}$. These results confirm that the proposed estimator works rather well.

6 Bandwidth parameter selection

In this section, we introduce a data-driven bandwidth selector for local linear fitting, which is based on the method given in Lee et al. [2012a]. They proposed to estimate the unknown quantities which appear in the optimal bandwidth minimizing the asymptotic mean integrated squared error by fitting some polynomial regression models. We simply adapt their method to the censored data context. From Theorem 2, the optimal bandwidth when local linear fitting is applied is given by $b_j^* n^{-1/5}$, where

$$b_j^* = \left(\frac{\int c_j(x_j)dx_j}{4\int d_j(x_j)^2 p_j(x_j)dx_j} \right)^{\frac{1}{5}},$$

$$c_j(x_j) = \frac{E(Z_j^2 \sigma_G^2(\mathbf{X},\mathbf{Z})|X_j = x_j)}{E(Z_j^2|X_j = x_j)^2} \mu_0(K^2), \text{ and} \tag{10}$$

$$d_j(x_j) = \frac{1}{2}\alpha_j''(x_j)\mu_2(K). \tag{11}$$

We estimate $\int \alpha_j''(x_j)^2 p_j(x_j)dx_j$ by $\frac{1}{n}\sum_{i=1}^n \hat{\alpha}_j''(X_{i,j})^2$ where $\hat{\alpha}_j''(x_j) = \sum_{k=2}^s k(k-1)c_{j,k}x_j^{k-2}$, with $c_{j,k}$ being the minimizers of

$$\sum_{i=1}^n \rho \left(Y_i^{\hat{G}} - \sum_{j=1}^d Z_{i,j} \left[\sum_{k=0}^s c_{j,k} X_{i,j}^k \right] \right), \tag{12}$$

and where ρ is a given loss function and s is the degree of the polynomial used to approximate $m(\mathbf{X}_i, \mathbf{Z}_i)$. Note that, to deal with censoring, the estimated synthetic response is used instead of the response itself. Other unknown quantities can be estimated in a similar manner. A natural choice for ρ would be the squared loss function $\rho(u) = u^2$. However, with this loss function, selected bandwidths produced unsatisfactory results. Note that, in formulae (10) and (11), only $E(Z_j^2\sigma_G^2(\mathbf{X},\mathbf{Z})|X_j = x_j)$ is affected by censoring, which means that, theoretically, other quantities are invariant regardless of the occurrence of censoring. Nevertheless, some large values of $Y_i^{\hat{G}}$ inflated by the unbiased transformation may cause a significant increase of the estimates of $\int \alpha_j''(x_j)^2 p_j(x_j)dx_j$ as the percentage of censoring increases. To address this problem, we use the following Huber loss function instead of the squared loss function for the estimation of $\int \alpha_j''(x_j)^2 p_j(x_j)dx_j$:

$$\rho_k(u) = \begin{cases} u^2/2 & \text{if } |u| < k \\ k(|u| - k/2) & \text{if } |u| \geq k \end{cases} .$$

This function is typically used in robust estimation. By employing this loss function, we expect that large values of $Y_i^{\hat{G}}$ can be prevented from having too much effect on estimating $\int \alpha_j''(x_j)^2 p_j(x_j)dx_j$.

Tables 3 and 4 show the performance of the above bandwidth selection procedure. We generate 500 random samples from the same model as in Section 5. The Gaussian kernel is used for the multivariate local linear kernel estimator, since the Epanechnikov kernel gives very poor estimates due to its compact support. To estimate the unknown quantities, we use a cubic polynomial for $\alpha_j(x_j)$ and a linear polynomial for the other functions. The tuning parameter k is set to $1.345\hat{\sigma}$ where $\hat{\sigma} = \text{MAD}/0.6745$, and MAD is the mean absolute deviation of the residuals.

Table 3 shows how the automatic bandwidth selector works. We compute the ratio of the MISE obtained with bandwidths $\hat{h}_{opt}$ and $\hat{h}_a$ respectively, that is, $\text{MISE}(\hat{h}_{opt})/\text{MISE}(\hat{h}_a)$. Here, $\hat{h}_{opt}$ is the optimal bandwidth described in Section 5 and $\hat{h}_a$ is the data-driven bandwidth proposed in this section. It follows from Table 3 that our bandwidth selector works reasonably well, since the values of the ratios are not so far from 1. The selection procedure is influenced by censoring, however, the noise level (γ) has no or only very limited effect. The ratios with censoring have relatively large values compared to the noncensored case. An interesting finding is that the ratios do not have an increasing trend in PC. It means that our selection procedure works well, and does not break down even with high PC. In Table 3, there is a ratio smaller than 1. Indeed, this can happen in finite sample studies, since our automatic bandwidth selector gives data-adaptive bandwidths for each sample whereas the optimal bandwidth is selected as the best one among a set of bandwidths that are the same for all samples.

We also compare our SBF estimator based on the above automatic bandwidth selector to the

Table 3: Ratio of the mean integrated squared error based on $\hat{h}_{opt}$ over the mean integrated squared error based on $\hat{h}_a$ for the local linear SBF estimator for $n = 200, 400$, where $\hat{h}_{opt}$ is the optimal bandwidth, and $\hat{h}_a$ is the data driven bandwidth. Here, γ is the standard deviation of the error, and PC is the percentage of censoring.

n	γ	PC(%)	Ratio	n	γ	PC(%)	Ratio
200	1	0	0.986	400	1	0	1.050
		10	1.153			10	1.127
		30	1.216			30	1.234
		50	1.191			50	1.190
	1.5	0	1.089		1.5	0	1.123
		10	1.200			10	1.192
		30	1.184			30	1.195
		50	1.114			50	1.145

Table 4: MISE of the local linear SBF estimator and the MK estimator with data driven bandwidth selectors for $n = 200$. Here, γ is the standard deviation of the error, PC is the percentage of censoring, IV is the integrated variance, and IB^2 is the integrated squared bias.

		SBF			MK		
γ	PC(%)	MISE	IV	IB^2	MISE	IV	IB^2
1	0	0.064	0.050	0.014	0.204	0.108	0.096
	10	0.219	0.181	0.038	0.585	0.435	0.149
	30	0.610	0.455	0.154	2.339	2.117	0.221
	50	1.221	0.796	0.425	5.625	5.193	0.432
1.5	0	0.138	0.087	0.052	0.294	0.129	0.165
	10	0.330	0.237	0.093	0.719	0.524	0.196
	30	0.793	0.565	0.228	3.259	2.954	0.305
	50	1.417	0.861	0.556	6.092	5.518	0.574

multivariate local linear kernel (MK) estimator based on the **np** package in R. The **np** package offers a bandwidth selector based on the cross-validation principle. In Table 4, we see that our SBF estimator outperforms the MK estimator. In particular, the integrated variance of the MK estimator increases very rapidly compared to the integrated bias as PC becomes high.

7 Discussion

In this paper, we propose a smooth backfitting (SBF) estimator for the coefficient functions in a varying coefficient model having different covariates as smoothing variables when there is random right censoring in the response. We focus on the case where the censoring does not depend on the covariates, which is the case, for example, when the censoring occurs at the end of the study. However, if there is some belief that censoring is affected by the characteristics of the subjects, then considering the dependency between the censoring variable and the covariates in the estimation procedure could be appealing. In this case, the synthetic response is given by

$$Y^{\hat{G}} = \frac{\delta\phi(T)}{1 - \hat{G}_{\mathbf{U}}(T-)},$$

where $G_{\mathbf{U}}$ denotes the conditional distribution of C given $\mathbf{U} = (\mathbf{X}^\top, \mathbf{Z}^\top)^\top$, that is, $G_{\mathbf{U}}(\cdot) = P(C \leq \cdot|\mathbf{U})$. Note that $\mathbf{U}$ rather than its value $\mathbf{u}$ is used here, since the SBF method is minimizing a global criterion induced by integration. This approach also preserves the conditional mean of Y given the covariates if we replace assumptions (A1) and (A2) by the conditional independence assumption between Y and C given $\mathbf{U}$. Similar ideas have been used in the literature. See Talamakrouni et al. [2012] for an example. In the dependent censoring case, the Beran estimator (Beran [1981]) can be used as an estimator of $G_{\mathbf{U}}$. Nevertheless, this may cause the well-known "curse of dimensionality" problem, because in our model the dimension of the covariates is large in general. Recall that the motivation for employing the SBF method is to avoid "curse of dimensionality" in fitting coefficient functions. In this case, it is possible to restrict attention to a proper subset of covariates as variables to estimate $G_{\mathbf{U}}$. Another alternative is to consider parametric or semiparametric models to avoid high dimensional smoothing. A final remark of this section is about the choice of the truncation point τ_0. This should be chosen carefully, since a poor choice may result in an estimator with large variance, or in too much truncation of the data.

Appendix

This section contains the proofs of the asymptotic results of Section 3. We start with the next lemma, which gives a uniform convergence result for kernel weighted averages.

Lemma A.1. *Let (X_i, Y_i) $i = 1, \ldots, n$ be independent and identically distributed random variables with joint density $f(x, y)$ and let K be a bounded, Lipschitz continuous and symmetric density function supported on a compact interval. Suppose that $E|Y_1|^s < \infty$ for some $s > 1$ and $\sup_{x\in\mathcal{X}} \int |y|^s f(x, y)dy < \infty$, where $\mathcal{X}$ is the support of X_1. Then, the following result holds with $K_h(u, v)$ defined in (3), assuming that $h \to 0$ and $n^\gamma h \to \infty$ for some $\gamma < 1 - s^{-1}$ as $n \to \infty$:*

$$\sup_{x\in\mathcal{X}} \left| \frac{1}{n} \sum_{i=1}^{n} \{K_h(x, X_i)Y_i - E(K_h(x, X_i)Y_i)\} \right| = o_p(1).$$

Proof. This follows from a slight modification of Proposition 4 in Mack and Silverman [1982], if we substitute the kernel function $(1/h)K((u - v)/h)$ therein by the boundary corrected kernel $K_h(u, v)$. □

Proof of Lemma 2. Write

$$\|\tilde{\boldsymbol{\alpha}}_j^{\hat{G}}(x_j) - \tilde{\boldsymbol{\alpha}}_j^{G}(x_j)\| \leq \max_{1\leq i\leq n} |Y_i^{\hat{G}} - Y_i^G| \; \|\hat{\mathbf{Q}}_j(x_j)^{-1}\|_2 \left\| \frac{1}{n} \sum_{i=1}^{n} (1, (X_{i,j} - x_j)/h_j)^\top K_{h_j}(x_j, X_{i,j}) Z_{i,j} \right\|,$$

where $\| \cdot \|_2$ denotes the spectral norm of a matrix. Note that

$$|Y_i^{\hat{G}} - Y_i^G| = \frac{\delta_i |\phi(T_i)|}{(1 - G(T_i-))(1 - \hat{G}(T_i-))} |\hat{G}(T_i-) - G(T_i-)|.$$

Therefore,

$$\begin{aligned} \max_{1\leq i\leq n} |Y_i^{\hat{G}} - Y_i^G| &\leq \sup_{t\leq\tau_0} \left\{ |\hat{G}(t) - G(t)| \frac{|\phi(t)|}{(1 - G(t))^2} \frac{(1 - G(t))}{(1 - \hat{G}(t))} \right\} \\ &\leq \sup_{t\leq\tau_0} |\hat{G}(t) - G(t)| O_p(1), \end{aligned}$$

by assumption (A3) and the fact $\sup_{t\leq\tau_0} \frac{1-G(t)}{1-\hat{G}(t)} = O_p(1)$; see e.g. Lemma A.1 in Lopez and Patilea [2009]. Now, it suffices to show that $\sup_{x_j\in[0,1]} \|\hat{\mathbf{Q}}_j(x_j)^{-1}\|_2$ and $\sup_{x_j\in[0,1]} \|\frac{1}{n}\sum_{i=1}^n (1, (X_{i,j} - x_j)/h_j)^\top K_{h_j}(x_j, X_{i,j})Z_{i,j}\|$ are bounded in probability. From Lemma A.1, $\hat{\mathbf{Q}}_j(x_j)$ converges uniformly in probability to $\mathbf{Q}_j(x_j)$ where

$$\mathbf{Q}_j(x_j) = \begin{pmatrix} 1 & 0 \\ 0 & \mu_2(K) \end{pmatrix} E(Z_j^2|X_j = x_j)p_j(x_j),$$

and where p_j is the marginal density of X_j. Then, we have

$$\begin{aligned}
\sup_{x_j\in[0,1]}\|\hat{\mathbf{Q}}_j(x_j)^{-1}\|_2 &= \frac{1}{\inf_{x_j\in[0,1]}\left\{\hat{\lambda}_{j0}(x_j)\wedge\hat{\lambda}_{j1}(x_j)\right\}} \\
&\leq \frac{1}{\inf_{x_j\in[0,1]}\left\{\hat{\lambda}_{j0}(x_j)\wedge\hat{\lambda}_{j1}(x_j)-\lambda_{j0}(x_j)\wedge\lambda_{j1}(x_j)\right\}+\inf_{x_j\in[0,1]}\left\{\lambda_{j0}(x_j)\wedge\lambda_{j1}(x_j)\right\}} \\
&\leq \frac{1}{-\sup_{x_j\in[0,1]}\left|\hat{\lambda}_{j0}(x_j)\wedge\hat{\lambda}_{j1}(x_j)-\lambda_{j0}(x_j)\wedge\lambda_{j1}(x_j)\right|+\inf_{x_j\in[0,1]}\left\{\lambda_{j0}(x_j)\wedge\lambda_{j1}(x_j)\right\}},
\end{aligned}$$

where $\hat{\lambda}_{jk}(x_j)$ and $\lambda_{jk}(x_j)$ $(k=0,1)$ are the eigenvalues of $\hat{\mathbf{Q}}_j(x_j)$ and $\mathbf{Q}_j(x_j)$, respectively. It follows that $\sup_{x_j\in[0,1]}\left\{\hat{\lambda}_{j0}(x_j)\wedge\hat{\lambda}_{j1}(x_j)-\lambda_{j0}(x_j)\wedge\lambda_{j1}(x_j)\right\}=o_p(1)$ from the continuity of eigenvalues, and that $\inf_{x_j\in[0,1]}\left\{\lambda_{j0}(x_j)\wedge\lambda_{j1}(x_j)\right\}\geq B$ for some constant $B>0$, since $\mathbf{Q}_j(x_j)$ is a positive definite matrix by assumptions (B1) and (B3). These imply $\sup_{x_j\in[0,1]}\|\hat{\mathbf{Q}}_j(x_j)^{-1}\|_2=O_p(1)$. Next, note that

$$\left\|\frac{1}{n}\sum_{i=1}^{n}(1,(X_{i,j}-x_j)/h_j)^{\top}K_{h_j}(x_j,X_{i,j})Z_{i,j}\right\|\leq\frac{\sqrt{2}}{n}\sum_{i=1}^{n}K_{h_j}(x_j,X_{i,j})|Z_{i,j}|\equiv a_n(x_j).$$

Then, we have

$$\sup_{x_j\in[0,1]}|a_n(x_j)| \leq \sup_{x_j\in[0,1]}|a_n(x_j)-E(a_n(x_j))|+\sup_{x_j\in[0,1]}|E(a_n(x_j))|. \tag{13}$$

It follows that $\sup_{x_j\in[0,1]}|a_n(x_j)-E(a_n(x_j))|=o_p(1)$ from Lemma A.1. For the second factor on the right hand side of (13), observe that,

$$\begin{aligned}
&E(|Z_{1,j}|K_{h_j}(x_j,X_{1,j})) \\
&\leq \sup_{u\in[0,1]}E(|Z_{1,j}||X_{1,j}=u)\cdot E\left(K_{h_j}(x_j,X_{1,j})\right)<\infty,
\end{aligned}$$

uniformly in $x_j\in[0,1]$ by assumptions (B2) and (B4). Therefore, we can conclude that $\sup_{x_j\in[0,1]}|a_n(x_j)|=O_p(1)$. This completes the proof. □

Proof of Theorem 1. First note that, by using similar arguments as in Lee et al. [2012a],

$$\|\hat{\mathbf{r}}^{\hat{G}}\|_{\mathbf{M}}\leq C\sum_{j=1}^{d}\left[\int\tilde{\alpha}_{j0}^{\hat{G}}(x_j)^2q_j(x_j)dx_j+\mu_2(K)\cdot\int\tilde{\alpha}_{j1}^{\hat{G}}(x_j)^2q_j(x_j)dx_j\right]^{\frac{1}{2}},$$

for some constant $C>0$ with probability tending to one, where $\tilde{\alpha}_{jk}^{\hat{G}}(x_j),\ k=0,1$, is the estimated version of $\tilde{\alpha}_{jk}^{G}(x_j)$ with G being replaced by $\hat{G}$. We only prove that $\int\tilde{\alpha}_{j0}^{\hat{G}}(x_j)^2q_j(x_j)dx_j<$

∞ with probability tending to one. The proof for $\int \tilde{\alpha}_{j1}^{\hat{G}}(x_j)^2 q_j(x_j)dx_j < \infty$ can be done similarly. For all $j = 1, \ldots, d$,

$$\int \tilde{\alpha}_{j0}^{\hat{G}}(x_j)^2 q_j(x_j)dx_j \quad \leq \quad 2\left(\int \tilde{\alpha}_{j0}^{G}(x_j)^2 q_j(x_j)dx_j + \int (\tilde{\alpha}_{j0}^{\hat{G}}(x_j) - \tilde{\alpha}_{j0}^{G}(x_j))^2 q_j(x_j)dx_j\right). \quad (14)$$

The fact that the first term on the right hand side of (14) is bounded with probability tending to one was established in Lee et al. [2012a]. For the second term, observe that

$$\sup_{x_j \in [0,1]} |\tilde{\alpha}_{j0}^{\hat{G}}(x_j) - \tilde{\alpha}_{j0}^{G}(x_j)| \leq \sup_{x_j \in [0,1]} \|\tilde{\boldsymbol{\alpha}}_j^{\hat{G}}(x_j) - \tilde{\boldsymbol{\alpha}}_j^{G}(x_j)\| = O_p\left(\sup_{t \leq \tau_0} |\hat{G}(t) - G(t)|\right) = o_p(1),$$

by Lemma 2. Therefore $\|\hat{\mathbf{r}}^{\hat{G}}\|_{\mathbf{M}} < \infty$ with probability tending to one. This completes the proof. $\square$

Proof of Lemma 3. Let $\hat{\boldsymbol{\alpha}}^{\hat{G}}(\mathbf{x}) = (\hat{\boldsymbol{\alpha}}_1^{\hat{G}}(x_1)^\top, \ldots, \hat{\boldsymbol{\alpha}}_d^{\hat{G}}(x_d)^\top)^\top$, $\hat{\boldsymbol{\alpha}}_j^{\hat{G}}(x_j) = (\hat{\alpha}_{j0}^{\hat{G}}(x_j), \hat{\alpha}_{j1}^{\hat{G}}(x_j))^\top$, $\hat{\boldsymbol{\alpha}}^{G}(\mathbf{x}) = (\hat{\boldsymbol{\alpha}}_1^{G}(x_1)^\top, \ldots, \hat{\boldsymbol{\alpha}}_d^{G}(x_d)^\top)^\top$ and $\hat{\boldsymbol{\alpha}}_j^{G}(x_j) = (\hat{\alpha}_{j0}^{G}(x_j), \hat{\alpha}_{j1}^{G}(x_j))^\top$. We will prove that, for all j, $\hat{\boldsymbol{\alpha}}_j^{\hat{G}}(x_j) - \hat{\boldsymbol{\alpha}}_j^{G}(x_j) = O_p(\sup_{t \leq \tau_0} |\hat{G}(t) - G(t)|) + o_p(n^{-2/5})$ for any $x_j \in [0,1]$. For this, we need to define some functions. First, let

$$\begin{aligned}
\tilde{\boldsymbol{\alpha}}_j^{\hat{G},A}(x_j) &= \hat{\mathbf{Q}}_j(x_j)^{-1} \frac{1}{n} \sum_{i=1}^{n} \begin{pmatrix} 1 \\ \frac{X_{i,j} - x_j}{h_j} \end{pmatrix} Z_{i,j}(Y_i^{\hat{G}} - m(\mathbf{X}_i, \mathbf{Z}_i)) K_{h_j}(x_j, X_{i,j}), \\
\tilde{\boldsymbol{\alpha}}_j^{G,A}(x_j) &= \hat{\mathbf{Q}}_j(x_j)^{-1} \frac{1}{n} \sum_{i=1}^{n} \begin{pmatrix} 1 \\ \frac{X_{i,j} - x_j}{h_j} \end{pmatrix} Z_{i,j}(Y_i^{G} - m(\mathbf{X}_i, \mathbf{Z}_i)) K_{h_j}(x_j, X_{i,j}),
\end{aligned}$$

and let $\hat{\boldsymbol{\alpha}}_j^{\hat{G},A}(x_j)$ and $\hat{\boldsymbol{\alpha}}_j^{G,A}(x_j)$ be the jth component vectors of $\hat{\boldsymbol{\alpha}}^{\hat{G},A}(\mathbf{x})$ and $\hat{\boldsymbol{\alpha}}^{G,A}(\mathbf{x})$ respectively. Here $\hat{\boldsymbol{\alpha}}^{\hat{G},A}$ and $\hat{\boldsymbol{\alpha}}^{G,A}$ are the projections of $\tilde{\boldsymbol{\alpha}}^{\hat{G},A}$ and $\tilde{\boldsymbol{\alpha}}^{G,A}$ onto $\mathcal{H}(\hat{\mathbf{M}})$, respectively, where

$$\begin{aligned}
\tilde{\boldsymbol{\alpha}}^{\hat{G},A}(\mathbf{x}) &= \hat{\mathbf{M}}(\mathbf{x})^{-1} \frac{1}{n} \sum_{i=1}^{n} \mathbf{v}(\mathbf{X}_i, \mathbf{Z}_i; \mathbf{x})(Y_i^{\hat{G}} - m(\mathbf{X}_i, \mathbf{Z}_i)) K_{\mathbf{h}}(\mathbf{x}, \mathbf{X}_i), \\
\tilde{\boldsymbol{\alpha}}^{G,A}(\mathbf{x}) &= \hat{\mathbf{M}}(\mathbf{x})^{-1} \frac{1}{n} \sum_{i=1}^{n} \mathbf{v}(\mathbf{X}_i, \mathbf{Z}_i; \mathbf{x})(Y_i^{G} - m(\mathbf{X}_i, \mathbf{Z}_i)) K_{\mathbf{h}}(\mathbf{x}, \mathbf{X}_i).
\end{aligned}$$

Now, note that

$$\begin{aligned}
\hat{\boldsymbol{\alpha}}_j^{\hat{G}}(x_j) - \hat{\boldsymbol{\alpha}}_j^{G}(x_j) &= \hat{\boldsymbol{\alpha}}_j^{\hat{G},A}(x_j) - \hat{\boldsymbol{\alpha}}_j^{G,A}(x_j) \\
&= (\hat{\boldsymbol{\alpha}}_j^{\hat{G},A}(x_j) - \tilde{\boldsymbol{\alpha}}_j^{\hat{G},A}(x_j)) + (\tilde{\boldsymbol{\alpha}}_j^{\hat{G},A}(x_j) - \tilde{\boldsymbol{\alpha}}_j^{G,A}(x_j)) + (\tilde{\boldsymbol{\alpha}}_j^{G,A}(x_j) - \hat{\boldsymbol{\alpha}}_j^{G,A}(x_j)) \\
&= R_{1n} + R_{2n} + R_{3n}.
\end{aligned}$$

We can see that for any $x_j \in [0,1]$, $R_{2n} = \tilde{\boldsymbol{\alpha}}_j^{\hat{G}}(x_j) - \tilde{\boldsymbol{\alpha}}_j^{G}(x_j) = O_p(\sup_{t \le \tau_0} |\hat{G}(t) - G(t)|)$ by Lemma 2, and that $R_{3n} = o_p(n^{-2/5})$ by the same arguments as in Lee et al. [2012a]. As for R_{1n}, following the lines of the proof of Theorem 2 in Lee et al. [2012a], it suffices to show that

$$\sup_{\mathbf{x} \in [0,1]^d} \left\| \sum_{l=1}^{\infty} \hat{U}^l \hat{\mathbf{r}}^{\hat{G},A}(\mathbf{x}) \right\| = O_p \left(\sup_{t \le \tau_0} |\hat{G}(t) - G(t)| \right) + o_p(n^{-2/5}),$$

where $\hat{\mathbf{r}}^{\hat{G},A} = (I - \hat{U})\tilde{\boldsymbol{\alpha}}^{\hat{G},A}$. This follows if we can show that

$$\sup_{x_k \in [0,1]} \left\| \int \hat{\mathbf{Q}}_k(x_k)^{-1} \hat{\mathbf{Q}}_{jk}(x_j, x_k) \tilde{\boldsymbol{\alpha}}_j^{\hat{G},A}(x_j) dx_j \right\| = O_p \left(\sup_{t \le \tau_0} |\hat{G}(t) - G(t)| \right) + o_p(n^{-2/5}),$$

for all $k = 1, \ldots, d$. By the triangle inequality,

$$\begin{aligned} &\sup_{x_k \in [0,1]} \left\| \int \hat{\mathbf{Q}}_k(x_k)^{-1} \hat{\mathbf{Q}}_{jk}(x_j, x_k) \tilde{\boldsymbol{\alpha}}_j^{\hat{G},A}(x_j) dx_j \right\| \\ &\le \sup_{x_k \in [0,1]} \left\| \int \hat{\mathbf{Q}}_k(x_k)^{-1} \hat{\mathbf{Q}}_{jk}(x_j, x_k) \tilde{\boldsymbol{\alpha}}_j^{G,A}(x_j) dx_j \right\| \\ &\quad + \sup_{x_k \in [0,1]} \left\| \int \hat{\mathbf{Q}}_k(x_k)^{-1} \hat{\mathbf{Q}}_{jk}(x_j, x_k) (\tilde{\boldsymbol{\alpha}}_j^{\hat{G},A}(x_j) - \tilde{\boldsymbol{\alpha}}_j^{G,A}(x_j)) dx_j \right\| \\ &= R'_{1n} + R'_{2n}. \end{aligned}$$

From Lee et al. [2012a], it follows that $R'_{1n} = o_p(n^{-2/5})$. As for R'_{2n}, note that

$$\begin{aligned} R'_{2n} \le \; & \sup_{x_j \in [0,1]} \left\| \tilde{\boldsymbol{\alpha}}_j^{\hat{G},A}(x_j) - \tilde{\boldsymbol{\alpha}}_j^{G,A}(x_j) \right\| \\ \times \; & \sup_{x_k \in [0,1]} \int \left(\|(\hat{\mathbf{Q}}_k(x_k)^{-1})_1\| + \|(\hat{\mathbf{Q}}_k(x_k)^{-1})_2\| \right) \left(\|(\hat{\mathbf{Q}}_{jk}(x_j, x_k))_1\| + \|(\hat{\mathbf{Q}}_{jk}(x_j, x_k))_2\| \right) dx_j \end{aligned} \tag{15}$$

where $(\mathbf{A})_l$ denotes the lth row of a matrix $\mathbf{A}$. The first factor on the right hand side of (15) is $O_p(\sup_{t \le \tau_0} |\hat{G}(t) - G(t)|)$ by Lemma 2. Now, it suffices to show that the second factor is bounded in probability. For this, note that $\hat{\mathbf{Q}}_{jk}(x_j, x_k)$ converge uniformly in probability to $\mathbf{Q}_{jk}(x_j, x_k)$ from Lemma A.1, where

$$\mathbf{Q}_{jk}(x_j, x_k) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} E(Z_j Z_k | X_j = x_j, X_k = x_k) p_{jk}(x_j, x_k),$$

and where p_{jk} is the joint density of X_j and X_k. Since $E(Z_j Z_k | X_j = x_j, X_k = x_k) p_{jk}(x_j, x_k)$ is bounded on $[0,1]^2$ by assumptions (B2) and (B3), $\|(\hat{\mathbf{Q}}_{jk}(x_j, x_k))_1\| + \|(\hat{\mathbf{Q}}_{jk}(x_j, x_k))_2\| = O_p(1)$ uniformly in $x_j, x_k \in [0,1]^2$. It follows that $\|(\hat{\mathbf{Q}}_k(x_k)^{-1})_1\| + \|(\hat{\mathbf{Q}}_k(x_k)^{-1})_2\| = O_p(1)$ uniformly in $x_j \in [0,1]$ from the fact that $\|(\hat{\mathbf{Q}}_k(x_k)^{-1})_l\| \le \|\hat{\mathbf{Q}}_k(x_k)^{-1}\|_2$ for $l = 1, 2$, where $\|\cdot\|_2$ denotes the spectral norm, and this is $O_p(1)$, as is shown in the proof of Lemma 2. □

Acknowledgements

S. J. Yang, A. El Ghouch and I. Van Keilegom all acknowledge financial support from the contract 'Projet d'Actions de Recherche Concertées' (ARC) 11/16-039 of the 'Communauté Française de Belgique', granted by the 'Académie Universitaire Louvain', and from IAP research network P7/06 of the Belgian Government (Belgian Science Policy). S. J. Yang and I. Van Keilegom also acknowledge financial support from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013)/ERC Grant agreement No.203650.

References

R. Beran. Nonparametric regression with randomly censored survival data. *Technical Report*, 1981.

F. Bravo. Varying coefficient partially linear models with randomly censored data. *Working Paper*, 2012.

J. De Uña Álvarez and J. Roca Pardiñas. Additive models in censored regression. *Computational Statistics & Data Analysis*, 53(9):3490–3501, 2009.

A. El Ghouch and I. Van Keilegom. Non-parametric regression with dependent censored data. *Scandinavian Journal of Statistics*, 35(2):228–247, 2008.

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

J. Fan and I. Gijbels. *Local polynomial modelling and its applications*. Chapman and Hall, New York, 1996.

J. Fan and T. Huang. Profile likelihood inferences on semiparametric varying-coefficient partially linear models. *Bernoulli*, 11(6):1031–1057, 2005.

T. Hastie and R. Tibshirani. Varying-coefficient models. *Journal of the Royal Statistical Society-Series B*, 55(4):757–796, 1993.

H. Koul, V. Susarla, and J. Van Ryzin. Regression analysis with randomly right-censored data. *The Annals of Statistics*, 9(6):1276–1288, 1981.

T. L. Lai, Z. Ying, and Z. Zheng. Asymptotic normality of a class of adaptive statistics with applications to synthetic data methods for censored regression. *Journal of Multivariate Analysis*, 52(2):259–279, 1995.

Y. K. Lee, E. Mammen, and B. U. Park. Projection-type estimation for varying coefficient regression models. *Bernoulli*, 18(1):177–205, 2012a.

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

S. Leurgans. Linear models, random censoring and synthetic data. *Biometrika*, 74(2):301–309, 1987.

S.-H. Lo and K. Singh. The product-limit estimator and the bootstrap: some asymptotic representations. *Probability Theory and Related Fields*, 71(3):455–465, 1986.

O. Lopez. Single-index regression models with right-censored responses. *Journal of Statistical Planning and Inference*, 139(3):1082–1097, 2009.

O. Lopez and V. Patilea. Nonparametric lack-of-fit tests for parametric mean-regression models with censored data. *Journal of Multivariate Analysis*, 100(1):210–230, 2009.

O. Lopez, V. Patilea, and I. Van Keilegom. Single index regression models in the presence of censoring depending on the covariates. *Bernoulli*, 19(3):721–747, 2013.

Y. P. Mack and B. W. Silverman. Weak and strong uniform consistency of kernel regression estimates. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 61(3):405–415, 1982.

E. Mammen, O. Linton, and J. Nielsen. The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. *The Annals of Statistics*, 27(5):1443–1490, 1999.

J. D. Opsomer. Asymptotic properties of backfitting estimators. *Journal of Multivariate Analysis*, 73(2):166–179, 2000.

C. Srinivasan and M. Zhou. Linear regression with censoring. *Journal of Multivariate Analysis*, 49(2):179–201, 1994.

M. Talamakrouni, A. El Ghouch, and I. Van Keilegom. Guided censored regression. *Scandinavian Journal of Statistics*, (under revision), 2012.

L. Yang, B. U. Park, L. Xue, and W. Härdle. Estimation and testing for varying coefficients in additive models with marginal integration. *Journal of the American Statistical Association*, 101(475):1212–1227, 2006.

Z. Zheng. A class of estimators of the parameters in linear regression with censored data. *Acta Mathematicae Applicatae Sinica*, 3(3):231–241, 1987.

M. Zhou. Asymptotic normality of the 'synthetic data' regression estimator for censored survival data. *The Annals of Statistics*, 20(2):1002–1021, 1992.