Weighted Local Linear Smoothing Method for Randomly Right Censored Varying Coefficient Models

Tizheng Li

School of Science,
Xi'an University of Architecture and Technology,
No. 13, Yanta Road,
Bielin district, Xi'an, China.
E-mail address: `tizhengli@xauat.edu.cn`

Abstract. In this paper, we propose a new estimation method for randomly right censored varying coefficient models and establish the asymptotic normality of the resulting estimator. Compared to the existing estimation method, the proposed estimation method is computationally simple and practically feasible since it requires no transformation of data. Simulation studies are conducted to assess the finite sample performance of the proposed estimation method against that of the existing estimation method. The simulation results demonstrate that the proposed estimation method performs much better than the existing estimation method.

1. Introduction

In recent years, there has been an upsurge of interest and effort in nonparametric regression analysis because of its flexibility and adaptability in modeling a regression relationship. However, a nonparametric regression model may fail to incorporate some prior information and the resulting estimator of the regression function tends to incur greater variance. The worse is the so-called "curse of dimensionality", which makes the commonly used nonparametric smoothing methods practically impotent when the number of the covariates is large. To overcome the "curse of dimensionality", many alternatives of the nonparametric models have been proposed in the literature. Some typical examples include additive models Hastie and Tibshirani (1990), low-dimensional interaction models (Friedman, 1991;

Gu and Wahba, 1993), single-index models (Härdle and Stoker, 1989), partially linear models (Engle et al., 1986), varying coefficient models (Cleveland et al., 1992; Hastie and Tibshirani, 1993), and their hybrids (Carroll et al., 1997; Wong et al., 2008; Zhang et al., 2002). One of the most important alternatives is varying coefficient models, which extend the classical linear models by allowing the regression coefficients to depend on certain covariates. This type of models has been very useful statistical tool to explore the dynamic pattern in many scientific areas, such as economics, finance, politics, epidemiology, medical science, ecology and so on. In general, a varying coefficient model assumes that the relationship between the response variable Y and the covariates $X_1, \cdots, X_p$ and U follows

$$Y_i = \mathbf{X}_i^{\mathrm{T}} \mathbf{a}(U_i) + \varepsilon_i, \quad i = 1, \cdots, n, \tag{1.1}$$

where Y_i, $\mathbf{X}_i = (X_{i1}, \cdots, X_{ip})^{\mathrm{T}}$ and U_i are, respectively, the observations of the response variable Y and the covariates $X_1, \cdots, X_p$ and U, $\mathbf{a}(\cdot) = (a_1(\cdot), \cdots, a_p(\cdot))^{\mathrm{T}}$ is a vector of unknown coefficient functions, and ε_i $(i = 1, \cdots, n)$ are the random error terms with $\mathrm{E}(\varepsilon_i | U_i, \mathbf{X}_i) = 0$ and $\mathrm{Var}(\varepsilon_i | U_i, \mathbf{X}_i) = \sigma^2(U_i, \mathbf{X}_i)$ $(i = 1, \cdots, n)$. This model is much more flexible than the classical linear model, since each coefficient function is modeled 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 function of the covariates $X_1, \cdots, X_p$. If all coefficient functions are constant functions, the model reduces to the classical linear model.

The varying coefficient model was originally introduced by Cleveland et al. (1992) and was further developed by Hastie and Tibshirani (1993). Due to its easy interpretation and its flexibility to explore the dynamic pattern of a regression relationship, model (1.1) has attracted a great deal of attention over the past two decades. When its response variable Y is fully observed, model (1.1) has been widely studied in estimation (Cai and Xu, 2009; Cai et al., 2000; Fan and Zhang, 1999; Hastie and Tibshirani, 1993; Huang et al., 2004; Lu et al., 2008; Qu and Li, 2006; Wang et al., 2009; Xue and Zhu, 2007), hypothesis testing (Cai et al., 2000; Fan and Zhang, 2000; Fan et al., 2001; Xu and Zhu, 2008; Zhang and Peng, 2010), and variable selection (Daye et al., 2012; Fan et al., 2014; Liu et al., 2014; Wang and Xia, 2009; Wang et al., 2008; Xue and Qu, 2012; Zhao and Xue, 2011).

However, in many applications of regression analysis, especially in biomedical studies, the response variable Y cannot be completely observed due to possible censoring, for instance, withdrawal of patients from a study or death from a cause unrelated to the specific disease of being studied. Among the censoring mechanisms, random right censoring is of the widest application backgrounds. Let C be the censoring variable. For randomly right censored data, we observe $\{(U_i, \mathbf{X}_i, Z_i, \delta_i)\}_{i=1}^n$, a random sample of $(U, \mathbf{X}, Z, \delta)$, where

$$Z = \min(Y, C) \text{ and } \delta = I(Y \leq C), \tag{1.2}$$

with $I(\cdot)$ being the indicator function. Throughout this paper, we assume that the censoring variable C is independent of the response variable Y. Furthermore, for ease of the technical proofs, we assume that C and Y are both the nonnegative random variables from now on.

Compared to the study on randomly right censored nonparametric regression models (Cai, 2003; Fan and Gijbels, 1994; Kim and Truong, 1998; Lopez and Patilea, 2009; Pardo-Fernández et al., 2007), less attention has been paid to randomly right

censored varying coefficient models. Recently, Luo et al. (2006) proposed a synthetic data method for randomly right censored varying coefficient models, in which the transformation technique proposed by Fan and Gijbels (1994) was first used to transform the censored data and then the local polynomial smoothing method was employed to fit the varying coefficient model with the transformed data. However, this estimation method needs to choose two parameters, one is the tuning parameter in the data transformation, which is used to reduce the variability of the transformed data, and the other is the bandwidth in the local polynomial smoothing, thus it is computationally demanding for applied researchers. More importantly, when the censoring distribution depends on the covariate vector $(U, \mathbf{X})$, the method needs to estimate the conditional censoring distribution function via the local Kaplan-Meier method in which another bandwidth is needed to be chosen, hence it is practically feasible only when the dimension of $(U, \mathbf{X})$ is small because of the well-known "curse of dimensionality".

The purpose of this paper is to propose a computationally simple and practically feasible estimation method for randomly right censored varying coefficient models. The proposed estimation method constructs a weighted local least squares function based on the idea of weighted least squares, which may be regarded as a generalization of the weighted least squares method originally introduced by Zhou (1992) for randomly right censored linear regression models and further extended by Stute (1999), Cai (2003) and Lopez (2009) for randomly right censored nonlinear regression models, randomly right censored nonparametric regression models and randomly right censored single-index models, respectively. Specifically, the weighted local least squares function is constructed by first expressing the kernel weighted local least squares function for the uncensored data $\{(U_i, \mathbf{X}_i, Y_i)\}_{i=1}^n$ as an integral with respect to the empirical distribution function of $\{(U_i, \mathbf{X}_i, Y_i)\}_{i=1}^n$ and then replacing the empirical distribution function for the uncensored data by its counterpart for the censored data $\{(U_i, \mathbf{X}_i, Z_i, \delta_i)\}_{i=1}^n$ (that is, Kaplan-Meier estimator). In such a way, no transformation is needed so that it avoids the selection of the tuning parameter and the estimation of the unconditional or conditional censoring distribution function. Therefore, the proposed estimation method not only has the advantage of computational simplicity and practical feasibility, but also is expected to produce more accurate coefficient estimates, which is vital to the application of the varying coefficient model because the coefficient estimates are generally taken as the main evidence to explore the dynamic pattern of the underlying regression relationship. We conduct simulation studies to examine the finite sample performance of the proposed estimation method and to make an empirical comparison with the existing estimation method in terms of the mean squared error of coefficient estimates. The simulation results show that the proposed estimation method indeed has much better finite sample performance than the existing one. In addition to Luo et al. (2006), another related paper by Yang et al. (2014) studied the problem of estimating the coefficient functions in randomly right censored varying coefficient models where different coefficient functions have different one-dimensional smoothing variable, which is more flexible than the model considered in this paper, in which the authors proposed an estimation method based on the synthetic data obtained by the unbiased transformation given by Koul et al. (1981) and the smooth back-fitting technique and studied the asymptotic normality of the resulting estimators of the coefficient functions. Although both the proposed

estimation method and the estimation method of Yang et al. (2014) employed the same assumption on the censoring mechanism (see Remark 2.3 below), the starting point of constructing estimator of the coefficient functions is completely different.

The remainder of this paper is organized as follows. In Section 2, we describe in detail the proposed estimation method for the randomly right censored varying coefficient model, establish the asymptotic normality of the resulting estimator of the coefficient function vector $\mathbf{a}(u)$, and discuss the issue of bandwidth selection. In the same section, a comparison with the synthetic data method of Luo et al. (2006) is made and an extension to randomly right censored varying coefficient models with multivariate smoothing variable is included. Section 3 presents numerical comparison. Conclusion remark is given in Section 4. Technical proofs are given in Section 5.

2. Weighted local linear smoothing method

2.1. *Weighted local linear estimator.* To better illustrate the idea of the proposed estimation method, we first consider the case where the response variable Y is fully observed. In this case, the local linear smoothing method (Fan and Gijbels, 1996) is commonly used to estimate the coefficient function vector $\mathbf{a}(\cdot)$, although other smoothing methods such as the Nadaraya-Watson kernel method and the spline methods are applicable. The main reason for preferring the local linear smoothing method is because it possesses many attractive properties such as high statistical efficiency in an asymptotic minimax sense, design adaptation, and automatic boundary corrections (for details, see Fan and Gijbels, 1996).

Assume that all the coefficient functions $a_1(\cdot), \cdots, a_p(\cdot)$ have continuous second order derivatives. Then for any given u_0 in the domain of the covariate U, it follows from the Taylor's expansion that

$$a_j(u) \approx a_j(u_0) + a_j'(u_0)(u - u_0), \ j = 1, \cdots, p,$$

in a neighborhood of u_0. The local linear smoothing method finds $\mathbf{a}(u_0)$ and $\mathbf{a}'(u_0) = (a_1'(u_0), \cdots, a_p'(u_0))^{\mathrm{T}}$ to minimize the following kernel weighted local least squares function

$$\frac{1}{n} \sum_{i=1}^{n} \left[Y_i - \mathbf{X}_i^{\mathrm{T}} \mathbf{a}(u_0) - (U_i - u_0) \mathbf{X}_i^{\mathrm{T}} \mathbf{a}'(u_0) \right]^2 K_h(U_i - u_0), \qquad (2.1)$$

where $K_h(\cdot) = K(\cdot/h)/h$ with $K(\cdot)$ being a kernel function and h being the bandwidth.

Let $\mathbf{Y} = (Y_1, \cdots, Y_n)^{\mathrm{T}}$, $\mathbf{M} = (\mathbf{X}_1^{\mathrm{T}} \mathbf{a}(U_1), \cdots, \mathbf{X}_n^{\mathrm{T}} \mathbf{a}(U_n))^{\mathrm{T}}$, $\boldsymbol{\Psi}(u_0) = (\mathbf{a}(u_0)^{\mathrm{T}}, h\mathbf{a}'(u_0)^{\mathrm{T}})^{\mathrm{T}}$, $\mathbf{W}(u_0) = \mathrm{diag}(K_h(U_1 - u_0), \cdots, K_h(U_n - u_0))$, and

$$\mathbf{X}(u_0) = \begin{pmatrix} \mathbf{X}_1^{\mathrm{T}} & h^{-1}(U_1 - u_0)\mathbf{X}_1^{\mathrm{T}} \\ \vdots & \vdots \\ \mathbf{X}_n^{\mathrm{T}} & h^{-1}(U_n - u_0)\mathbf{X}_n^{\mathrm{T}} \end{pmatrix}.$$

With the above notations, the solution of the weighted least squares problem (2.1), that is, the estimator of $\boldsymbol{\Psi}(u_0)$, can be expressed as

$$\widetilde{\boldsymbol{\Psi}}(u_0) = [\mathbf{X}(u_0)^{\mathrm{T}} \mathbf{W}(u_0) \mathbf{X}(u_0)]^{-1} \mathbf{X}(u_0)^{\mathrm{T}} \mathbf{W}(u_0) \mathbf{Y}. \qquad (2.2)$$

Consequently, the estimator of the coefficient function vector $\mathbf{a}(u)$ at u_0 is

$$\widetilde{\mathbf{a}}(u_0) = (\mathbf{I}_p, \mathbf{0}_{p \times p})[\mathbf{X}(u_0)^{\mathrm{T}}\mathbf{W}(u_0)\mathbf{X}(u_0)]^{-1}\mathbf{X}(u_0)^{\mathrm{T}}\mathbf{W}(u_0)\mathbf{Y}, \qquad (2.3)$$

where $\mathbf{0}_{p \times p}$ is the $p \times p$ zero matrix.

Taking u_0 in (2.3) to be $U_1, \cdots, U_n$, respectively, we can obtain $\widetilde{\mathbf{M}}$, the fitted vector of $\mathbf{M}$, as

$$\widetilde{\mathbf{M}} = \begin{pmatrix} \mathbf{X}_1^{\mathrm{T}}\widetilde{\mathbf{a}}(U_1) \\ \vdots \\ \mathbf{X}_n^{\mathrm{T}}\widetilde{\mathbf{a}}(U_n) \end{pmatrix} = \widetilde{\mathbf{S}}\mathbf{Y}, \qquad (2.4)$$

where

$$\widetilde{\mathbf{S}} = \begin{pmatrix} (\mathbf{X}_1^{\mathrm{T}}, \mathbf{0}_{1 \times p}) \left[\mathbf{X}(U_1)^{\mathrm{T}}\mathbf{W}(U_1)\mathbf{X}(U_1)\right]^{-1} \mathbf{X}(U_1)^{\mathrm{T}}\mathbf{W}(U_1) \\ \vdots \\ (\mathbf{X}_n^{\mathrm{T}}, \mathbf{0}_{1 \times p}) \left[\mathbf{X}(U_n)^{\mathrm{T}}\mathbf{W}(U_n)\mathbf{X}(U_n)\right]^{-1} \mathbf{X}(U_n)^{\mathrm{T}}\mathbf{W}(U_n) \end{pmatrix}.$$

It is clear that the kernel weighted local least squares function (2.1) can be expressed as the following form

$$\int \left[y - \mathbf{x}^{\mathrm{T}}\mathbf{a}(u_0) - (u - u_0)\mathbf{x}^{\mathrm{T}}\mathbf{a}'(u_0)\right]^2 K_h(u - u_0)\,\mathrm{d}F_n(u, \mathbf{x}, y), \qquad (2.5)$$

where $F_n(u, \mathbf{x}, y)$ is the empirical distribution function of $\{(U_i, \mathbf{X}_i, Y_i)\}_{i=1}^n$, a consistent estimator of the joint distribution function $F(u, \mathbf{x}, y) = P(U \leq u, \mathbf{X} \leq \mathbf{x}, Y \leq y)$ of $(U, \mathbf{X}, Y)$.

However, in the situation of random right censoring, the local linear estimator of the coefficient function vector $\mathbf{a}(\cdot)$ cannot be obtained through minimizing (2.5) since $F_n(u, \mathbf{x}, y)$ relies on the uncensored observations $\{(U_i, \mathbf{X}_i, Y_i)\}_{i=1}^n$. In this case, a natural way to proceed is to replace $F_n(u, \mathbf{x}, y)$ with other consistent estimator of $F(u, \mathbf{x}, y)$. To this purpose, let

$$W_{in} = \frac{\delta_{(i)}}{n - i + 1} \prod_{j=1}^{i-1} \left(\frac{n - j}{n - j + 1}\right)^{\delta_{(j)}}, \qquad (2.6)$$

where $Z_{(1)} \leq \cdots \leq Z_{(n)}$ are the order statistics of $Z_1, \cdots, Z_n$ and $\delta_{(i)}$ is the δ associated with $Z_{(i)}$. Following Stute and Wang (1993), the Kaplan-Meier estimator $\widehat{F}_n(y)$ of the distribution function $F(y) = P(Y \leq y)$ of the response variable Y can be expressed as $\widehat{F}_n(y) = \sum_{i=1}^n W_{in} I(Z_{(i)} \leq y)$. In the presence of the covariates, Stute (1993) extended the Kaplan-Meier estimator to include the covariates and proposed an estimator for the joint distribution function $F(u, \mathbf{x}, y)$, namely,

$$\widehat{F}_n(u, \mathbf{x}, y) = \sum_{i=1}^n W_{in} I(U_{(i)} \leq u, \mathbf{X}_{(i)} \leq \mathbf{x}, Z_{(i)} \leq y),$$

where $U_{(i)}$ and $\mathbf{X}_{(i)}$ are the U and $\mathbf{X}$ associated with $Z_{(i)}$, respectively. It follows from Corollary 1.5 in Stute (1993) that $\widehat{F}_n(u, \mathbf{x}, y)$ is a consistent estimator of $F(u, \mathbf{x}, y)$. Like that in Cai (2003), we replace $F_n(u, \mathbf{x}, y)$ in (2.5) with $\widehat{F}_n(u, \mathbf{x}, y)$

to obtain the following weighted local least squares function

$$\int \left[y - \mathbf{x}^{\mathrm{T}} \mathbf{a}(u_0) - (u - u_0) \mathbf{x}^{\mathrm{T}} \mathbf{a}'(u_0) \right]^2 K_h(u - u_0) \, \mathrm{d}\widehat{F}_n(u, \mathbf{x}, y)$$

$$= \sum_{i=1}^{n} W_{in}^* \left[Z_i - \mathbf{X}_i^{\mathrm{T}} \mathbf{a}(u_0) - (U_i - u_0) \mathbf{X}_i^{\mathrm{T}} \mathbf{a}'(u_0) \right]^2 K_h(U_i - u_0), \qquad (2.7)$$

where W_{in}^* is the W_{in}-value associated with the natural order of Z_i. By minimizing (2.7) with respect to $\mathbf{a}(u_0)$ and $\mathbf{a}'(u_0)$, we obtain the weighted local linear estimators of $\boldsymbol{\Psi}(u_0)$ and $\mathbf{a}(u_0)$ as

$$\widehat{\boldsymbol{\Psi}}(u_0) = [\mathbf{X}(u_0)^{\mathrm{T}} \mathbf{W}^* \mathbf{W}(u_0) \mathbf{X}(u_0)]^{-1} \mathbf{X}(u_0)^{\mathrm{T}} \mathbf{W}^* \mathbf{W}(u_0) \mathbf{Z} \qquad (2.8)$$

and

$$\widehat{\mathbf{a}}(u_0) = (\mathbf{I}_p, \mathbf{0}_{p \times p}) [\mathbf{X}(u_0)^{\mathrm{T}} \mathbf{W}^* \mathbf{W}(u_0) \mathbf{X}(u_0)]^{-1} \mathbf{X}(u_0)^{\mathrm{T}} \mathbf{W}^* \mathbf{W}(u_0) \mathbf{Z}, \qquad (2.9)$$

where $\mathbf{W}^* = \mathrm{diag}(W_{1n}^*, \cdots, W_{nn}^*)$ and $\mathbf{Z} = (Z_1, \cdots, Z_n)^{\mathrm{T}}$.

Taking u_0 in (2.9) to be $U_1, \cdots, U_n$, respectively, we can obtain $\widehat{\mathbf{M}}$, the fitted vector of $\mathbf{M}$, as

$$\widehat{\mathbf{M}} = \begin{pmatrix} \mathbf{X}_1^{\mathrm{T}} \widehat{\mathbf{a}}(U_1) \\ \vdots \\ \mathbf{X}_n^{\mathrm{T}} \widehat{\mathbf{a}}(U_n) \end{pmatrix} = \widehat{\mathbf{S}} \mathbf{Z}, \qquad (2.10)$$

where

$$\widehat{\mathbf{S}} = \begin{pmatrix} (\mathbf{X}_1^{\mathrm{T}}, \mathbf{0}_{1 \times p}) \left[\mathbf{X}(U_1)^{\mathrm{T}} \mathbf{W}^* \mathbf{W}(U_1) \mathbf{X}(U_1) \right]^{-1} \mathbf{X}(U_1)^{\mathrm{T}} \mathbf{W}^* \mathbf{W}(U_1) \\ \vdots \\ (\mathbf{X}_n^{\mathrm{T}}, \mathbf{0}_{1 \times p}) \left[\mathbf{X}(U_n)^{\mathrm{T}} \mathbf{W}^* \mathbf{W}(U_n) \mathbf{X}(U_n) \right]^{-1} \mathbf{X}(U_n)^{\mathrm{T}} \mathbf{W}^* \mathbf{W}(U_n) \end{pmatrix}.$$

Remark 2.1. There are two weights in the last expression of (2.7). The first weight W_{in}^* is used to compensate for the censoring while the second weight $K(\cdot)$ is used to control the amount of smoothing. When there is no censoring, $W_{in}^* = \frac{1}{n}$ and $Z_i = Y_i$, then (2.7) reduces to (2.1), which is the ordinary kernel weighted local least squares function.

2.2. *Asymptotic normality of the weighted local linear estimator.* To facilitate the presentation, we first introduce some notations. Denote by $G(\cdot)$ the distribution function of the censoring variable C. Define $\tau_F = \inf\{y | F(y) = 1\}$ and $\tau_G = \inf\{t | G(t) = 1\}$. Let $\mu_j = \int u^j K(u) \, \mathrm{d}u$, $\nu_j = \int u^j K^2(u) \, \mathrm{d}u$, $b_k(u, \mathbf{x}) = \mathrm{E}(Y^k \{1 - G(Y)\}^{-1} | U = u, \mathbf{X} = \mathbf{x})$, $\sigma^{*2}(u, \mathbf{x}) = \mathrm{Var}(Z_G | U = u, \mathbf{X} = \mathbf{x})$, $\boldsymbol{\Gamma}(u) = \mathrm{E}(\mathbf{X}\mathbf{X}^{\mathrm{T}} | U = u)$, and $\boldsymbol{\Gamma}_1(u) = \mathrm{E}(\mathbf{X}^{\otimes 4} | U = u)$, where $\mathbf{X}^{\otimes 4} = (\mathbf{X}\mathbf{X}^{\mathrm{T}})(\mathbf{X}\mathbf{X}^{\mathrm{T}})$ and $Z_G = \delta Z \{1 - G(Z-)\}^{-1}$ with $G(t-)$ being the left limit of $G(t)$ at t. Further, let $\boldsymbol{\Gamma}^*(u) = \mathrm{E}(\sigma^{*2}(U, \mathbf{X}) \mathbf{X}\mathbf{X}^{\mathrm{T}} | U = u)$, $\boldsymbol{\Gamma}_k^*(u) = \mathrm{E}(b_k(U, \mathbf{X}) \mathbf{X}\mathbf{X}^{\mathrm{T}} | U = u)$, $\boldsymbol{\Gamma}_{1k}^*(u) = \mathrm{E}(b_k(U, \mathbf{X}) \mathbf{X}^{\otimes 4} | U = u)$, and $\varepsilon_G = Z_G - \mathrm{E}(Z_G | U, \mathbf{X})$.

The following technical conditions are imposed to establish the asymptotic normality of the weighted local linear estimator, although some of them might not be the weakest possible.

  C1. The covariate U has a bounded support Ω and its density function $f(u)$ is continuous and is bounded away from 0 on its support Ω.

  C2. The $p \times p$ matrix $\boldsymbol{\Gamma}(u)$ is nonsingular for each $u \in \Omega$. The matrices $\boldsymbol{\Gamma}(u)$, $\boldsymbol{\Gamma}(u)^{-1}$, $\boldsymbol{\Gamma}_1(u)$, $\boldsymbol{\Gamma}^*(u)$, $\boldsymbol{\Gamma}_k^*(u)$, and $\boldsymbol{\Gamma}_{1k}^*(u)$ are all continuous.

C3. (i) $P(Y{\leq}C|U, \mathbf{X}, Y) = P(Y{\leq}C|Y)$,

    (ii) $\tau_F{\leq}\tau_G$.

C4. All of the coefficient functions $a_1(\cdot), \cdots, a_p(\cdot)$ have continuous second order derivatives.

C5. The kernel function $K(\cdot)$ is a symmetric density function and has a compact support, say $[-1, 1]$.

C6. When $n{\to}\infty$, we have $h{\to}0$, $nh{\to}\infty$.

C7. There exists some $\gamma > 0$, such that $\mathrm{E}(|\varepsilon_G|^{2+\gamma}|U = u, \mathbf{X} = \mathbf{x}) < \infty$ for all $\mathbf{x} \in \mathbb{R}^p$ and u in a neighborhood of u_0 and $\mathrm{E}(|\mathbf{X}|^{2+\gamma}|U = u)$ is continuous.

Remark 2.2. The conditions listed above are quite common in the literature. For example, the condition C3(i) is also imposed by Stute (1993, 1996, 1999) for the randomly right censored parametric regression models, Cai (2003) and Lopez and Patilea (2009) for the randomly right censored nonparametric regression models, and Lopez (2009) for the randomly right censored single-index models. As pointed out by Stute (1996), the condition C3(i), together with the independence of Y and C, guarantee that the joint distribution function $F(u, \mathbf{x}, y)$ of $(U, \mathbf{X}, Y)$ can be theoretically derived from that of $(U, \mathbf{X}, Z, \delta)$ and hence can be consistently estimated from the censored observations $\{(U_i, \mathbf{X}_i, Z_i, \delta_i)\}_{i=1}^n$. A major case for which the condition C3(i) holds is the case where the censoring variable C is independent of the covariate vector $(U, \mathbf{X})$. However, the condition C3(i) is flexible enough to allow for a dependence structure between C and $(U, \mathbf{X})$. A detailed discussion on situations in which the condition C3(i) holds without assuming the independence of C and $(U, \mathbf{X})$ can be found in Stute (1999). When $\tau_G < \tau_F$, in general, there is no way to consistently estimate the joint distribution function $F(u, \mathbf{x}, y)$ since relevant information about Y on (τ_G, τ_F) is always cut off due to random right censoring. The condition C3(ii) allows one to avoid this case. It is easy to show that the asymptotic optimal bandwidth $h_{\mathrm{opt}} = O(n^{-1/5})$ (see Remark 2.4 below) satisfies the condition C6. The condition C7 is imposed to establish the asymptotic normality of $\widehat{\boldsymbol{\Psi}}(u_0)$ and $\widehat{\mathbf{a}}(u_0)$ and it is widely used in the literature of the censored regression analysis (see, for example, Cai, 2003; Fan and Gijbels, 1994; Luo et al., 2006). Furthermore, as pointed out by Hall et al. (1999) and Cai and Ould-Saïd (2003), the requirement that the kernel function $K(\cdot)$ has compact support in the condition C5 can be removed at the expense of lengthier arguments used in the proof of Theorem 2.1. In particular, the Gaussian kernel is allowed.

Remark 2.3. In the literature of randomly right censored regression models, any estimation method relies on some conditions on the censoring mechanism. Along the line of research initiated by Stute (1993), this paper assumes that the response variable Y is independent of the censoring variable C and $P(Y{\leq}C|U, \mathbf{X}, Y) = P(Y{\leq}C|Y)$. As emphasized in Remark 2.2, this assumption assures that the joint distribution function $F(u, \mathbf{x}, y)$ of $(U, \mathbf{X}, Y)$ can be theoretically derived from that of $(U, \mathbf{X}, Z, \delta)$ and thus can be consistently estimated from the censored observations $\{(U_i, \mathbf{X}_i, Z_i, \delta_i)\}_{i=1}^n$. Although the proposed estimation method is free of the dimension of the covariate vector $(U, \mathbf{X})$ because no any type of kernel estimator of $F(u, \mathbf{x}, y)$ is used, it has the drawback that it assumes that the censoring variable C depends on $(U, \mathbf{X})$ in a very particular way. This type of dependence might hold true when the censoring is purely 'administrative' (censoring at the end of the study), but when the censoring is caused by other factors such as death due

to another disease or change of treatment, then less restrictive assumptions on the censoring mechanism are required. Following Akritas (1994) and Van Keilegom and Akritas (1999), we may replace the above assumption on the censoring mechanism with a more flexible assumption that the response variable Y is conditionally independent of the censoring variable C given $(U, \mathbf{X})$, and propose an estimator for the joint distribution function $F(u, \mathbf{x}, y)$ of $(U, \mathbf{X}, Y)$ by averaging the local Kaplan-Meier estimates of the conditional distribution function $F(y|u, \mathbf{x})$ of Y given $U = u$ and $\mathbf{X} = \mathbf{x}$ over a range of values of $(u, \mathbf{x})$. However, this is not recommended in practice since the local Kaplan-Meier estimator of $F(y|u, \mathbf{x})$ will suffer from the "curse of dimensionality" problem when the dimension of $(U, \mathbf{X})$ is high.

The following theorem gives the asymptotic normality of $\widehat{\mathbf{\Psi}}(u_0)$ and $\widehat{\mathbf{a}}(u_0)$.

Theorem 2.1. *Suppose that the conditions C1–C7 hold. Then, for any $u_0 \in \Omega$, we have*

$$\sqrt{nh}\left[\widehat{\mathbf{\Psi}}(u_0) - \mathbf{\Psi}(u_0) - 2^{-1}h^2\mu_2\begin{pmatrix}\mathbf{a}''(u_0)\\\mathbf{0}\end{pmatrix} + o_P(h^2)\right] \xrightarrow{D} N(\mathbf{0}, \mathbf{\Sigma}(u_0)),$$

where

$$\mathbf{\Sigma}(u_0) = f^{-1}(u_0)(\mathbf{\Gamma}^{-1}(u_0)\mathbf{\Gamma}^*(u_0)\mathbf{\Gamma}^{-1}(u_0)) \otimes \begin{pmatrix}\nu_0 & 0\\0 & \nu_2/\mu_2^2\end{pmatrix}$$

with $\otimes$ being the Kronecker product. In particular, we have

$$\sqrt{nh}\left[\widehat{\mathbf{a}}(u_0) - \mathbf{a}(u_0) - 2^{-1}h^2\mu_2\mathbf{a}''(u_0) + o_P(h^2)\right] \xrightarrow{D} N(\mathbf{0}, \mathbf{\Sigma}_1(u_0)),$$

where

$$\mathbf{\Sigma}_1(u_0) = \nu_0 f^{-1}(u_0)\mathbf{\Gamma}^{-1}(u_0)\mathbf{\Gamma}^*(u_0)\mathbf{\Gamma}^{-1}(u_0).$$

Remark 2.4. We can see from Theorem 2.1 that the asymptotic bias for the censored case is the same as that for the uncensored situation, while the asymptotic variance for the censored case is larger than its counterpart for the uncensored situation. In other words, the asymptotic bias does not depend on the censoring scheme but the asymptotic variance does rely on the censoring structure. This is not surprising because the asymptotic bias comes from the linear approximation. Furthermore, it is easy to see that the asymptotic mean square error (AMSE) is given by

$$\text{AMSE} = \frac{h^4}{4}\mu_2^2\|\mathbf{a}''(u)\|^2 + \frac{\nu_0}{nhf(u)}\text{tr}(\mathbf{\Gamma}^{-1}(u)\mathbf{\Gamma}^*(u)\mathbf{\Gamma}^{-1}(u)).$$

By minimizing the AMSE with respect to h, we obtain the asymptotic optimal bandwidth as

$$h_{\text{opt}} = \left\{\frac{\nu_0\text{tr}(\mathbf{\Gamma}^{-1}(u)\mathbf{\Gamma}^*(u)\mathbf{\Gamma}^{-1}(u))}{f(u)\mu_2^2\|\mathbf{a}''(u)\|^2}\right\}^{1/5} n^{-\frac{1}{5}}.$$

Hence, as expected, the optimal convergence rate of AMSE is of the order of $n^{-4/5}$. Finally, in the absence of random right censoring, the asymptotic variance $\mathbf{\Sigma}_1(u_0)$ reduces to $\nu_0 f^{-1}(u_0)\mathbf{\Gamma}^{-1}(u_0)\text{E}(\sigma^2(U, \mathbf{X})\mathbf{X}\mathbf{X}^{\mathrm{T}}|U = u)\mathbf{\Gamma}^{-1}(u_0)$, which is the asymptotic variance without censoring (see Cai et al., 2000).

Remark 2.5. Although this paper focuses on the case where the smoothing variable U in model (1.1) is one-dimensional, the proposed estimation method and its asymptotic theory continue to hold for the situation of multidimensional smoothing variable but more and complicated notations involve, see Ruppert and Wand

(1994). Nevertheless, such an extension is not recommended in practice because of the well-known "curse of dimensionality" problem.

2.3. *Selection of bandwidth.* In this section, we address how to select the bandwidth in the proposed estimation method, which is an important issue related to the practical implementation of the proposed estimation method. Like that in Cai (2003), we employ the corrected version of Akaike information criterion (AICC) proposed by Hurvich et al. (1998) to choose the optimal value of the bandwidth h. Specifically, the AICC score for the uncensored case is defined as

$$\text{AICC}(h) = \log\left\{\frac{1}{n}\sum_{i=1}^{n}[Y_i - \mathbf{X}_i^{\text{T}}\widetilde{\mathbf{a}}(U_i)]^2\right\} + \frac{n + \text{tr}(\widetilde{\mathbf{S}})}{n - 2 - \text{tr}(\widetilde{\mathbf{S}})}. \qquad (2.11)$$

However, in the presence of random right censoring, as suggested by Cai (2003), the AICC score is of the following form

$$\text{AICC}(h) = \log\left\{\sum_{i=1}^{n} W_{in}[Z_i - \mathbf{X}_i^{\text{T}}\widehat{\mathbf{a}}(U_i)]^2\right\} + \frac{n + \text{tr}(\widehat{\mathbf{S}})}{n - 2 - \text{tr}(\widehat{\mathbf{S}})}. \qquad (2.12)$$

2.4. *Comparison with the synthetic data method of Luo et al. (2006).* As mentioned in the introduction, Luo et al. (2006) proposed a synthetic data method for randomly right censored varying coefficient models based on the unbiased data transformation technique proposed by Fan and Gijbels (1994) and the local polynomial smoothing method and studied the asymptotic normality of the resulting estimators of the coefficient functions. Unfortunately, the authors did not investigate the finite sample performance of their estimation method through simulation studies. In what follows, we will briefly introduce the synthetic data method of Luo et al. (2006) and compare it with the proposed estimation method. The synthetic data method of Luo et al. (2006) consists of the following two steps.

Transformation of data. Let $G(t|u, \mathbf{x})$ be the conditional distribution function of the censoring variable C given $U = u$ and $\mathbf{X} = \mathbf{x}$. When $G(t|u, \mathbf{x})$ is unknown, assume that $\widehat{G}(t|u, \mathbf{x})$ is an estimator of $G(t|u, \mathbf{x})$, for example, the Kaplan-Meier estimator when the censoring variable C is independent of the covariate vector $(U, \mathbf{X})$ or the local Kaplan-Meier estimator when C depends on $(U, \mathbf{X})$. Luo et al. (2006) constructed the following synthetic data or pseudo-responses

$$Z_{i\widehat{G}} = (1 + \phi)L_{i\widehat{G}} - \phi K_{i\widehat{G}}, \quad i = 1, \cdots, n, \qquad (2.13)$$

where $L_{i\widehat{G}} = \int_0^{Z_i}\{1 - \widehat{G}(t - |u, \mathbf{x})\}^{-1}\text{d}t$, $K_{i\widehat{G}} = Z_i\delta_i\{1 - \widehat{G}(Z_i - |U_i, \mathbf{X}_i)\}^{-1}$, ϕ is a tuning parameter which controls the weights put on the censored and uncensored observations, $\widehat{G}(t - |u, \mathbf{x})$ is the left limit of $\widehat{G}(t|u, \mathbf{x})$ at t. Under the assumption that the response variable Y is conditionally independent of the censoring variable C given $(U, \mathbf{X})$, it is easy to show that $\text{E}(Z_{iG}|U_i, \mathbf{X}_i) = \text{E}(Y_i|U_i, \mathbf{X}_i)$, hence the above data transformation is unbiased when $G(t|u, \mathbf{x})$ is known. This type of transformation was originally introduced by Fan and Gijbels (1994) in randomly right censored nonparametric regression models. Note that $\phi = -1$ and $\phi = 0$ corresponds to the Koul et al. (1981) transformation $K_{i\widehat{G}}$ (abbreviated as the KSV transformation) and the Leurgans (1987) transformation $L_{i\widehat{G}}$, respectively. As pointed out by Fan and Gijbels (1994), an appropriate choice of ϕ can reduce the variability of the transformed data. In practice, the choice of $\phi > 0$ focuses more on the censored

observations than on the uncensored observations, which is more intuitive than Leurgans's equal choice with $\phi = 0$. Fan and Gijbels (1994) recommended the following choice of ϕ:

$$\widehat{\phi} = \min_{\{i:\delta_i=1\}} \frac{\int_0^{Z_i} \{1 - \widehat{G}(t - |u, \mathbf{x})\}^{-1} \mathrm{d}\mathbf{t} - Z_i}{Z_i \{1 - \widehat{G}(Z_i - |U_i, \mathbf{X}_i)\}^{-1} - \int_0^{Z_i} \{1 - \widehat{G}(t - |u, \mathbf{x})\}^{-1} \mathrm{d}\mathbf{t}}. \tag{2.14}$$

We will use $\widehat{\phi}$ in our simulation studies, although any ϕ between 0 and $\widehat{\phi}$ should work.

Application of local polynomial smoothing method. Based on the transformed data $\{U_i, \mathbf{X}_i, Z_{i\widehat{G}}\}_{i=1}^n$, Luo et al. (2006) employed the commonly used local polynomial smoothing method to estimate the coefficient functions. Without loss of generality and for the convenience of comparison, here we only focus on the case of local linear smoothing. In this case, the loss function to be minimized for the estimation of the coefficient function vector $\mathbf{a}(\cdot)$ is similar to (2.1) and is of the following form

$$\frac{1}{n} \sum_{i=1}^n \left[Z_{i\widehat{G}} - \mathbf{X}_i^{\mathrm{T}} \mathbf{a}(u_0) - (U_i - u_0) \mathbf{X}_i^{\mathrm{T}} \mathbf{a}'(u_0) \right]^2 K_h(U_i - u_0). \tag{2.15}$$

Similarly to (2.3), the estimator of $\mathbf{a}(u)$ at u_0 can be expressed as

$$\widehat{\mathbf{a}}_{\widehat{G}}(u_0) = (\mathbf{I}_p, \mathbf{0}_{p \times p}) [\mathbf{X}(u_0)^{\mathrm{T}} \mathbf{W}(u_0) \mathbf{X}(u_0)]^{-1} \mathbf{X}(u_0)^{\mathrm{T}} \mathbf{W}(u_0) \mathbf{Z}_{\widehat{G}}, \tag{2.16}$$

where $\mathbf{Z}_{\widehat{G}} = (Z_{1\widehat{G}}, \cdots, Z_{n\widehat{G}})^{\mathrm{T}}$. Note that in our simulation studies, to make a fair comparison, we employed the AICC method to choose the bandwidth in the synthetic data method of Luo et al. (2006), although other bandwidth selection methods such as the CV method and the GCV method are applicable. In this case, by replacing Y_i and $\widetilde{\mathbf{a}}(U_i)$ in (2.11) by $Z_{i\widehat{G}}$ and $\widehat{\mathbf{a}}_{\widehat{G}}(U_i)$, respectively, we can obtain the AICC score as

$$\mathrm{AICC}(h) = \log \left\{ \frac{1}{n} \sum_{i=1}^n [Z_{i\widehat{G}} - \mathbf{X}_i^{\mathrm{T}} \widehat{\mathbf{a}}_{\widehat{G}}(U_i)]^2 \right\} + \frac{n + \mathrm{tr}(\widetilde{\mathbf{S}})}{n - 2 - \mathrm{tr}(\widetilde{\mathbf{S}})}. \tag{2.17}$$

Remark 2.6. The proposed estimation method differs from the synthetic data method of Luo et al. (2006) in the following aspects. First, the condition on the censoring mechanism is different. In the synthetic data method of Luo et al. (2006), the condition on the censoring mechanism is that the response variable Y is conditionally independent of the censoring variable C given $(U, \mathbf{X})$, while in the proposed estimation method the condition on the censoring mechanism is that Y is unconditionally independent of C and $P(Y \leq C | U, \mathbf{X}, Y) = P(Y \leq C | Y)$. When the censoring variable C is independent of the covariate vector $(U, \mathbf{X})$, which is the case, for example, when the censoring occurs at the end of the study, the condition on the censoring mechanism in the synthetic data method of Luo et al. (2006) reduces to that in the proposed estimation method since C is also independent of $(U, \mathbf{X}, Y)$ in this case. Second, the proposed estimation method is computationally simple and easy to implement compared to the synthetic data method of Luo et al. (2006). Unlike the synthetic data method of Luo et al. (2006), no data transformation is needed to account for the censoring in the proposed estimation method, thus the proposed estimation method not only enjoys the advantage of computational simplicity and practical feasibility, but also is expected to give better finite

sample performance than the synthetic data method of Luo et al. (2006), which is confirmed by the simulation studies conducted in Section 3.

2.5. *Extension to randomly right censored varying coefficient models with different smoothing variables.* In this paper, we consider such a kind of varying coefficient models where all coefficient functions share the same smoothing variable. However, the assumption of all coefficient functions sharing the same smoothing variable is strict and has limited applications. In this section, we extend the estimation method proposed in Section 2 to randomly right censored varying coefficient models with different smoothing variables. The sample form of this kind of varying coefficient models is

$$Y_i = \sum_{j=1}^{p} a_j(U_{ij})X_{ij} + \varepsilon_i, \quad i = 1, \cdots, n, \tag{2.18}$$

where U_{ij} is the observation of j-th smoothing variable U_j, the other symbols are the same as those in (1.1). When the response variable Y is fully observed, model (2.18) has been studied in estimation (Zhang et al., 2002; Zhang and Li, 2007) and hypothesis testing (Ip et al., 2007). When the response variable Y is subject to randomly right censoring, Yang et al. (2014) proposed an estimation method for (2.18) based on synthetic data obtained by the unbiased transformation given by Koul et al. (1981) and the smooth back-fitting technique and studied the asymptotic normality of the resulting estimators of the coefficient functions. Our estimation method for model (2.18) is briefly described as follows. In the absence of randomly right censoring, via approximating $a_j(u_j)$ by $a_j(u_{0j}) + a_j'(u_{0j})(u_j - u_{0j})$ in a neighborhood of a given u_{0j} in the domain of the j-th smoothing variable U_j, we obtain the following kernel weighted local least squares function

$$\frac{1}{n} \sum_{i=1}^{n} \left\{ Y_i - \sum_{j=1}^{p} [a_j(u_{0j}) + a_j'(u_{0j})(U_{ij} - u_{0j})]X_{ij} \right\}^2 K_{\mathbf{h}}(\mathbf{U}_i - \mathbf{u}_0)$$

$$= \int \left\{ y - \sum_{j=1}^{p} [a_j(u_{0j}) + a_j'(u_{0j})(u_j - u_{0j})]x_j \right\}^2 K_{\mathbf{h}}(\mathbf{u} - \mathbf{u}_0) \, \mathrm{d}F_n(\mathbf{u}, \mathbf{x}, y), \tag{2.19}$$

where $K_{\mathbf{h}}(\mathbf{U}_i - \mathbf{u}_0) = \prod_{j=1}^{p} K_{h_j}(U_{ij} - u_{0j})$, $K_{h_j}(\cdot) = K(\cdot/h_j)/h_j$ with $K(\cdot)$ being an univariate kernel function, $\mathbf{h} = (h_1, \cdots, h_p)^{\mathrm{T}}$ is the bandwidth vector, $\mathbf{U}_i = (U_{i1}, \cdots, U_{ip})^{\mathrm{T}}$, $\mathbf{u}_0 = (u_{01}, \cdots, u_{0p})^{\mathrm{T}}$, $\mathbf{u} = (u_1, \cdots, u_p)^{\mathrm{T}}$, $\mathbf{x} = (x_1, \cdots, x_p)^{\mathrm{T}}$, $F_n(\mathbf{u}, \mathbf{x}, y)$ is the empirical distribution function of $\{(\mathbf{U}_i, \mathbf{X}_i, Y_i)\}_{i=1}^{n}$.

In the presence of randomly right censoring, by replacing $F_n(\mathbf{u}, \mathbf{x}, y)$ in (2.19) with $\widehat{F}_n(\mathbf{u}, \mathbf{x}, y)$ defined by

$$\widehat{F}_n(\mathbf{u}, \mathbf{x}, y) = \sum_{i=1}^{n} W_{in} I(\mathbf{U}_{(i)} \leq \mathbf{u}, \mathbf{X}_{(i)} \leq \mathbf{x}, Z_{(i)} \leq y),$$

where W_{in} is defined by (2.6), $\mathbf{U}_{(i)}$ and $\mathbf{X}_{(i)}$ are the $\mathbf{U}$ and $\mathbf{X}$ associated with $Z_{(i)}$, respectively, we get the following weighted local least squares function

$$\int \left\{ y - \sum_{j=1}^{p} [a_j(u_{0j}) + a'_j(u_{0j})(u_j - u_{0j})] x_j \right\}^2 K_{\mathbf{h}}(\mathbf{u} - \mathbf{u}_0) \, \mathrm{d}\widehat{F}_n(\mathbf{u}, \mathbf{x}, y)$$

$$= \sum_{i=1}^{n} W_{in}^* \left\{ Z_i - \sum_{j=1}^{p} [a_j(u_{0j}) + a'_j(u_{0j})(U_{ij} - u_{0j})] X_{ij} \right\}^2 K_{\mathbf{h}}(\mathbf{U}_i - \mathbf{u}_0), \quad (2.20)$$

where W_{in}^* is the W_{in}-value associated with the natural order of Z_i.

Let $\{\widetilde{a}_j(\mathbf{u}_0)\}_{j=1}^{p}$ be the first p entries of the minimizer of (2.20), $\mathbf{W}(\mathbf{u}_0) = \mathrm{diag}(K_{\mathbf{h}}(\mathbf{U}_1 - \mathbf{u}_0), \cdots, K_{\mathbf{h}}(\mathbf{U}_n - \mathbf{u}_0))$ and

$$\mathbf{X}(\mathbf{u}_0) = \begin{pmatrix} X_{11} & \cdots & X_{1p} & h_1^{-1}(U_{11} - u_{01})X_{11} & \cdots & h_p^{-1}(U_{1p} - u_{0p})X_{1p} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ X_{n1} & \cdots & X_{np} & h_1^{-1}(U_{n1} - u_{01})X_{n1} & \cdots & h_p^{-1}(U_{np} - u_{0p})X_{np} \end{pmatrix}.$$

Then it follows from the least squares theory that

$$\widetilde{a}_j(\mathbf{u}_0) = \mathbf{e}_{j,2p}^{\mathrm{T}} [\mathbf{X}(\mathbf{u}_0)^{\mathrm{T}} \mathbf{W}^* \mathbf{W}(\mathbf{u}_0) \mathbf{X}(\mathbf{u}_0)]^{-1} \mathbf{X}(\mathbf{u}_0)^{\mathrm{T}} \mathbf{W}^* \mathbf{W}(\mathbf{u}_0) \mathbf{Z}, \; j = 1, \cdots, p, \tag{2.21}$$

where $\mathbf{e}_{j,2p}$ is a $2p{\times}1$ unit vector with 1 at its j-th position and $\mathbf{W}^*$ and $\mathbf{Z}$ are the same as those in Section 2.1.

Like Cai and Fan (2000) and Linton and Härdle (1996), we employ the average method to obtain the estimator of $a_j(u_j)$ at u_{0j} as

$$\widehat{a}_j(u_{0j}) = \frac{1}{n} \sum_{i=1}^{n} \widetilde{a}_j(U_{i1}, \cdots, U_{i,j-1}, u_{0j}, U_{i,j+1}, \cdots, U_{i,n}), \; j = 1, \cdots, p. \tag{2.22}$$

Remark 2.7. Although the proposed estimation method can be extended to the situation where different coefficient functions depend on different smoothing variables, the theoretical justification of the asymptotic normality of the resulting estimators of the coefficient functions is not obvious because of the complexity of the resulting estimator. This is beyond the scope of the current paper, and is warranted as a future topic of research. Furthermore, since $\widehat{a}_j(u_{0j})$ is no longer the linear combinations of $Z_1, \cdots, Z_n$, the AICC bandwidth selector proposed in Section 2.3 can not be used to choose the bandwidth vector $\mathbf{h}$. One possible solution to this issue is to employ the rule-of-thumb (ROT) method, that is, $h_j = S_{u_j} n^{-0.2}$, where S_{u_j} is the sample standard deviation of the observations $U_{1j}, \cdots, U_{nj}$ of U_j, $j = 1, \cdots, p$.

3. Simulation studies

As emphasized in the introduction, the accurate estimate of the coefficient functions plays an important role in using varying coefficient models to explore the dynamic pattern of a regression relationship. In this section, we conduct simulation studies to assess the finite sample performance of the proposed estimation method. The experiments mainly focus on the accuracy of the coefficient estimates.

3.1. *Design of the experiment.* The experimental data were generated from the following model

$$Y = a_1(U)X_1 + a_2(U)X_2 + a_3(U)X_3 + \varepsilon,$$

where $X_1 \equiv 1$, $U \sim U(0,1)$, $(X_2, X_3)^{\mathrm{T}}$ was distributed as a bivariate normal distribution with mean vector $(0,0)^{\mathrm{T}}$ and covariance matrix $\left(\begin{smallmatrix} 1 & 0.5 \\ 0.5 & 1 \end{smallmatrix} \right)$, and the error term ε was generated from a normal distribution $N(0, 0.25)$ and was independent of (U, X_2, X_3). The coefficient functions $a_1(u)$, $a_2(u)$ and $a_3(u)$ are, respectively, chosen from the following two groups of functions:

Group 1. $a_1(u) = 0.2 \exp(-0.6 + 3.0u)$, $a_2(u) = 1 + (1 - 2u)^3$, $a_3(u) = \log_2(1 + 3u)$;

Group 2. $a_1(u) = 8u(1 - u)$, $a_2(u) = 1 + \cos(2\pi u)$, $a_3(u) = 3.5[\exp(-(4u - 1)^2) + \exp(-(4u - 3)^2)] - 1.5$.

The coefficient functions in Group 1 are monotone, while they are not monotone in Group 2.

The censoring variable C was distributed as the uniform distribution $U(0, c_1)$, where the parameter c_1 was used to control the censoring rate. In our simulation, we considered three different censoring rates, namely, 15%, 30% and 45% of censoring.

We used the weighted local linear smoothing (WLLS) method and the synthetic data (SD) method proposed by Luo et al. (2006) to estimate the coefficient functions and cross compared the results in terms of the accuracy of the coefficient estimates. In both methods, the kernel function was taken to be the Gaussian kernel function $K(u) = \frac{1}{\sqrt{2\pi}} \exp(-u^2/2)$ and the bandwidth was selected by the AICC procedure described in Section 2.3.

The accuracy of coefficient estimates is evaluated by using mean square error (MSE), defined by

$$\mathrm{MSE}_j = n_0^{-1} \sum_{k=1}^{n_0} [\widehat{a}_j(u_k) - a_j(u_k)]^2, \quad j = 1, 2, 3,$$

where $\{u_k, k = 1, 2, \cdots, n_0\}$ is a set of grid points uniformly placed on $(0, 1)$ and $\widehat{a}_j(\cdot)$ is an estimate of $a_j(\cdot)$ obtained by using any one of the WLLS and SD methods. In our simulation, we took $n_0 = 51$.

The sample size n was set to 200 and 400. For each combination of the two sample sizes, the two groups of coefficient functions, and the three censoring rates, the simulation was repeated 500 times.

3.2. *Simulation results with analysis.* The mean and standard deviation of the MSEs over 500 simulations are summarized in Table 3.1, in which the numbers in parentheses are the standard deviations. Furthermore, we depict in Figures 3.1–3.4 the mean curves of the two groups of coefficient functions estimated by the WLLS and SD methods, respectively.

We can see from Table 3.1 that the means and standard deviations for the WLLS method are consistently much smaller than the corresponding means and standard deviations for the SD method for the two groups of coefficient functions, the two sample sizes, and the three censoring rates under investigation, which shows that the WLLS method produces more accurate estimates of the coefficient functions than the SD method. More details can be seen from Figures 3.1–3.4. For each group of coefficient functions, the WLLS method retrieves the true coefficient curves very well, but the SD method produces less accurate estimates of the coefficient functions especially when the censoring rate is large. We also find from Table 3.1 that the means and standard deviations for the two estimation methods both decrease as the sample size n increases. However, by further comparing the estimated coefficient functions under $n = 200$ with those under $n = 400$, it seems that the degree of

Coefficient group	Censoring rate(%)	c_1	n	WLLS			SD		
				$a_1(\cdot)$	$a_2(\cdot)$	$a_3(\cdot)$	$a_1(\cdot)$	$a_2(\cdot)$	$a_3(\cdot)$
Group 1	15	8.23	200	0.0111	0.0169	0.0139	0.0361	0.0654	0.0871
				(0.0076)	(0.0108)	(0.0109)	(0.0475)	(0.0965)	(0.1304)
			400	0.0052	0.0081	0.0071	0.0179	0.0291	0.0416
				(0.0031)	(0.0049)	(0.0046)	(0.0205)	(0.0282)	(0.0282)
	30	3.86	200	0.0203	0.0241	0.0271	0.0776	0.1259	0.1363
				(0.0157)	(0.0175)	(0.0290)	(0.1285)	(0.1571)	(0.2016)
			400	0.0112	0.0119	0.0137	0.0390	0.0703	0.0736
				(0.0087)	(0.0075)	(0.0104)	(0.0378)	(0.0896)	(0.1135)
	45	1.84	200	0.0801	0.0449	0.0667	0.1306	0.1086	0.2247
				(0.0731)	(0.0436)	(0.0760)	(0.0753)	(0.0830)	(0.1471)
			400	0.0456	0.0209	0.0318	0.1275	0.0791	0.2187
				(0.0303)	(0.0184)	(0.0238)	(0.0566)	(0.0437)	(0.0957)
Group 2	15	11.73	200	0.0157	0.0247	0.0323	0.0403	0.0817	0.1145
				(0.0100)	(0.0187)	(0.0182)	(0.0366)	(0.0847)	(0.1750)
			400	0.0100	0.0151	0.0172	0.0214	0.0493	0.0574
				(0.0047)	(0.0079)	(0.0088)	(0.0155)	(0.1304)	(0.0595)
	30	5.75	200	0.0325	0.0504	0.0585	0.1114	0.2126	0.2813
				(0.0225)	(0.0451)	(0.0727)	(0.1357)	(0.3123)	(0.3230)
			400	0.0160	0.0225	0.0245	0.0569	0.1130	0.1494
				(0.0063)	(0.0134)	(0.0117)	(0.0471)	(0.1003)	(0.1560)
	45	3.40	200	0.0944	0.1119	0.1238	0.1150	0.2386	0.3065
				(0.0604)	(0.1462)	(0.1131)	(0.1394)	(0.4359)	(0.4571)
			400	0.0402	0.0362	0.0430	0.0749	0.1225	0.1928
				(0.0169)	(0.0204)	(0.0222)	(0.0938)	(0.1129)	(0.1058)

TABLE 3.1. Mean and standard deviation (in parentheses) of the mean square errors over 500 simulations

improvement for both estimation methods decreases as the censoring rate increases, we owe this observation to the referee. This is reasonable because the larger the censoring rate is, the more information in the data will be lost, which makes it more difficult to further improve the finite sample performance of both estimation methods especially as the sample size n increases from a large value 200 to a larger value 400.

4. Conclusion remark

In this paper, we proposed a new estimation method for randomly right censored varying coefficient models and established the asymptotic normality of the resulting estimator. Different from the existing estimation method, the proposed estimation method is computationally simple and practically feasible since it requires no transformation of data. This superiority becomes more evident when the censoring distribution function depends on the covariate vector $(U, \mathbf{X})$ and the dimension of $(U, \mathbf{X})$ is large. Simulation studies were conducted to assess the finite sample performance of the proposed estimation method by comparing the results with those obtained by the existing estimation method. The simulation results demonstrate that the proposed estimation method indeed has better finite sample performance than the existing estimation method.

5. Proofs

To facilitate the presentation, we introduce some notations, which will be used in the proofs of some Lemmas and Theorem 2.1. Let $W_G = \frac{\delta}{1-G(Z-)}$, $W_{iG} = \frac{\delta_i}{1-G(Z_i-)}$, $W_{i\widehat{G}} = \frac{\delta_i}{1-\widehat{G}(Z_i-)}$, $Z_{iG} = \frac{\delta_i Z_i}{1-G(Z_i-)}$, $Z_{i\widehat{G}} = \frac{\delta_i Z_i}{1-\widehat{G}(Z_i-)}$, and $\varepsilon_{iG} = Z_{iG} - $
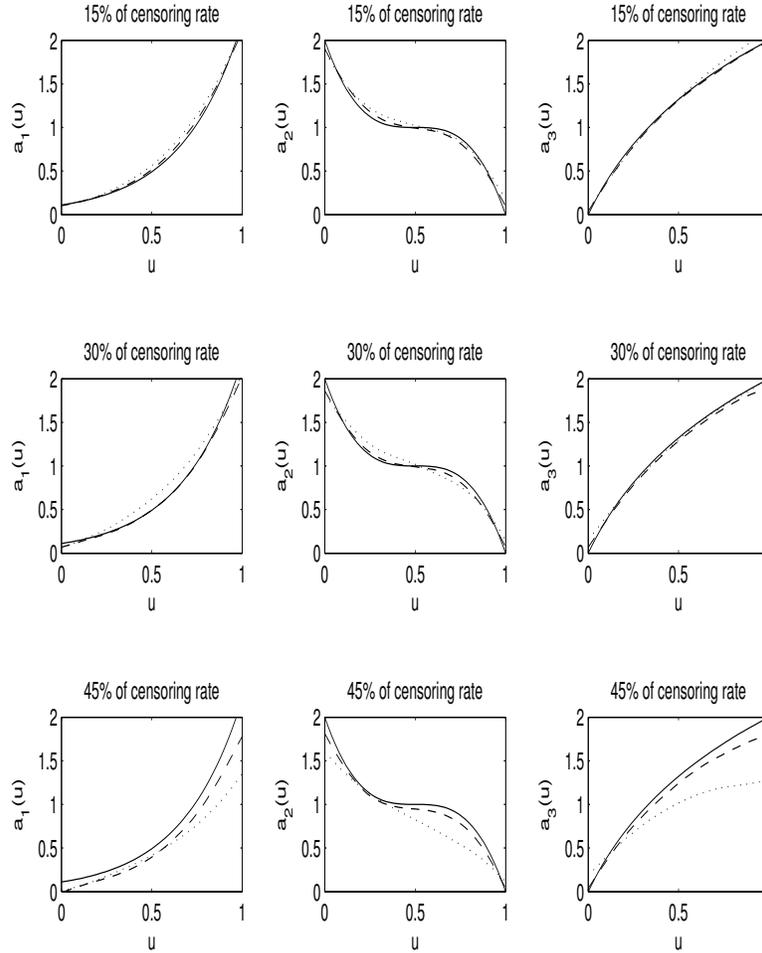
FIGURE 3.1. The mean curves of the coefficient functions (solid line) in Group 1 estimated by the WLLS method (dashed line) and the SD method (dotted line) with $n = 200$.

$\mathrm{E}(Z_{iG}|U_i, \mathbf{X}_i)$, where $\widehat{G}(\cdot)$ is the Kaplan-Meier estimator of $G(\cdot)$. Furthermore, let

$$\mathbf{S}_{n,l}(u_0) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{X}_i \mathbf{X}_i^{\mathrm{T}} h^{-l}(U_i - u_0)^l K_h(U_i - u_0),$$

$$\mathbf{S}_{n,lG}(u_0) = \frac{1}{n} \sum_{i=1}^{n} W_{iG} \mathbf{X}_i \mathbf{X}_i^{\mathrm{T}} h^{-l}(U_i - u_0)^l K_h(U_i - u_0),$$

$$\mathbf{S}_{n,l\widehat{G}}(u_0) = \frac{1}{n} \sum_{i=1}^{n} W_{i\widehat{G}} \mathbf{X}_i \mathbf{X}_i^{\mathrm{T}} h^{-l}(U_i - u_0)^l K_h(U_i - u_0),$$
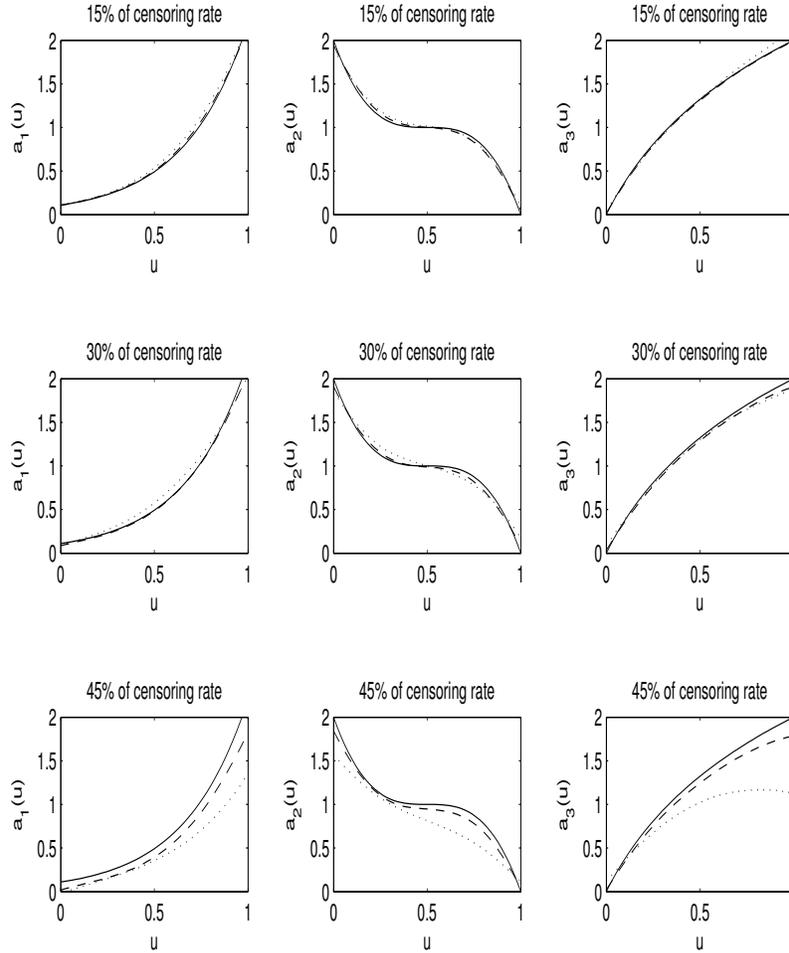
FIGURE 3.2. The mean curves of the coefficient functions (solid line) in Group 1 estimated by the WLLS method (dashed line) and the SD method (dotted line) with $n = 400$.

$$\mathbf{T}_{n,lG}(u_0) \;=\; \frac{1}{n}\sum_{i=1}^{n} Z_{iG}\mathbf{X}_i h^{-l}(U_i - u_0)^l K_h(U_i - u_0),$$

$$\mathbf{T}_{n,l\widehat{G}}(u_0) \;=\; \frac{1}{n}\sum_{i=1}^{n} Z_{i\widehat{G}}\mathbf{X}_i h^{-l}(U_i - u_0)^l K_h(U_i - u_0),$$

and

$$\mathbf{T}_{n,lG}^{*}(u_0) = \frac{1}{n}\sum_{i=1}^{n} \varepsilon_{iG}\mathbf{X}_i h^{-l}(U_i - u_0)^l K_h(U_i - u_0).$$
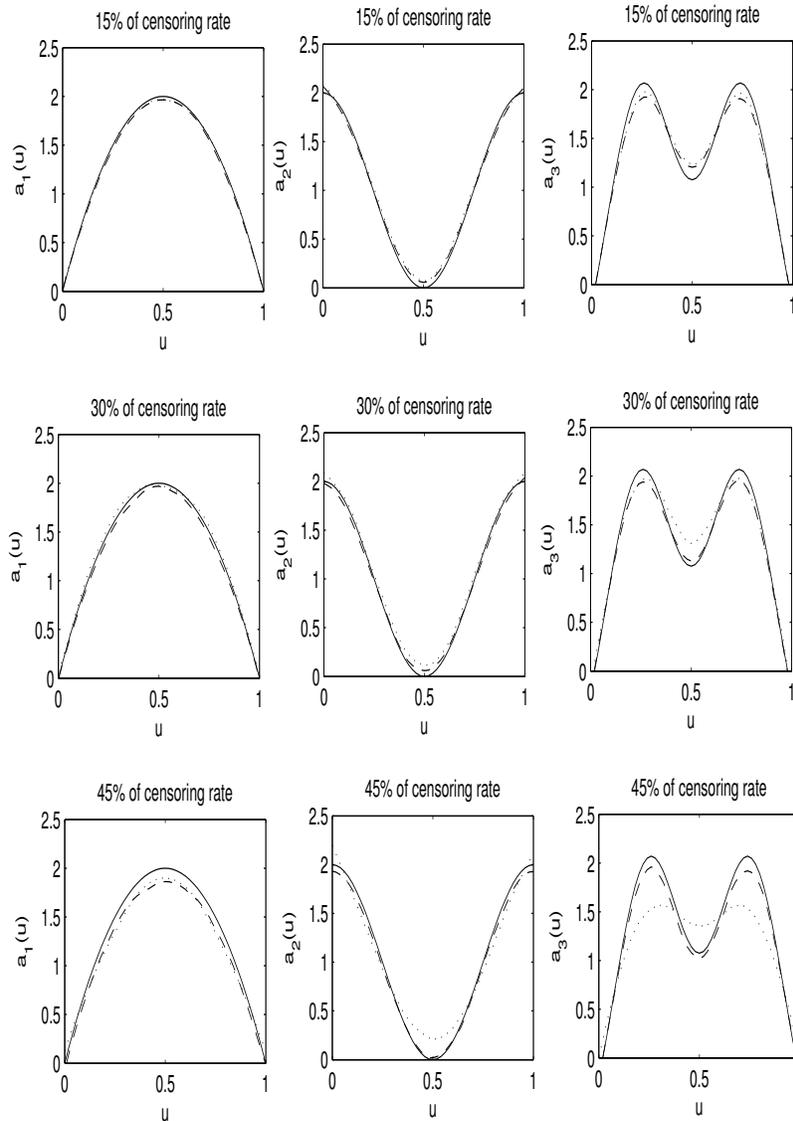
FIGURE 3.3. The mean curves of the coefficient functions (solid line) in Group 2 estimated by the WLLS method (dashed line) and the SD method (dotted line) with $n = 200$.

Lemma 5.1. *Under the conditions C1–C6, we have*

$$\mathbf{S}_{nG}(u_0) = \begin{pmatrix} \mathbf{S}_{n,0G}(u_0) & \mathbf{S}_{n,1G}(u_0) \\ \mathbf{S}_{n,1G}(u_0) & \mathbf{S}_{n,2G}(u_0) \end{pmatrix} = f(u_0)\mathbf{\Gamma}(u_0)\otimes\begin{pmatrix} 1 & 0 \\ 0 & \mu_2 \end{pmatrix}[1 + o_P(1)] \quad (5.1)$$

and

$$\mathbf{S}_{n\widehat{G}}(u_0) = \begin{pmatrix} \mathbf{S}_{n,0\widehat{G}}(u_0) & \mathbf{S}_{n,1\widehat{G}}(u_0) \\ \mathbf{S}_{n,1\widehat{G}}(u_0) & \mathbf{S}_{n,2\widehat{G}}(u_0) \end{pmatrix} = f(u_0)\mathbf{\Gamma}(u_0)\otimes\begin{pmatrix} 1 & 0 \\ 0 & \mu_2 \end{pmatrix}[1 + o_P(1)]. \quad (5.2)$$
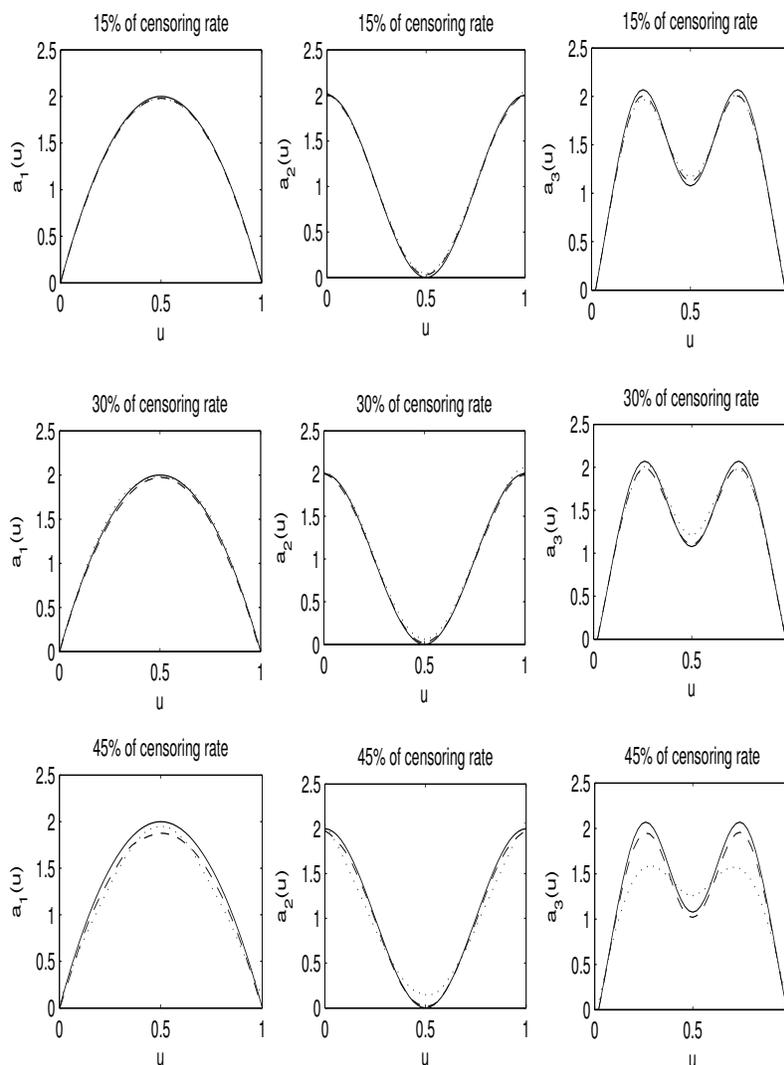
FIGURE 3.4. The mean curves of the coefficient functions (solid line) in Group 2 estimated by the WLLS method (dashed line) and the SD method (dotted line) with $n = 400$.

Proof: It follows from $\mathrm{E}(W_G|U, \mathbf{X}) = 1$ and $\mathrm{E}(W_G^2|U, \mathbf{X}) = b_0(U, \mathbf{X})$ that

$$
\begin{aligned}
\mathrm{E}(\mathbf{S}_{n,lG}(u_0)) &= \mathrm{E}[W_G \mathbf{X}\mathbf{X}^{\mathrm{T}} h^{-l}(U - u_0)^l K_h(U - u_0)] \\
&= \mathrm{E}[\mathrm{E}(W_G|U, \mathbf{X})\mathbf{X}\mathbf{X}^{\mathrm{T}} h^{-l}(U - u_0)^l K_h(U - u_0)] \\
&= \mathrm{E}[\mathbf{X}\mathbf{X}^{\mathrm{T}} h^{-l}(U - u_0)^l K_h(U - u_0)] \\
&= \mathrm{E}[\mathrm{E}(\mathbf{X}\mathbf{X}^{\mathrm{T}}|U) h^{-l}(U - u_0)^l K_h(U - u_0)] \\
&= \mathrm{E}[\boldsymbol{\Gamma}(U) h^{-l}(U - u_0)^l K_h(U - u_0)] \\
&= \int_{u_0-h}^{u_0+h} \boldsymbol{\Gamma}(u) h^{-l-1}(u - u_0)^l K((u - u_0)/h) f(u)\, \mathrm{d}u \\
&= \int_{-1}^{1} \boldsymbol{\Gamma}(u_0 + ht) f(u_0 + ht) t^l K(t)\, \mathrm{d}t \\
&= f(u_0)\boldsymbol{\Gamma}(u_0)\mu_l[1 + o(1)]
\end{aligned}
$$

and

$$
\begin{aligned}
\mathrm{Var}(\mathbf{S}_{n,lG}(u_0)) &= n^{-1}\mathrm{Var}[W_G \mathbf{X}\mathbf{X}^{\mathrm{T}} h^{-l}(U-u_0)^l K_h(U-u_0)] \\
&\le n^{-1}\mathrm{E}[W_G^2 \mathbf{X}^{\otimes 4} h^{-2l}(U-u_0)^{2l} K_h^2(U-u_0)] \\
&= n^{-1}\mathrm{E}[\mathrm{E}(W_G^2|U,\mathbf{X})\mathbf{X}^{\otimes 4} h^{-2l}(U-u_0)^{2l} K_h^2(U-u_0)] \\
&= n^{-1}\mathrm{E}[b_0(U,\mathbf{X})\mathbf{X}^{\otimes 4} h^{-2l}(U-u_0)^{2l} K_h^2(U-u_0)] \\
&= n^{-1}\mathrm{E}[\mathrm{E}(b_0(U,\mathbf{X})\mathbf{X}^{\otimes 4}|U)h^{-2l}(U-u_0)^{2l} K_h^2(U-u_0)] \\
&= n^{-1}\mathrm{E}[\boldsymbol{\Gamma}_{10}^*(U)h^{-2l}(U-u_0)^{2l} K_h^2(U-u_0)] \\
&= n^{-1}\int_{u_0-h}^{u_0+h}\boldsymbol{\Gamma}_{10}^*(u)h^{-2l-2}(u-u_0)^{2l}K^2((u-u_0)/h)f(u)\mathrm{d}u \\
&= (nh)^{-1}\int_{-1}^{1}\boldsymbol{\Gamma}_{10}^*(u_0+ht)f(u_0+ht)t^{2l}K^2(t)\,\mathrm{d}t \\
&= (nh)^{-1}f(u_0)\boldsymbol{\Gamma}_{10}^*(u_0)\nu_{2l}[1+o(1)] = O((nh)^{-1}).
\end{aligned}
$$

Thus

$$
\mathbf{S}_{n,lG}(u_0) = \mathrm{E}(\mathbf{S}_{n,lG}(u_0)) + O_P\left(\sqrt{\mathrm{Var}(\mathbf{S}_{n,lG}(u_0))}\right) = f(u_0)\boldsymbol{\Gamma}(u_0)\mu_l[1+o_P(1)].
$$

This completes the proof of (5.1). □

Next, we prove (5.2). Note that

$$
W_{i\widehat{G}} - W_{iG} = W_{iG}\frac{\widehat{G}(Z_i-) - G(Z_i-)}{1-G(Z_i-)}\frac{1-G(Z_i-)}{1-\widehat{G}(Z_i-)}
$$

and the fact, which follows immediately from Theorem 2.2 in Zhou (1991), that

$$
\sup_{0\le t\le Z_{(n)}}\left|\frac{\widehat{G}(t-)-G(t-)}{1-G(t-)}\right| = O_P(n^{-1/2}) \tag{5.3}
$$

and

$$
\sup_{0\le t\le Z_{(n)}}\left|\frac{1-G(t-)}{1-\widehat{G}(t-)}\right| = O_P(1), \tag{5.4}
$$

where $Z_{(n)} = \max\{Z_1,\cdots,Z_n\}$. This, together with the conclusion that $\mathbf{S}_{n,lG}(u_0) = O_P(1)$ by (5.1), imply that

$$
\begin{aligned}
\mathbf{S}_{n,l\widehat{G}}(u_0) - \mathbf{S}_{n,lG}(u_0) &= \frac{1}{n}\sum_{i=1}^{n}(W_{i\widehat{G}}-W_{iG})\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}h^{-l}(U_i-u_0)^l K_h(U_i-u_0) \\
&\le \mathbf{S}_{n,lG}(u_0)\sup_{0\le t\le Z_{(n)}}\left|\frac{\widehat{G}(t-)-G(t-)}{1-G(t-)}\right|\sup_{0\le t\le Z_{(n)}}\left|\frac{1-G(t-)}{1-\widehat{G}(t-)}\right| \\
&= O_P(n^{-1/2}) = o_P(1).
\end{aligned}
$$

Hence

$$
\mathbf{S}_{n,l\widehat{G}}(u_0) = (\mathbf{S}_{n,l\widehat{G}}(u_0) - \mathbf{S}_{n,lG}(u_0)) + \mathbf{S}_{n,lG}(u_0) = \mathbf{S}_{n,lG}(u_0)[1+o_P(1)].
$$

This along with (5.1) yield

$$
\mathbf{S}_{n\widehat{G}}(u_0) = \mathbf{S}_{nG}(u_0)[1+o_P(1)] = f(u_0)\boldsymbol{\Gamma}(u_0)\otimes\begin{pmatrix}1 & 0 \\ 0 & \mu_2\end{pmatrix}[1+o_P(1)].
$$

As a consequence of Lemma 5.1, we can show that

$$\mathbf{S}_n(u_0) = \begin{pmatrix} \mathbf{S}_{n,0}(u_0) & \mathbf{S}_{n,1}(u_0) \\ \mathbf{S}_{n,1}(u_0) & \mathbf{S}_{n,2}(u_0) \end{pmatrix} = f(u_0)\mathbf{\Gamma}(u_0)\otimes \begin{pmatrix} 1 & 0 \\ 0 & \mu_2 \end{pmatrix}[1 + o_P(1)]. \quad (5.5)$$

Lemma 5.2. *Under the conditions C1–C6, we have*

$$\mathbf{T}_{nG}(u_0) = \begin{pmatrix} \mathbf{T}_{n,0G}(u_0) \\ \mathbf{T}_{n,1G}(u_0) \end{pmatrix} = f(u_0)\mathbf{\Gamma}(u_0)\mathbf{a}(u_0)\otimes \begin{pmatrix} 1 \\ 0 \end{pmatrix}[1 + o_P(1)] \quad (5.6)$$

and

$$\mathbf{T}_{n\widehat{G}}(u_0) = \begin{pmatrix} \mathbf{T}_{n,0\widehat{G}}(u_0) \\ \mathbf{T}_{n,1\widehat{G}}(u_0) \end{pmatrix} = \mathbf{T}_{nG}(u_0) + f(u_0)\mathbf{\Gamma}(u_0)\mathbf{a}(u_0)\otimes \begin{pmatrix} 1 \\ 0 \end{pmatrix} O_P(n^{-1/2}).$$
$$(5.7)$$

Proof: It is easy to show that $\mathrm{E}(Z_G|U,\mathbf{X}) = \mathrm{E}(Y|U,\mathbf{X}) = \mathbf{X}^{\mathrm{T}}\mathbf{a}(U)$ and $\mathrm{E}(Z_G^2|U,\mathbf{X}) = \mathrm{E}(Y^2\{1 - G(Y)\}^{-1}|U,\mathbf{X}) = b_2(U,\mathbf{X})$. This implies that

$$
\begin{aligned}
\mathrm{E}(\mathbf{T}_{n,lG}(u_0)) &= \mathrm{E}[Z_G\mathbf{X}h^{-l}(U - u_0)^l K_h(U - u_0)] \\
&= \mathrm{E}[\mathrm{E}(Z_G|U,\mathbf{X})\mathbf{X}h^{-l}(U - u_0)^l K_h(U - u_0)] \\
&= \mathrm{E}[\mathbf{X}\mathbf{X}^{\mathrm{T}}\mathbf{a}(U)h^{-l}(U - u_0)^l K_h(U - u_0)] \\
&= \mathrm{E}[\mathrm{E}(\mathbf{X}\mathbf{X}^{\mathrm{T}}|U)\mathbf{a}(U)h^{-l}(U - u_0)^l K_h(U - u_0)] \\
&= \mathrm{E}[\mathbf{\Gamma}(U)\mathbf{a}(U)h^{-l}(U - u_0)^l K_h(U - u_0)] \\
&= \int_{u_0-h}^{u_0+h} \mathbf{\Gamma}(u)\mathbf{a}(u)h^{-l-1}(u - u_0)^l K((u - u_0)/h)f(u)\,\mathrm{d}u \\
&= \int_{-1}^{1} \mathbf{\Gamma}(u_0 + ht)\mathbf{a}(u_0 + ht)f(u_0 + ht)t^l K(t)\,\mathrm{d}t \\
&= f(u_0)\mathbf{\Gamma}(u_0)\mathbf{a}(u_0)\mu_l[1 + o(1)]
\end{aligned}
$$

and

$$
\begin{aligned}
\mathrm{Var}(\mathbf{T}_{n,lG}(u_0)) &= n^{-1}\mathrm{Var}[Z_G\mathbf{X}h^{-l}(U - u_0)^l K_h(U - u_0)] \\
&\leq n^{-1}\mathrm{E}[Z_G^2\mathbf{X}\mathbf{X}^{\mathrm{T}}h^{-2l}(U - u_0)^{2l} K_h^2(U - u_0)] \\
&= n^{-1}\mathrm{E}[\mathrm{E}(Z_G^2|U,\mathbf{X})\mathbf{X}\mathbf{X}^{\mathrm{T}}h^{-2l}(U - u_0)^{2l} K_h^2(U - u_0)] \\
&= n^{-1}\mathrm{E}[b_2(U,\mathbf{X})\mathbf{X}\mathbf{X}^{\mathrm{T}}h^{-2l}(U - u_0)^{2l} K_h^2(U - u_0)] \\
&= n^{-1}\mathrm{E}[\mathrm{E}(b_2(U,\mathbf{X})\mathbf{X}\mathbf{X}^{\mathrm{T}}|U)h^{-2l}(U - u_0)^{2l} K_h^2(U - u_0)] \\
&= n^{-1}\mathrm{E}[\mathbf{\Gamma}_2^*(U)h^{-2l}(U - u_0)^{2l} K_h^2(U - u_0)] \\
&= n^{-1}\int_{u_0-h}^{u_0+h} \mathbf{\Gamma}_2^*(u)h^{-2l-2}(u - u_0)^{2l} K^2((u - u_0)/h)f(u)\,\mathrm{d}u \\
&= (nh)^{-1}\int_{-1}^{1} \mathbf{\Gamma}_2^*(u_0 + ht)f(u_0 + ht)t^{2l} K^2(t)\,\mathrm{d}t \\
&= (nh)^{-1}f(u_0)\mathbf{\Gamma}_2^*(u_0)\nu_{2l}[1 + o(1)] = O((nh)^{-1}).
\end{aligned}
$$

Combining the last two results yields $\mathbf{T}_{n,lG}(u_0) = f(u_0)\mathbf{\Gamma}(u_0)\mathbf{a}(u_0)\mu_l[1 + o_P(1)]$. This proves (5.6).

We next prove (5.7). Observe that

$$Z_{i\widehat{G}} - Z_{iG} = Z_{iG}\frac{\widehat{G}(Z_i-) - G(Z_i-)}{1 - G(Z_i-)}\frac{1 - G(Z_i-)}{1 - \widehat{G}(Z_i-)}.$$

This, together with (5.3), (5.4) and (5.6), imply that

$$\mathbf{T}_{n,l\widehat{G}}(u_0) - \mathbf{T}_{n,lG}(u_0) = \frac{1}{n}\sum_{i=1}^{n}(Z_{i\widehat{G}} - Z_{iG})\mathbf{X}_i h^{-l}(U_i - u_0)^l K_h(U_i - u_0)$$

$$\leq \mathbf{T}_{n,lG}(u_0)\sup_{0\leq t\leq Z_{(n)}}\left|\frac{\widehat{G}(t-) - G(t-)}{1 - G(t-)}\right|\sup_{0\leq t\leq Z_{(n)}}\left|\frac{1 - G(t-)}{1 - \widehat{G}(t-)}\right|$$

$$= f(u_0)\mathbf{\Gamma}(u_0)\mathbf{a}(u_0)\mu_l O_P(n^{-1/2}).$$

This completes the proof of (5.7). □

Proof of Theorem 2.1: Since the coefficient functions $a_1(u), \cdots, a_p(u)$ are all smooth in the neighborhood of $|U_i - u_0| < h$, it follows from the Taylor's expansion that

$$\mathbf{X}_i^{\mathrm{T}}\mathbf{a}(U_i) = \mathbf{X}_i^{\mathrm{T}}\mathbf{a}(u_0) + \left(\frac{U_i - u_0}{h}\right)\mathbf{X}_i^{\mathrm{T}}h\mathbf{a}'(u_0)$$

$$+ \frac{h^2}{2}\left(\frac{U_i - u_0}{h}\right)^2\mathbf{X}_i^{\mathrm{T}}\mathbf{a}''(u_0) + o_P(h^2).$$

Then

$$\mathbf{T}_{n,0G}(u_0) - \mathbf{T}_{n,0G}^*(u_0) = \mathbf{S}_{n,0}(u_0)\mathbf{a}(u_0) + h\mathbf{S}_{n,1}(u_0)\mathbf{a}'(u_0)$$

$$+ \frac{h^2}{2}\mathbf{S}_{n,2}(u_0)\mathbf{a}''(u_0) + o_P(h^2)$$

and

$$\mathbf{T}_{n,1G}(u_0) - \mathbf{T}_{n,1G}^*(u_0) = \mathbf{S}_{n,1}(u_0)\mathbf{a}(u_0) + h\mathbf{S}_{n,2}(u_0)\mathbf{a}'(u_0)$$

$$+ \frac{h^2}{2}\mathbf{S}_{n,3}(u_0)\mathbf{a}''(u_0) + o_P(h^2).$$

Combining the last two results with (5.5) gives

$$\mathbf{T}_{nG}(u_0) - \mathbf{T}_{nG}^*(u_0)$$

$$= \begin{pmatrix} \mathbf{T}_{n,0G}(u_0) - \mathbf{T}_{n,0G}^*(u_0) \\ \mathbf{T}_{n,1G}(u_0) - \mathbf{T}_{n,1G}^*(u_0) \end{pmatrix}$$

$$= \begin{pmatrix} \mathbf{S}_{n,0}(u_0) & \mathbf{S}_{n,1}(u_0) \\ \mathbf{S}_{n,1}(u_0) & \mathbf{S}_{n,2}(u_0) \end{pmatrix}\begin{pmatrix} \mathbf{a}(u_0) \\ h\mathbf{a}'(u_0) \end{pmatrix} + \frac{h^2}{2}\begin{pmatrix} \mathbf{S}_{n,2}(u_0) \\ \mathbf{S}_{n,3}(u_0) \end{pmatrix}\mathbf{a}''(u_0) + o_P(h^2)$$

$$= f(u_0)\mathbf{\Gamma}(u_0)\otimes\begin{pmatrix} 1 & 0 \\ 0 & \mu_2 \end{pmatrix}\mathbf{\Psi}(u_0) + \frac{h^2}{2}f(u_0)\mathbf{\Gamma}(u_0)\otimes\begin{pmatrix} \mu_2 \\ 0 \end{pmatrix}\mathbf{a}''(u_0) + o_P(h^2) \quad (5.8)$$

Since $W_{in}^* = \frac{1}{n}\frac{\delta_i}{1-\widehat{G}(Z_i-)} = \frac{1}{n}W_{i\widehat{G}}$ (for details, see Satten and Datta, 2001), it is easy to show that the minimizer of (2.7), that is $\widehat{\mathbf{\Psi}}(u_0)$, can be re-writen as

$$
\begin{aligned}
\widehat{\mathbf{\Psi}}(u_0) &= \begin{pmatrix} \mathbf{S}_{n,0\widehat{G}}(u_0) & \mathbf{S}_{n,1\widehat{G}}(u_0) \\ \mathbf{S}_{n,1\widehat{G}}(u_0) & \mathbf{S}_{n,2\widehat{G}}(u_0) \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{T}_{n,0\widehat{G}}(u_0) \\ \mathbf{T}_{n,1\widehat{G}}(u_0) \end{pmatrix} \\
&= \mathbf{S}_{n\widehat{G}}^{-1}(u_0)\mathbf{T}_{n\widehat{G}}(u_0) \\
&= \mathbf{S}_{n\widehat{G}}^{-1}(u_0)[(\mathbf{T}_{n\widehat{G}}(u_0) - \mathbf{T}_{nG}(u_0)) + (\mathbf{T}_{nG}(u_0) - \mathbf{T}_{nG}^*(u_0)) + \mathbf{T}_{nG}^*(u_0)] \\
&= \mathbf{S}_{n\widehat{G}}^{-1}(u_0)\mathbf{T}_{nG}^*(u_0) + \mathbf{S}_{n\widehat{G}}^{-1}(u_0)(\mathbf{T}_{nG}(u_0) - \mathbf{T}_{nG}^*(u_0)) + \\
&\quad \mathbf{S}_{n\widehat{G}}^{-1}(u_0)(\mathbf{T}_{n\widehat{G}}(u_0) - \mathbf{T}_{nG}(u_0)) \\
&= I_1 + I_2 + I_3. \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (5.9)
\end{aligned}
$$

By (5.2) and (5.8), we obtain

$$
I_2 = \mathbf{\Psi}(u_0) + 2^{-1}h^2\mu_2\begin{pmatrix} \mathbf{a}''(u_0) \\ \mathbf{0} \end{pmatrix} + o_P(h^2). \qquad\qquad (5.10)
$$

It follows from (5.2) and (5.7) that

$$
I_3 = \mathbf{a}(u_0)\otimes(1,0)^{\mathrm{T}}[1 + o_P(1)]O_P(n^{-1/2}) = O_P(n^{-1/2}). \qquad (5.11)
$$

By (5.9), (5.10), and (5.11), we have

$$
\sqrt{nh}\left[\widehat{\mathbf{\Psi}}(u_0) - \mathbf{\Psi}(u_0) - 2^{-1}h^2\mu_2\begin{pmatrix} \mathbf{a}''(u_0) \\ \mathbf{0} \end{pmatrix} + o_P(h^2)\right] = \sqrt{nh}I_1 + o_P\left(\sqrt{h}\right).
$$
$$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (5.12)$$

Noting that

$$
\mathbf{S}_{n\widehat{G}}^{-1}(u_0) = f(u_0)^{-1}\mathbf{\Gamma}(u_0)^{-1}\otimes\begin{pmatrix} 1 & 0 \\ 0 & \mu_2^{-1} \end{pmatrix}[1 + o_P(1)]
$$

and $o_P\left(\sqrt{h}\right) = o_P(1)$ by the condition C5, it suffices to prove Theorem 2.1 by showing that

$$
\sqrt{nh}\mathbf{T}_{nG}^*(u_0) = n^{-\frac{1}{2}}h^{\frac{1}{2}}\sum_{i=1}^{n}\begin{pmatrix} \mathbf{X}_iK_h(U_i - u_0)\varepsilon_{iG} \\ \mathbf{X}_i\left(\frac{U_i-u_0}{h}\right)K_h(U_i - u_0)\varepsilon_{iG} \end{pmatrix} \xrightarrow{\mathrm{D}} N(\mathbf{0}, \mathbf{\Sigma}_2(u_0)),
$$
$$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (5.13)$$

where

$$
\mathbf{\Sigma}_2(u_0) = f(u_0)\mathbf{\Gamma}^*(u_0)\otimes\begin{pmatrix} \nu_0 & 0 \\ 0 & \nu_2 \end{pmatrix}.
$$

Let $\mathbf{c} = (\mathbf{c}_1^{\mathrm{T}}, \mathbf{c}_2^{\mathrm{T}})^{\mathrm{T}}$ be an arbitrary unit vector with $\mathbf{c}_1$ and $\mathbf{c}_2$ both being the $p\times 1$ vectors, and

$$
\begin{aligned}
\zeta_i &= n^{-\frac{1}{2}}h^{\frac{1}{2}}\mathbf{c}^{\mathrm{T}}\begin{pmatrix} \mathbf{X}_iK_h(U_i - u_0)\varepsilon_{iG} \\ \mathbf{X}_i\left(\frac{U_i-u_0}{h}\right)K_h(U_i - u_0)\varepsilon_{iG} \end{pmatrix} \\
&= n^{-\frac{1}{2}}h^{\frac{1}{2}}\varepsilon_{iG}\mathbf{X}_i^{\mathrm{T}}(\mathbf{c}_1 + \mathbf{c}_2(U_i - u_0)/h)K_h(U_i - u_0).
\end{aligned}
$$

By the Cramér-Wold device, it suffices to prove

$$
q_n = \sqrt{nh}\mathbf{c}^{\mathrm{T}}\mathbf{T}_{nG}^*(u_0) = \sum_{i=1}^{n}\zeta_i \xrightarrow{\mathrm{D}} N\left(0, \mathbf{c}^{\mathrm{T}}\mathbf{\Sigma}_2(u_0)\mathbf{c}\right). \qquad (5.14)
$$

Since q_n is the sum of independent and nonidentically distributed random variables and has $\mathrm{E}(q_n) = 0$ and

$$
\begin{aligned}
\mathrm{Var}(q_n) \;=\;& h\mathbf{c}^{\mathrm{T}}\mathrm{E}\left[\varepsilon_G^2 \begin{pmatrix} \mathbf{X}\mathbf{X}^{\mathrm{T}} & \mathbf{X}\mathbf{X}^{\mathrm{T}}\left(\frac{U-u_0}{h}\right) \\ \mathbf{X}\mathbf{X}^{\mathrm{T}}\left(\frac{U-u_0}{h}\right) & \mathbf{X}\mathbf{X}^{\mathrm{T}}\left(\frac{U-u_0}{h}\right)^2 \end{pmatrix} K_h^2(U-u_0)\right]\mathbf{c} \\
\;=\;& h\mathbf{c}^{\mathrm{T}}\mathrm{E}\left[\mathrm{E}(\varepsilon_G^2|U,\mathbf{X}) \begin{pmatrix} \mathbf{X}\mathbf{X}^{\mathrm{T}} & \mathbf{X}\mathbf{X}^{\mathrm{T}}\left(\frac{U-u_0}{h}\right) \\ \mathbf{X}\mathbf{X}^{\mathrm{T}}\left(\frac{U-u_0}{h}\right) & \mathbf{X}\mathbf{X}^{\mathrm{T}}\left(\frac{U-u_0}{h}\right)^2 \end{pmatrix} K_h^2(U-u_0)\right]\mathbf{c}
\end{aligned}
$$

$$
\begin{aligned}
\;=\;& h\mathbf{c}^{\mathrm{T}}\mathrm{E}\left[\sigma^{*2}(U,\mathbf{X}) \begin{pmatrix} \mathbf{X}\mathbf{X}^{\mathrm{T}} & \mathbf{X}\mathbf{X}^{\mathrm{T}}\left(\frac{U-u_0}{h}\right) \\ \mathbf{X}\mathbf{X}^{\mathrm{T}}\left(\frac{U-u_0}{h}\right) & \mathbf{X}\mathbf{X}^{\mathrm{T}}\left(\frac{U-u_0}{h}\right)^2 \end{pmatrix} K_h^2(U-u_0)\right]\mathbf{c} \\
\;=\;& h\mathbf{c}^{\mathrm{T}}\mathrm{E}\left[\mathrm{E}(\sigma^{*2}(U,\mathbf{X})\mathbf{X}\mathbf{X}^{\mathrm{T}}|U) \begin{pmatrix} 1 & \left(\frac{U-u_0}{h}\right) \\ \left(\frac{U-u_0}{h}\right) & \left(\frac{U-u_0}{h}\right)^2 \end{pmatrix} K_h^2(U-u_0)\right]\mathbf{c} \\
\;=\;& f(u_0)\mathbf{c}^{\mathrm{T}}\mathbf{\Gamma}^*(u_0)\otimes \begin{pmatrix} \nu_0 & 0 \\ 0 & \nu_2 \end{pmatrix}\mathbf{c}[1+o(1)] \\
\;=\;& \mathbf{c}^{\mathrm{T}}\mathbf{\Sigma}_2(u_0)\mathbf{c}[1+o(1)],
\end{aligned}
$$

it suffices to prove (5.14) by showing that the Lyapounov condition holds, that is for some $\gamma > 0$,

$$
\sum_{i=1}^{n}\mathrm{E}(|\zeta_i|^{2+\gamma}) = \sum_{i=1}^{n}\mathrm{E}(|n^{-\frac{1}{2}}h^{\frac{1}{2}}\varepsilon_{iG}\mathbf{X}_i^{\mathrm{T}}(\mathbf{c}_1+\mathbf{c}_2(U_i-u_0)/h)K_h(U_i-u_0)|^{2+\gamma}) = o(1).
$$

(5.15)

It follows from the condition C7 that

$$
\begin{aligned}
& \mathrm{E}(|n^{-\frac{1}{2}}h^{\frac{1}{2}}\varepsilon_{iG}\mathbf{X}_i^{\mathrm{T}}(\mathbf{c}_1+\mathbf{c}_2(U_i-u_0)/h)K_h(U_i-u_0)|^{2+\gamma}) \\
=\;& n^{-\frac{2+\gamma}{2}}h^{\frac{2+\gamma}{2}}\mathrm{E}(|\varepsilon_{iG}\mathbf{X}_i^{\mathrm{T}}(\mathbf{c}_1+\mathbf{c}_2(U_i-u_0)/h)K_h(U_i-u_0)|^{2+\gamma}) \\
=\;& n^{-\frac{2+\gamma}{2}}h^{\frac{2+\gamma}{2}}\mathrm{E}(\mathrm{E}(|\varepsilon_{iG}|^{2+\gamma}|U_i,\mathbf{X}_i)|\mathbf{X}_i^{\mathrm{T}}(\mathbf{c}_1+\mathbf{c}_2(U_i-u_0)/h)K_h(U_i-u_0)|^{2+\gamma}) \\
=\;& n^{-\frac{2+\gamma}{2}}h^{\frac{2+\gamma}{2}}O(\mathrm{E}(|\mathbf{X}_i^{\mathrm{T}}(\mathbf{c}_1+\mathbf{c}_2(U_i-u_0)/h)K_h(U_i-u_0)|^{2+\gamma})). \quad\quad (5.16)
\end{aligned}
$$

By the C_r inequality, we can show that

$$\mathrm{E}(|\mathbf{X}_i^{\mathrm{T}}(\mathbf{c}_1 + \mathbf{c}_2(U_i - u_0)/h)K_h(U_i - u_0)|^{2+\gamma})$$

$$\leq 2^{1+\gamma}\mathrm{E}(|\mathbf{X}_i^{\mathrm{T}}\mathbf{c}_1 K_h(U_i - u_0)|^{2+\gamma}) + 2^{1+\gamma}\mathrm{E}(|\mathbf{X}_i^{\mathrm{T}}\mathbf{c}_2 h^{-1}(U_i - u_0)K_h(U_i - u_0)|^{2+\gamma})$$

$$= 2^{1+\gamma}\mathrm{E}(\mathrm{E}(|\mathbf{X}_i^{\mathrm{T}}\mathbf{c}_1|^{2+\gamma}|U_i)K_h^{2+\gamma}(U_i - u_0)) +$$

$$\quad 2^{1+\gamma}\mathrm{E}(\mathrm{E}(|\mathbf{X}_i^{\mathrm{T}}\mathbf{c}_2|^{2+\gamma}|U_i)h^{-2-\gamma}|U_i - u_0|^{2+\gamma}K_h^{2+\gamma}(U_i - u_0))$$

$$= 2^{1+\gamma}\int_{u_0-h}^{u_0+h}\mathrm{E}(|\mathbf{X}_i^{\mathrm{T}}\mathbf{c}_1|^{2+\gamma}|U_i = u)h^{-2-\gamma}K^{2+\gamma}((u - u_0)/h)f(u)\,\mathrm{d}u +$$

$$\quad 2^{1+\gamma}\int_{u_0-h}^{u_0+h}\mathrm{E}(|\mathbf{X}_i^{\mathrm{T}}\mathbf{c}_2|^{2+\gamma}|U_i = u)h^{-2-\gamma}|(u - u_0)/h|^{2+\gamma}K^{2+\gamma}((u - u_0)/h)f(u)\,\mathrm{d}u$$

$$= 2^{1+\gamma}h^{-1-\gamma}f(u_0)\mathrm{E}(|\mathbf{X}_i^{\mathrm{T}}\mathbf{c}_1|^{2+\gamma}|U_i = u_0)\int_{-1}^{1}K^{2+\gamma}(t)\,\mathrm{d}t[1 + o(1)] +$$

$$\quad 2^{1+\gamma}h^{-1-\gamma}f(u_0)\mathrm{E}(|\mathbf{X}_i^{\mathrm{T}}\mathbf{c}_2|^{2+\gamma}|U_i = u_0)\int_{-1}^{1}|t|^{2+\gamma}K^{2+\gamma}(t)\,\mathrm{d}t[1 + o(1)]$$

$$= O(h^{-1-\gamma}).$$

$$(5.17)$$

Combining (5.16) with (5.17) yields

$$\sum_{i=1}^{n}\mathrm{E}(|\zeta_i|^{2+\gamma}) = O(n^{-\frac{\gamma}{2}}h^{-\frac{\gamma}{2}}) = O((nh)^{-\frac{\gamma}{2}}) = o(1).$$

This proves (5.15). Thus, the proof of Theorem 2.1 is completed.

$\square$

Acknowledgements

References

M. G. Akritas. Nearest neighbor estimation of a bivariate distribution under random censoring. *Ann. Statist.* **22** (3), 1299–1327 (1994). MR1311977.

Z. Cai. Weighted local linear approach to censored nonparametric regression. In *Recent advances and trends in nonparametric statistics*, pages 217–231. Elsevier B. V., Amsterdam (2003). MR2498243.

Z. Cai and J. Fan. Average regression surface for dependent data. *J. Multivariate Anal.* **75** (1), 112–142 (2000). MR1787404.

Z. Cai, J. Fan and Q. Yao. Functional-coefficient regression models for nonlinear time series. *J. Amer. Statist. Assoc.* **95** (451), 941–956 (2000). MR1804449.

Z. Cai and E. Ould-Saïd. Local M-estimator for nonparametric time series. *Statist. Probab. Lett.* **65** (4), 433–449 (2003). MR2039887.

Z. Cai and X. Xu. Nonparametric quantile estimations for dynamic smooth coefficient models. *J. Amer. Statist. Assoc.* **104** (485), 371–383 (2009). MR2504383.

R. J. Carroll, J. Fan, I. Gijbels and M. P. Wand. Generalized partially linear single-index models. *J. Amer. Statist. Assoc.* **92** (438), 477–489 (1997). MR1467842.

W. S. Cleveland, E. Grosse and W. M. Shyu. Local regression models. In *Statistical Models in S*, pages 309–376. Wadsworth & Brooks, Pacific Grove, CA (1992).

Z. J. Daye, J. Xie and H. Li. A sparse structured shrinkage estimator for nonparametric varying-coefficient model with an application in genomics. *J. Comput. Graph. Statist.* **21** (1), 110–133 (2012). MR2913359.

R. Engle, C. Granger, J. Rice and A. Weiss. Semiparametric estimates of the relation between weather and electricity sales. *Journal of the American Statistical Association* **81** (394), 310–320 (1986). DOI: 10.2307/22892184.

J. Fan and I. Gijbels. Censored regression: local linear approximations and their applications. *J. Amer. Statist. Assoc.* **89** (426), 560–570 (1994). MR1294083.

J. Fan and I. Gijbels. *Local polynomial modelling and its applications*, volume 66 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London (1996). ISBN 0-412-98321-4. MR1383587.

J. Fan, Y. Ma and W. Dai. Nonparametric independence screening in sparse ultra-high-dimensional varying coefficient models. *J. Amer. Statist. Assoc.* **109** (507), 1270–1284 (2014). MR3265696.

J. Fan, C. Zhang and J. Zhang. Generalized likelihood ratio statistics and Wilks phenomenon. *Ann. Statist.* **29** (1), 153–193 (2001). MR1833962.

J. Fan and W. Zhang. Statistical estimation in varying coefficient models. *Ann. Statist.* **27** (5), 1491–1518 (1999). MR1742497.

J. Fan and W. Zhang. Simultaneous confidence bands and hypothesis testing in varying-coefficient models. *Scand. J. Statist.* **27** (4), 715–731 (2000). MR1804172.

J. H. Friedman. Multivariate adaptive regression splines. *Ann. Statist.* **19** (1), 1–141 (1991). MR1091842.

C. Gu and G. Wahba. Smoothing spline ANOVA with component-wise Bayesian "confidence intervals". *J. Comput. Graph. Statist.* **2** (1), 97–117 (1993). MR1272389.

P. Hall, R. C. L. Wolff and Q. Yao. Methods for estimating a conditional distribution function. *J. Amer. Statist. Assoc.* **94** (445), 154–163 (1999). MR1689221.

W. Härdle and T. M. Stoker. Investigating smooth multiple regression by the method of average derivatives. *J. Amer. Statist. Assoc.* **84** (408), 986–995 (1989). MR1134488.

T. Hastie and R. Tibshirani. *Generalized additive models*, volume 43 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, Ltd., London (1990). ISBN 0-412-34390-8. MR1082147.

T. Hastie and R. Tibshirani. Varying-coefficient models. *J. Roy. Statist. Soc. Ser. B* **55** (4), 757–796 (1993). MR1229881.

J. Z. Huang, C. O. Wu and L. Zhou. Polynomial spline estimation and inference for varying coefficient models with longitudinal data. *Statist. Sinica* **14** (3), 763–788 (2004). MR2087972.

C. M. Hurvich, J. S. Simonoff and C.-L. Tsai. Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **60** (2), 271–293 (1998). MR1616041.

W.-C. Ip, H. Wong and R. Zhang. Generalized likelihood ratio test for varying-coefficient models with different smoothing variables. *Comput. Statist. Data Anal.* **51** (9), 4543–4561 (2007). MR2364464.

H. T. Kim and Y. K. Truong. Nonparametric regression estimates with censored data: local linear smoothers and their applications. *Biometrics* **54** (4), 1434–1444 (1998). MR1671594.

H. Koul, V. Susarla and J. Van Ryzin. Regression analysis with randomly right-censored data. *Ann. Statist.* **9** (6), 1276–1288 (1981). MR630110.

Sue Leurgans. Linear models, random censoring and synthetic data. *Biometrika* **74** (2), 301–309 (1987). MR903130.

O. B. Linton and W. Härdle. Estimation of additive regression models with known links. *Biometrika* **83** (3), 529–540 (1996). MR1423873.

J. Liu, R. Li and R. Wu. Feature selection for varying coefficient models with ultrahigh-dimensional covariates. *J. Amer. Statist. Assoc.* **109** (505), 266–274 (2014). MR3180562.

O. Lopez. Single-index regression models with right-censored responses. *J. Statist. Plann. Inference* **139** (3), 1082–1097 (2009). MR2479851.

O. Lopez and V. Patilea. Nonparametric lack-of-fit tests for parametric mean-regression models with censored data. *J. Multivariate Anal.* **100** (1), 210–230 (2009). MR2460488.

Y. Lu, R. Zhang and L. Zhu. Penalized spline estimation for varying-coefficient models. *Comm. Statist. Theory Methods* **37** (13-15), 2249–2261 (2008). MR2446666.

X. H. Luo, Z. H. Yang and Y. Zhou. Varying-coefficient regression models with censored data. *Acta Math. Appl. Sin.* **29** (3), 415–427 (2006). MR2253127.

J. C. Pardo-Fernández, I. Van Keilegom and W. González-Manteiga. Goodness-of-fit tests for parametric models in censored regression. *Canad. J. Statist.* **35** (2), 249–264 (2007). MR2393608.

A. Qu and R. Li. Quadratic inference functions for varying-coefficient models with longitudinal data. *Biometrics* **62** (2), 379–391 (2006). MR2227487.

D. Ruppert and M. P. Wand. Multivariate locally weighted least squares regression. *Ann. Statist.* **22** (3), 1346–1370 (1994). MR1311979.

G. A. Satten and S. Datta. The Kaplan-Meier estimator as an inverse-probability-of-censoring weighted average. *Amer. Statist.* **55** (3), 207–210 (2001). MR1947266.

W. Stute. Consistent estimation under random censorship when covariables are present. *J. Multivariate Anal.* **45** (1), 89–103 (1993). MR1222607.

W. Stute. Distributional convergence under random censorship when covariables are present. *Scand. J. Statist.* **23** (4), 461–471 (1996). MR1439707.

W. Stute. Nonlinear censored regression. *Statist. Sinica* **9** (4), 1089–1102 (1999). MR1744826.

W. Stute and J.-L. Wang. The strong law under random censorship. *Ann. Statist.* **21** (3), 1591–1607 (1993). MR1241280.

I. Van Keilegom and M. G. Akritas. Transfer of tail information in censored regression models. *Ann. Statist.* **27** (5), 1745–1784 (1999). MR1742508.

H. Wang and Y. Xia. Shrinkage estimation of the varying coefficient model. *J. Amer. Statist. Assoc.* **104** (486), 747–757 (2009). MR2541592.

L. Wang, Bo Kai and R. Li. Local rank inference for varying coefficient models. *J. Amer. Statist. Assoc.* **104** (488), 1631–1645 (2009). MR2597005.

L. Wang, H. Li and J. Z. Huang. Variable selection in nonparametric varying-coefficient models for analysis of repeated measurements. *J. Amer. Statist. Assoc.* **103** (484), 1556–1569 (2008). MR2504204.

H. Wong, W.-C. Ip and R. Zhang. Varying-coefficient single-index model. *Comput. Statist. Data Anal.* **52** (3), 1458–1476 (2008). MR2422748.

W.-L. Xu and L.-X. Zhu. Goodness-of-fit testing for varying-coefficient models. *Metrika* **68** (2), 129–146 (2008). MR2434309.

L. Xue and A. Qu. Variable selection in high-dimensional varying-coefficient models with global optimality. *J. Mach. Learn. Res.* **13**, 1973–1998 (2012). MR2956349.

L. Xue and L. Zhu. Empirical likelihood for a varying coefficient model with longitudinal data. *J. Amer. Statist. Assoc.* **102** (478), 642–654 (2007). MR2370858.

S. J. Yang, A. El Ghouch and I. Van Keilegom. Varying coefficient models having different smoothing variables with randomly censored data. *Electron. J. Stat.* **8** (1), 226–252 (2014). MR3189554.

R. Zhang and G. Li. Averaged estimation of functional-coefficient regression models with different smoothing variables. *Statist. Probab. Lett.* **77** (4), 455–461 (2007). MR2339051.

W. Zhang, S.-Y. Lee and X. Song. Local polynomial fitting in semivarying coefficient model. *J. Multivariate Anal.* **82** (1), 166–188 (2002). MR1918619.

W. Zhang and H. Peng. Simultaneous confidence band and hypothesis test in generalised varying-coefficient models. *J. Multivariate Anal.* **101** (7), 1656–1680 (2010). MR2610738.

P. Zhao and L. Xue. Variable selection for varying coefficient models with measurement errors. *Metrika* **74** (2), 231–245 (2011). MR2822159.

M. Zhou. Some properties of the Kaplan-Meier estimator for independent non-identically distributed random variables. *Ann. Statist.* **19** (4), 2266–2274 (1991). MR2434309.

M. Zhou. *M*-estimation in censored linear models. *Biometrika* **79** (4), 837–841 (1992). MR1209482.