Empirical likelihood and estimation in single-index varying-coefficient models with censored data

Liugen Xue (lgxue@bjut.edu.cn)
Henan University

Research Article

Keywords: Single-index varying-coefficient model, Empirical likelihood, Bias correction method, Confidence region, Pointwise confidence interval

Posted Date: January 2nd, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2417349/v1

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Statistics and Computing on October 19th, 2023. See the published version at https://doi.org/10.1007/s11222-023-10310-7.

Empirical likelihood and estimation in single-index varying-coefficient models with censored data

Liugen Xue

School of Mathematics and Statistics, Henan University, Kaifeng 475004, Henan, China

Abstract In this paper, we investigate the empirical likelihood and estimation of parameters of interest in single-index varying coefficient models with right censored data. A bias-corrected empirical log-likelihood ratio statistic for the regression parameter is proposed. It is shown the the statistic is asymptotically standard chi-squared, and thus the confidence region of the regression parameter is constructed. The estimators for both the regression parameter and the coefficient functions are constructed, their asymptotic distributions are obtained, and the consistent estimators for the asymptotic variances are given. The obtained results can be directly used to construct the confidence regions of the regression parameter and the pointwise confidence intervals of the coefficient functions. Our approach is to directly calibrate the empirical log-likelihood ratio, so that the resulting ratio is asymptotically chi-squared, undersmoothing of the coefficient functions is avoided, and the existing data-driven methods can effectively select the optimal bandwidth. The finite-sample behavior of the new methods is evaluated through simulation studies, and applications to a real data are illustrated.

Keywords Single-index varying-coefficient model; Empirical likelihood; Bias correction method; Confidence region; Pointwise confidence interval.

---

[0]Corresponding author: Liugen Xue

[0]E-mail address: lgxue@bjut.edu.cn

1 Introduction

Consider the single-index varying-coefficient model of the form

$$Y = \boldsymbol{g}^T(\beta_0^T X)Z + \varepsilon, \tag{1.1}$$

where $(X, Z) \in R^p \times R^q$ are covariates of the response variable Y, β_0 is a $p \times 1$ vector of unknown parameters, $\boldsymbol{g}(\cdot)$ is a $q \times 1$ vector of unknown coefficient functions, ε is a random error with $E(\varepsilon|X, Z) = 0$ almost surely. To assure the identifiability of model (1.1), we assume that $\|\beta_0\| = 1$ and that the first nonzero component of β_0 is positive, where $\|\cdot\|$ denotes the Euclidean metric. Generally, the first component of Z may be 1.

This paper assumes that response Y is a positive random variable and Y is right censored. That is, instead of Y, we observe $V = \min\{Y, C\}$ and the indicator $\Delta = I(Y \leq C)$ of the event $(Y \leq C)$, where C is a random variable with distribution function $G(t) = P(C \leq t)$. Let Y has distribution function $F(y) = P(Y \leq y)$, and let $G(t)$ and $F(y)$ be continuous but unknown. Suppose that, given X and Z, the variable C is independent of Y. For the distribution function $G(t)$, let $\overline{G}(t) = 1 - G(t)$ and $b_G = \sup\{t|\, G(t) < 1\}$. The two symbols can also be used for other distribution functions. We assume throughout the paper that $b_F \leq b_G$.

The major advantage of (1.1) is that it does not suffer from the curse of dimensionality which is often encountered in multivariate nonparametric settings, since $\boldsymbol{g}(\cdot)$ is a vector of univariate functions. Model (1.1) includes a class of important statistical models. Several special examples of this model are given below. If $Z = 1$, (1.1) reduces to the single-index models, which were investigated by many scholars, see for example, Härdle et al. (1993), Ichimura (1993), Weisberg and Welsh (1994), Zhu and Fang (1996), Chiou and Müller (1998), Hristache et al. (2001), Xue and Zhu (2006), and Cui et al. (2011). If $\beta_0 = 1$, (1.1) becomes the varying-coefficient models studied by many scholars, some works include: Chen and Tsay (1993), Hastie and Tibshirani (1993), Wu et al. (1998), Fan and Zhang (1999), Cai et al. (2000a, 2000b), and Xue and Zhu (2007). If the last component of β_0 to be non-zero and $Z = (1, \breve{X}^T)^T$ where $\breve{X}$ is the remaining vector of X with its pth component deleted, (1.1) becomes the adaptive varying-coefficient linear model proposed by Fan et al. (2003). The model was also studied by Lu et al. (2007). Xia and Li (1999) considered the generic case for this model. owing to model (1.1) takes into account the characteristics of single exponential model and variable coefficient model, so it is easier to interpret in real application.

In no censored case, Xue and Wang (2012) developed statistical inference techniques for the unknown coefficient functions and regression parameter in model (1.1). We first estimated the coefficient functions via the local linear fitting, then constructed the estimated and adjusted empirical likelihood (EL) ratios and a maximum EL estimator for the regression parameter, and proved their asymptotic properties. Xue and Pang (2013) investigated the estimators of parameters of interest for model (1.1). We constructed the estimators for the regression parameter and the coefficient functions, and proved their asymptotic properties. Huang and Zhang (2013) studied the profile EL inferences for model (1.1). They constructed the profile EL ratio for each component of the relevant parameter as well as the corrected EL ratio for coefficient functions, and proved the resulting statistics are asymptotically chi-squared.

It is interesting to model the right censored data with model (1.1). Since the semi-parametric regression model with the censored data has both parametric components and unknown functions, and the distribution function $G(t)$ is also unknown, it needs to replace it with its estimator when constructing EL ratio, which brings a bias. Without eliminating the bias, the constructed EL ratio is not standard chi-square, and it cannot be directly used to construct the region of the parameter of interest. Therefore, we need to bias-correct the EL ratio. That is our motivation to study this question. This paper aims to propose a bias correction method to construct the EL ratio and estimation of the parameters of interest in model (1.1). We propose a bias-corrected empirical log-likelihood ratio statistic for the regression parameter, and prove Wilks' phenomenon. Therefore, the confidence region of the regression parameter can be constructed. We construct the estimators of the regression parameter and the coefficient functions. Their asymptotic distributions are obtained, and the consistent estimators of the asymptotic variance are presented. The obtained results can be directly used to construct the confidence region of the regression parameter and the pointwise confidence intervals of the coefficient functions.

The following two desired features deserve mentioning. First, we directly calibrate the EL ratio so that the resulting EL ratio is asymptotically chi-squared. The ratio does not need to be multiplied by an adjustment factor. This avoids estimating the unknown adjustment factor, and thus improves the accuracy of the confidence region/interval. Second, by using bias correction method in constructing the EL ratio and estimator, undersmoothing of the coefficient functions is avoided, so that existing data-driven algorithms can be used to select the optimal bandwidth of the estimators of coefficient functions.

The structure of the rest of this paper is as follows. Section 2 is the methodology, which constructs the bias-corrected EL ratio and the maximum EL estimate of the regression parameter, and constructs a local linear estimates for the coefficient functions. Section 3 gives some theoretical results. Section 4 shows the simulation studies and a real data analysis. Section 5 is the conclusion remark. The proofs of the theorems are placed in Appendix A.

2 Methodology

In this section, we first construct a local linear estimator of the coefficient function $\boldsymbol{g}(u)$, then construct a bias-corrected EL ratio and a maximum EL estimator for the regression parameter β_0. Throughout this paper, we assume that the sample $\{(X_i, Z_i, V_i, \Delta_i), 1 \le i \le n\}$ from (X, Z, V, Δ) are independent and identically distributed. Then, model (1.1) can be written as

$$Y_i = \boldsymbol{g}^T(\beta_0^T X_i) Z_i + \varepsilon_i, \quad i = 1, \ldots, n, \tag{2.1}$$

where $X_i = (X_{i1}, \ldots, X_{ip})^T$, $Z_i = (Z_{i1}, \ldots, Z_{iq})^T$, and $E(\varepsilon_i | X_i, Z_i) = 0$ almost surely. Write $W = (X^T, Z^T)^T$ and $W_i = (X_i^T, Z_i^T)^T$, $i = 1, \ldots, n$.

2.1 Bias-corrected EL ratio

Let $\mathcal{B} = \{\beta \in R^p : \|\beta\| = 1, \text{ and the first non-zero element is positive}\}$. Then β_0 is an inner point of the set $\mathcal{B}$. Therefore, we only need to search for β_0 over $\mathcal{B}$. It is easy to obtain $E\left[\{\Delta/\overline{G}(V)\}\{V - \boldsymbol{g}^T(\beta_0^T X)Z\}|X, Z\right] = E(\varepsilon|X, Z) = 0$. If $\boldsymbol{g}(\cdot)$ is known, the single-index direction β_0 minimizes

$$Q(\beta) \equiv E\left[\{\Delta/\overline{G}(V)\}\{V - \boldsymbol{g}^T(\beta^T X)Z\}\right]^2 \quad \text{subject to} \quad \|\beta\| = 1. \tag{2.2}$$

If we use the Newton's algorithm to find the minimum point of $Q(\beta)$, we need to calculate the derivative of $Q(\beta)$ at point β_0. However, each component of $\boldsymbol{g}(\beta^T X)$ has no a derivative at the point β_0, because $\|\beta_0\| = 1$ means that β_0 is the boundary point on the unit sphere. For this, we suggest to reparametrize β_0, so that we can search the direction β_0 over a region in the Euclidean space R^{p-1}. A "reparametrization" method for β_0 can be applied; see, for example Yu and Ruppert (2002), Wang et al. (2010), and Xue and Pang (2013). Without loss of generality, we may assume that the rth component for the true parameter β_0 is positive. For $\beta = (\beta_1, \ldots, \beta_p)^T$, let $\beta^{(r)} = (\beta_1, \ldots, \beta_{r-1}, \beta_{r+1}, \ldots, \beta_p)^T$ be a $p-1$ dimensional parameter vector after removing the rth component β_r in β. Then, the new parameter $\beta_0^{(r)}$ must satisfy the constraint $\|\beta_0^{(r)}\| < 1$, and hence β is infinitely differentiable in a neighborhood of $\beta_0^{(r)}$. Noting that $\beta_r = (1 - \|\beta^{(r)}\|^2)^{1/2}$, the Jacobian matrix of β with respect to $\beta^{(r)}$ is obtained by

$$J_{\beta^{(r)}} = \frac{\partial \beta}{\partial \beta^{(r)}} = (\gamma_1, \ldots, \gamma_p)^T, \tag{2.3}$$

where γ_s $(1 \leq s \leq p, s \neq r)$ is a $p-1$ dimensional unit vector with sth component 1, and $\gamma_r = -(1 - \|\beta^{(r)}\|^2)^{-1/2}\beta^{(r)}$.

Now we consider the issue of minimizing $Q(\beta)$ in (2.2). It is clear that minimizing $Q(\beta)$ is equivalent to solving the estimating equations

$$\begin{cases} E\left[\{\Delta/\overline{G}(V)\}\{V - \boldsymbol{g}^T(\beta^T X)Z\}\dot{\boldsymbol{g}}^T(\beta^T X)ZJ^T_{\beta^{(r)}}X\right] = 0, \\ \|\beta\| - 1 = 0, \end{cases} \tag{2.4}$$

where $\dot{\boldsymbol{g}}(\cdot)$ stands for the derivative of the coefficient function $\boldsymbol{g}(\cdot)$. By (2.4), we introduce the following auxiliary random vectors:

$$\eta_i^*(\beta^{(r)}) = \{\Delta_i/\overline{G}(V_i)\}\{V_i - \boldsymbol{g}^T(\beta^T X_i)Z_i\}\dot{\boldsymbol{g}}^T(\beta^T X_i)Z_i J^T_{\beta^{(r)}}X_i, \quad i = 1, \ldots, n, \tag{2.5}$$

Note that $E\{\eta_i^*(\beta_0^{(r)})\} = 0$. Hence, we can define an empirical log-likelihood ratio of $\beta_0^{(r)}$, say $l^*(\beta_0^{(r)})$, which is asymptotically chi-squared (Owen, 1990). However, $l^*(\beta_0^{(r)})$ cannot be directly used to make statistical inference on $\beta_0^{(r)}$ because it contains the unknown functions $G(\cdot)$, $\boldsymbol{g}(\cdot)$ and $\dot{\boldsymbol{g}}(\cdot)$. A natural way is to replace $G(\cdot)$, $\boldsymbol{g}(\cdot)$ and $\dot{\boldsymbol{g}}(\cdot)$ in $l^*(\beta_0^{(r)})$ by their estimators, respectively. We use the Kaplan-Meier estimator $G_n(v)$ of $G(v)$. That is,

$$G_n(v) = 1 - \prod_{i=1}^{n} \left(\frac{n-i}{n-i-1}\right)^{I\{V_{(i)} \leq v, \Delta_{(i)} = 0\}}, \tag{2.6}$$

where $V_{(1)} \leq \ldots \leq V_{(n)}$ are the order statistics of the V-sample, and $\Delta_{(i)}$ is the Δ associated with $V_{(i)}$, $i = 1, \ldots, n$. Let $V_{iG} = V_i\Delta_i/\{1 - G(V_i)\}$. We can get $E(V_{iG}|X_i, Z_i) = E(Y_i|X_i, Z_i)$,

$i = 1, \ldots, n$. Hence, under model (1.1), we have

$$V_{iG} = \boldsymbol{g}^T(\beta_0^T X_i) Z_i + e_i, \quad i = 1, \ldots, n,$$

where $e_i = V_{iG} - E(V_{iG}|X_i, Z_i), \quad i = 1, \ldots, n$. For the above mode and the fixed point β_0, we apply the local linear fitting technique to estimate the coefficient function $\boldsymbol{g}(u) = (g_1(u), \ldots, g_q(u))^T$ and its derivative function $\dot{\boldsymbol{g}}(u) = (g_1'(u), \ldots, g_q'(u))^T$. For any U in a small neighborhood of u, one can approximate $g_j(U)$ locally by the linear functions

$$g_j(U) \approx g_j(u) + g_j'(u)(U - u) \equiv a_j + b_j(U - u), \quad j = 1, \ldots, q.$$

Let $\{(\hat{a}_j, \hat{b}_j), j = 1, \ldots, q\}$ be the solution to the weighted least-squares problem

$$\sum_{i=1}^n \left[V_{iG_n} - \sum_{j=1}^q \left\{ a_j + b_j(\beta_0^T X_i - u) \right\} Z_{ij} \right]^2 K_h(\beta_0^T X_i - u),$$

where $G_n(\cdot)$ is defined in (2.6), $K_h(\cdot) = h^{-1}K(\cdot/h)$, $K(\cdot)$ is a kernel function, $h = h_n$ is a bandwidth and $h > 0$. Then the local linear estimators for $g_j(u)$ and $g_j'(u)$ are defined as $\hat{g}_j(u; \beta_0) = \hat{a}_j$ and $\hat{g}_j'(u; \beta_0) = \hat{b}_j$ at a given β_0, It follows from the least squares theory that

$$\left(\hat{\boldsymbol{g}}^T(u; \beta_0), h\hat{\dot{\boldsymbol{g}}}^T(u; \beta_0) \right)^T = \{D^T(u; \beta_0)\Omega(u; \beta_0)D(u; \beta_0)\}^{-1} D^T(u; \beta_0)\Omega(u; \beta_0)\boldsymbol{V}_{G_n}, \tag{2.7}$$

where $\hat{\boldsymbol{g}}(u; \beta_0) = (\hat{g}_1(u; \beta_0), \ldots, \hat{g}_q(u; \beta_0))^T$, $\hat{\dot{\boldsymbol{g}}}(u; \beta_0) = (\hat{g}_1'(u; \beta_0), \ldots, \hat{g}_q'(u; \beta_0))^T$,

$$D(u; \beta_0) = \begin{pmatrix} Z_1^T & h^{-1}(\beta_0^T X_1 - u) Z_1^T \\ \vdots & \vdots \\ Z_n^T & h^{-1}(\beta_0^T X_n - u) Z_n^T \end{pmatrix}, \tag{2.8}$$

$\Omega(u; \beta_0) = \mathrm{diag}(K_h(\beta_0^T X_1 - u), \ldots, K_h(\beta_0^T X_n - u))$ and $\boldsymbol{V}_{G_n} = (V_{1G_n} \ldots, V_{nG_n})^T$. From (2.7) we obtain the initial estimators of $\boldsymbol{g}(u)$ and $\dot{\boldsymbol{g}}(u)$, namely,

$$\hat{\boldsymbol{g}}(u; \beta_0) = \sum_{i=1}^n W_{ni}(u; \beta_0) V_{iG_n} \quad \text{and} \quad \hat{\dot{\boldsymbol{g}}}(u; \beta_0) = \sum_{i=1}^n \widetilde{W}_{ni}(u; \beta_0) V_{iG_n}, \tag{2.9}$$

where

$$(W_{n1}(u; \beta_0), \ldots, W_{nn}(u; \beta_0)) = (\boldsymbol{I}_q, \boldsymbol{0}_q)\{D^T(u; \beta_0)\Omega(u; \beta_0)D(u; \beta_0)\}^{-1} D^T(u; \beta_0)\Omega(u; \beta_0),$$

$$(\widetilde{W}_{n1}(u; \beta_0), \ldots, \widetilde{W}_{nn}(u; \beta_0)) = h^{-1}(\boldsymbol{0}_q, \boldsymbol{I}_q)\{D^T(u; \beta_0)\Omega(u; \beta_0)D(u; \beta_0)\}^{-1} D^T(u; \beta_0)\Omega(u; \beta_0),$$

$\boldsymbol{I}_q$ is $p \times q$ identity matrix and $\mathbf{0}_q$ is $q \times q$ zero matrix. Therefore, we can obtain a random vector $\tilde{\eta}_i^*(\beta_0)$ by substituting $G(\cdot)$, $\boldsymbol{g}(\cdot)$ and $\dot{\boldsymbol{g}}(\cdot)$ of $\eta_i^*(\beta_0)$ with $G_n(\cdot)$, $\hat{\boldsymbol{g}}(u;\beta_0)$ and $\hat{\dot{\boldsymbol{g}}}(u;\beta_0)$, which leads to an empirical log-likelihood ratio of β_0, say $\tilde{l}^*(\beta_0)$. It is can proved that the ratio has a asymptotic distribution of a weighted sum of independent chi-square distributions. each with one degree of freedom and an unknown weight. It cannot be directly used for the statistical inference of β_0. Thus, $\tilde{l}^*(\beta_0)$ needs to be adjusted by multiplying an adjustment factor.

Below we adopt an alternative approach to construct the EL ratio function of $\beta^{(r)}$. We can introduce the random vectors

$$\tilde{\eta}_i(\beta^{(r)}) = \frac{\Delta_i}{\overline{G}_n(V_i-)}\{V_i - \hat{\boldsymbol{g}}^T(\beta^T X_i;\beta)Z_i\}\hat{\dot{\boldsymbol{g}}}^T(\beta^T X_i;\beta)Z_i J^T_{\beta^{(r)}} X_i, \quad i = 1, \cdots, n,$$

and construct an empirical log-likelihood ratio of $\beta_0^{(r)}$, say $\tilde{l}(\beta_0^{(r)})$. However, the asymptotic distribution of $\tilde{l}(\beta_0^{(r)})$ is not standard chi-squared. Actually, $\tilde{l}(\beta_0^{(r)})$ is asymptotically a weighted sum of independent χ^2-variables. It cannot be directly used for the statistical inference of $\beta_0^{(r)}$.

Now, we directly construct a bias-corrected empirical log-likelihood ratio statistic of $\beta_0^{(r)}$ such that the statistic converges in distribution to a standard χ^2-variable with $p-1$ degrees of freedom. Since the estimators $G_n(\cdot)$ and $\hat{\boldsymbol{g}}(\cdot;\beta)$ are used, there exist the biases $G_n(\cdot) - G(\cdot)$ and $\hat{\boldsymbol{g}}(\cdot;\beta) - \boldsymbol{g}(\cdot)$ in $\tilde{\eta}_i(\beta)$. To reduce the biases, we use the bias correction method. Let

$$\begin{aligned}\varphi_n(W_i, V_i; \beta^{(r)}) = {} & w(\beta^T X_i)\{V_i - \hat{\boldsymbol{g}}^T(\beta^T X_i;\beta)Z_i\}\hat{\dot{\boldsymbol{g}}}^T(\beta^T X_i;\beta)Z_i \\ & \times J^T_{\beta^{(r)}}\{X_i - \hat{\mu}^T(\beta^T X_i, Z_i;\beta)\}, \end{aligned} \tag{2.10}$$

where $W_i = (X_i^T, Z_i^T)^T$, $i = 1, \ldots, n$, $J_{\beta^{(r)}}$, $\hat{\boldsymbol{g}}(\cdot;\beta)$ and $\hat{\dot{\boldsymbol{g}}}(\cdot;\beta)$ are defined in (2.3) and (2.9), respectively, $w(\cdot)$ is an indicator function on the bounded support of the distribution of $\beta_0^T X$, which is used to control the boundary effect in the estimations $\hat{\boldsymbol{g}}(\cdot;\beta)$, $\hat{\dot{\boldsymbol{g}}}(\cdot;\beta)$ and $\hat{\mu}(\cdot, z;\beta_0)$, and $\hat{\mu}(\cdot, z;\beta_0)$ is a Nadaraya-Watson estimation of $\mu(u, z) = E(X|\beta_0^T X = u, Z = z)$. That is,

$$\hat{\mu}(u, z;\beta_0) = \sum_{i=1}^n \mathcal{W}_{ni}(u, z) X_i, \tag{2.11}$$

where

$$\mathcal{W}_{ni}(u, z;\beta_0) = \frac{\mathcal{K}\left(\frac{\beta_0^T X_i - u}{b^{1/(q+1)}}, \frac{Z_i - z}{b^{1/(q+1)}}\right)}{\sum_{i=1}^n \mathcal{K}\left(\frac{\beta_0^T X_i - u}{b^{1/(q+1)}}, \frac{Z_i - z}{b^{1/(q+1)}}\right)},$$

$\mathcal{K}(\cdot,\cdot)$ is a kernel function on R^{q+1}, and $b = b_n$ is a bandwidth with $0 < b < 1$ and $b \to 0$. Let

$$\tilde{\varphi}_n(s;\beta^{(r)}) = \int_{v\geq s}\int_{w\in R^{p+q+1}} \varphi_n(w,y;\beta^{(r)})F_n(\mathrm{d}w,\mathrm{d}v), \tag{2.12}$$

where

$$F_n(w,v) = \int_{t\leq v}\int_{s\leq w} \frac{1}{\overline{G}_n(t-)}F_n^*(\mathrm{d}s,\mathrm{d}t). \tag{2.13}$$

and

$$F_n^*(w,v) = \frac{1}{n}\sum_{i=1}^n I(W_i \leq w, V_i \leq v, \Delta_i = 1). \tag{2.14}$$

Here for any two vectors $\boldsymbol{a} = (a_1,\dots,a_d)^T$ and $\boldsymbol{b} = (b_1,\dots,b_d)^T$, the inequality $\boldsymbol{a} \leq \boldsymbol{b}$, means that $a_i \leq b_i$, $i = 1,\dots,d$. We introduce the auxiliary random vectors

$$\begin{aligned}\hat{\eta}_i(\beta^{(r)}) = &\frac{\Delta_i}{\overline{G}_n(V_i-)}\varphi_n(W_i,V_i;\beta^{(r)}) + \frac{1-\Delta_i}{\overline{H}_n(V_i-)}\tilde{\varphi}_n(V_i;\beta^{(r)}) \\ &-\int_{-\infty}^{\infty}\frac{I(V_i\geq s)}{\overline{H}_n^2(s-)}\tilde{\varphi}_n(s;\beta^{(r)})\mathrm{d}H_{0n}(s), \quad i=1,\dots,n,\end{aligned} \tag{2.15}$$

where $\overline{H}_n(\cdot) = 1 - H_n(\cdot)$,

$$\begin{aligned}H_n(v) &= \frac{1}{n}\sum_{i=1}^n I(V_i \leq v), \\ H_{0n}(v) &= \frac{1}{n}\sum_{i=1}^n I(V_i \leq v, \Delta_i = 0),\end{aligned}$$

and $G_n(\cdot)$, $\varphi_n(\cdot,\cdot;\beta^{(r)})$ and $\tilde{\varphi}_n(\cdot;\beta^{(r)})$ are defined in (2.6), (2.10) and (2.12), respectively. Thus, a bias-corrected empirical log-likelihood ratio function for $\beta^{(r)}$ is defined as

$$\hat{l}(\beta^{(r)}) = -2\max_{p_1,\dots,p_n}\left\{\sum_{i=1}^n \log(np_i)\Big| p_i\geq 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i\hat{\eta}_i(\beta^{(r)}) = 0\right\}. \tag{2.16}$$

Remark 1. The latter two items on the right side of (2.15) play a very important role in the bias correction, which is mainly used to reduce the bias caused by $G_n(\cdot) - G(\cdot)$. The centralization of X in $\varphi_n(W_i,V_i;\beta^{(r)})$ also plays a important role in the bias correction, which is mainly used to reduce the bias caused by $\hat{\boldsymbol{g}}(\cdot;\beta) - \boldsymbol{g}(\cdot)$. See the proof of Lemma 4 in the supporting materials. Just because we use the bias correction method to construct the EL ratio $\hat{l}(\beta^{(r)})$, the asymptotic distribution of $\hat{l}(\beta_0^{(r)})$ is a standard chi-square distribution with p degrees of freedom. The result is given in Theorem 2 of Section 3.

2.2 The estimation of regression parameter

We can maximize $\{-\hat{l}(\beta^{(r)})\}$ to obtain an estimator of $\beta_0^{(r)}$, say $\hat{\beta}^{(r)}$. According to Qin and Lawless (1994), $\hat{\beta}^{(r)}$ is asymptotically equivalent to the solution of the estimating equation $\sum_{i=1}^{n} \hat{\eta}_i(\beta^{(r)}) = 0$. From (2.15), we have

$$\begin{aligned}
\widehat{Q}(\beta^{(r)}) &\equiv \frac{1}{n}\sum_{i=1}^{n} \hat{\eta}_i(\beta^{(r)}) \\
&= \frac{1}{n}\sum_{i=1}^{n} \frac{\Delta_i}{\overline{G}_n(V_i-)}\varphi_n(W_i, V_i; \beta^{(r)}) + \int_{-\infty}^{\infty} \frac{1}{\overline{H}_n(s-)}\tilde{\varphi}_n(s; \beta^{(r)})\mathrm{d}H_{0n}(s) \\
&\quad - \int_{-\infty}^{\infty} \frac{\overline{H}_n(s-)}{\overline{H}_n^2(s-)}\tilde{\varphi}_n(s; \beta^{(r)})\mathrm{d}H_{0n}(s) \\
&= \frac{1}{n}\sum_{i=1}^{n} \frac{\Delta_i}{\overline{G}_n(V_i-)}\varphi_n(W_i, V_i; \beta^{(r)}) \equiv \widehat{Q}^*(\beta^{(r)}). \qquad (2.17)
\end{aligned}$$

From $\|\beta_0\| = 1$, we can obtain an estimator of β_0, say $\hat{\beta}$. Therefore, $\hat{\beta}$ is asymptotically equivalent to the solution of the estimating equations

$$\begin{cases} \widehat{Q}^*(\beta^{(r)}) = 0, \\ \|\beta\| - 1 = 0. \end{cases} \qquad (2.18)$$

An iterative algorithm is widely used for solving the estimating equations (2.18). From (A.25) in the proof of Theorem 3, we can give an iteration formula for calculating the estimate of β_0. That is,

$$\breve{\beta} = \tilde{\beta} + J_{\tilde{\beta}^{(r)}}\sqrt{n}\tilde{A}^{-}(\tilde{\beta}^{(r)})\hat{Q}^*(\tilde{\beta}^{(r)}), \qquad (2.19)$$

where $\widehat{Q}^*(\cdot)$ is defined by (2.17),

$$\tilde{A}(\tilde{\beta}^{(r)}) = \frac{1}{n}\sum_{i=1}^{n} \frac{\Delta_i w(\tilde{\beta}^T X_i)}{\overline{G}_n(V_i-)}\{\hat{\dot{\boldsymbol{g}}}^T(\tilde{\beta}^T X_i; \tilde{\beta})Z_i\}^2 J^T_{\tilde{\beta}^{(r)}}\{X_i - \hat{\mu}(\tilde{\beta}^T X_i, Z_i; \tilde{\beta})\}^{\otimes 2} J_{\tilde{\beta}^{(r)}}, \qquad (2.20)$$

$G_n(\cdot)$, $\hat{\dot{\boldsymbol{g}}}(\cdot; \tilde{\beta})$ and $\hat{\mu}(\cdot, \cdot; \tilde{\beta})$ are defined in (2.6), (2.9) and (2.11), respectively, $\tilde{A}^{-}$ is a generalized inverse of $\hat{A}$ and $\xi^{\otimes 2} = \xi\xi^T$ for any vector ξ. This iterative algorithm solves (2.18) and is identical to the Fisher's method of scoring version of the Newton-Raphson algorithm for solving the estimating equation.

We now outline the algorithm for estimation β_0 and $\boldsymbol{g}(\cdot)$.

Step 0 (Initialization step): Specify an initial value $\hat{\beta}_0$ with $\|\hat{\beta}_0\| = 1$.

Step 1: For a given β, compute estimate $\boldsymbol{g}(\cdot)$ and $\dot{\boldsymbol{g}}(\cdot)$ by formula (2.9).

Step 2: Estimate β using the formula (2.19), and get an estimate $\hat{\beta}$.

Step 3: Using the estimate $\hat{\beta}$ to obtain the final estimate of $\boldsymbol{g}(\cdot)$, say $\hat{\boldsymbol{g}}^*(u) = \hat{\boldsymbol{g}}(u, \hat{\beta})$.

For narrative purposes, we call the above method the reparametric EL method (RELM).

Remark 2. The basic idea behinds the foregoing algorithm is simple: estimate $\boldsymbol{g}(\cdot)$ and $\dot{\boldsymbol{g}}(\cdot)$ locally via (2.9), and then use all of the data and (2.19) to estimate β_0, with $\hat{\boldsymbol{g}}(\cdot;\beta)$ and $\hat{\dot{\boldsymbol{g}}}(\cdot;\beta)$ replacing $\boldsymbol{g}(\cdot)$ and $\dot{\boldsymbol{g}}(\cdot)$. The estimation procedure involves choosing the bandwidth. In step 1, we first select the bandwidth using the initial value $\hat{\beta}_0$, then use the estimate of β_0 in the each iteration to select the bandwidth, and use the bandwidth to estimate $\boldsymbol{g}(\cdot)$ and $\dot{\boldsymbol{g}}(\cdot)$ for a given β using formula (2.9).

Remark 3. It is worth mentioning that the reparametrization approach plays a very important role in constructing the estimator of β_0. Since we use this approach for β_0, the resulting estimator $\hat{\beta}$ is more efficient than other estimators. To compare the advantages of the above estimation methods, we present an alternative method to estimate β_0.

We may solve the estimating equation

$$\sum_{i=1}^{n} \tilde{\eta}_i(\beta) = 0 \tag{2.21}$$

to get an estimator of β_0, say $\hat{\beta}^*$, where

$$\tilde{\eta}_i(\beta) = \frac{1}{n}\sum_{i=1}^{n} \frac{\Delta_i w(\tilde{\beta}^T X_i)}{\overline{G}_n(V_i-)}\{V_i - \hat{\boldsymbol{g}}^T(\beta^T X_i;\beta) Z_i\}\hat{\dot{\boldsymbol{g}}}^T(\beta^T X_i;\beta) Z_i\{X_i - \hat{\mu}(\beta^T X_i, Z_i;\beta),$$

$G_n(\cdot)$, $\hat{\boldsymbol{g}}(\cdot;\beta)$, $\hat{\dot{\boldsymbol{g}}}(\cdot;\beta)$, $w(\cdot)$ and $\hat{\mu}(\cdot,\cdot;\beta)$ are defined by (2.6), (2.9), (2.10) and (2.11), respectively. In Theorem 4 of Section 3, we will show the asymptotic normality of $\hat{\beta}^*$.

For narrative purposes, we call the above method the reparametrize estimating equation method (NREEM). We will compare the execution effects of RELM and NREEM in the simulation study.

3 Main results

To obtain the asymptotic behaviors of the estimators, we first give the following conditions.

(C1) The density function of $\beta^T X$, $f_\beta(u)$, is bounded away from zero for $u \in \mathcal{U}_w$ and β near β_0, and satisfies the Lipschitz condition of order 1 on $\mathcal{U}_w$, where $\mathcal{U}_w$ is a compact support of $w(u)$. The distribution of Z has a compact support $\mathcal{Z}$.

(C2) The functions $g_j(u)$, $1 \le j \le q$, have bounded and continuous second derivatives on $\mathcal{U}_w$, where $g_j(u)$ is the jth component of $\boldsymbol{g}(u)$.

(C3) For β near β_0, the matrix $\Phi(u;\beta) = E(ZZ^T|\beta^T X = u)$ is positive definite, the (j,k)-th element of $\Phi(u;\beta)$, $1 \le j, k \le q$, is bounded away from zero, and satisfy the Lipschitz condition of order 1 on $\mathcal{U}_w$. Also, the (j,k)-th element $\omega_{jk}(u;\beta)$ of the matrix $\Omega(u;\beta)$, $1 \le j, k \le p$, is continuous at $u_0 \in \mathcal{U}_w$, where $\Omega(u;\beta)$ and $\mathcal{U}_w$ are defined in Theorem 6 and condition (C1), respectively.

(C4) The joint density function of $(\beta^T X, Z)$, $f_\beta(u,z)$, is bounded away from zero on $\mathcal{U}_w \times \mathcal{Z}$, where $\mathcal{U}_w$ and $\mathcal{Z}$ are defined in condition (C1). The functions $f_\beta(u,z)$ and $\mu_j(u,z)$ have bounded partial derivatives up to order $2(q+1)$, where $\mu_j(u,z)$ is the jth components of $\mu(u,z) = E(X|\beta^T X = u, Z = z)$ for β near β_0, and $j = 1, \ldots, p$.

(C5) The kernel $K(u)$ is a bounded symmetric probability density function with a support $[-1,1]$, satisfying the Lipschitz condition, and $\int_{-1}^{1} u^2 K(u) \mathrm{d}u \neq 0$. The kernel $\mathcal{K}(u,z)$ is a real-valued function of bounded variation, and is a right continuous probability density function of order $q+1$ with support contained $[-1,1]^{q+1}$.

(C6) The bandwidths h and b satisfy $h = c_1 n^{-1/5}$ and $b = c_2 n^{-1/5}$ for $c_1 > 0$ and $c_2 > 0$.

(C7) $\sup_{u,z} E[\{\|X\|^2/\overline{G}(Y)\}|\,\beta^T X = u, Z = z] < \infty$, $\sup_{x,z} E[\{\varepsilon^4/\overline{G}(Y)\}|\,X = x, Z = z] < \infty$, where β near β_0.

Remark 4. Condition (C1) is used to bound the density function of $\beta^T X$ away from zero. This ensures that the denominators of $\hat{\boldsymbol{g}}(u;\beta)$ and $\hat{\dot{\boldsymbol{g}}}(u;\beta)$ are, with high probability, bounded away from 0 for $u \in \mathcal{U}_w$. Conditions (C2)–(C4) are the standard smoothness and continuous conditions. Condition (C5) is the common assumption for the kernel functions. Condition (C6) gives the optimal bandwidth for estimating $\boldsymbol{g}(u)$ and β_0. Condition (C7) is the usual component condition.

We first state the uniform convergence rates of the estimators of $\boldsymbol{g}(\cdot)$ and $\dot{\boldsymbol{g}}(\cdot)$.

Theorem 1. Suppose that conditions (C1)–(C3) and (C5)–(C7) hold. Then

$$\sup_{u\in\mathcal{U}_w,\beta\in\mathcal{B}_n} \|\hat{\boldsymbol{g}}(u;\beta)-\boldsymbol{g}(u)\| = O_P\left(n^{-2/5}\sqrt{\log n}\right)$$

and

$$\sup_{u\in\mathcal{U}_w,\beta\in\mathcal{B}_n} \|\hat{\dot{\boldsymbol{g}}}(u;\beta)-\dot{\boldsymbol{g}}(u)\| = O_P\left(n^{-1/5}\sqrt{\log n}\right),$$

where $\mathcal{B}_n = \{\beta| \, \|\beta-\beta_0\| \le c_1 n^{-1/2}\}$ for some positive constant c_1, $\hat{\boldsymbol{g}}(u;\beta)$ and $\hat{\dot{\boldsymbol{g}}}(u;\beta)$ are defined in (2.9), and $\dot{\boldsymbol{g}}(u) = (g_1'(u),\ldots,g_q'(u))^T$.

The asymptotic property of $\hat{l}(\beta_0)$ defined by (2.16) is as follows.

Theorem 2. Suppose that conditions (C1)–(C7) hold. Then $\hat{l}(\beta_0^{(r)}) \xrightarrow{D} \chi_q^2$, where " $\xrightarrow{D}$ " represents the convergence in distribution, χ_q^2 is a chi-square variable with q degrees of freedom.

Using Theorem 2, an approximate $1-\alpha$ confidence region of $\beta_0^{(r)}$ is defined as

$$\left\{\beta^{(r)}|\, \hat{l}(\beta^{(r)}) \le \chi_q^2(1-\alpha)\right\},$$

where $\chi_q^2(1-\alpha)$ is the $(1-\alpha)$th quantile of the χ_q^2 distribution and $0<\alpha<1$.

The following theorem shows that the estimator $\hat{\beta}$ has asymptotic normality.

Theorem 3. Suppose that conditions (C1)–(C7) hold. Then

$$\sqrt{n}\left(\hat{\beta}^{(r)}-\beta_0^{(r)}\right) \xrightarrow{D} N\left(0, A^- B A^-\right),$$

and

$$\sqrt{n}\left(\hat{\beta}-\beta_0\right) \xrightarrow{D} N\left(0, J_{\beta_0^{(r)}} A^- B A^- J_{\beta_0^{(r)}}^T\right),$$

where $J_{\beta_0^{(r)}}$ is defined by (2.3),

$$\begin{aligned}
A &= E\left\{w(\beta_0^T X)\{\dot{\boldsymbol{g}}^T(\beta_0^T X)Z\}^2 J_{\beta_0^{(r)}}^T \{X-\mu(\beta_0^T X,Z)\}^{\otimes 2} J_{\beta_0^{(r)}}\right\},\\
B &= E\left[\frac{1}{\overline{G}(Y)}\{\varphi(W,Y;\beta_0^{(r)})\}^{\otimes 2}\right] - E\left[\frac{1}{\overline{F}(C)\overline{G}^2(C)}\{\tilde{\varphi}(C;\beta_0^{(r)})\}^{\otimes 2}\right],\\
\varphi(W,Y;\beta_0^{(r)}) &= w(\beta_0^T X)\{Y-\boldsymbol{g}^T(\beta_0^T X)Z\}\dot{\boldsymbol{g}}^T(\beta_0^T X)Z J_{\beta^{(r)}}^T\{X-\mu(\beta_0^T X,Z)\}, \qquad (3.1)\\
\tilde{\varphi}(y;\beta_0^{(r)}) &= E\left\{\varphi(W,Y;\beta_0^{(r)})I(Y\ge y)\right\}, \qquad (3.2)
\end{aligned}$$

and $\xi\xi^{\otimes 2}$ for any vector ξ.

To apply Theorem 3 to construct the confidence regions of $\beta_0^{(r)}$ and β_0, we need to estimate A and B. The estimator of A is defined as $\hat{A} = \tilde{A}(\hat{\beta}^{(r)})$, where $\tilde{A}(\cdot)$ is defined by (2.20). The estimator of B is defined as

$$\widehat{B} = \frac{1}{n} \sum_{i=1}^{n} \hat{\eta}(\hat{\beta}^{(r)}) \hat{\eta}^T(\hat{\beta}^{(r)}),$$

where $\hat{\eta}(\cdot)$ is defined in (2.15). By Theorem 3, we have

$$\left\{\hat{A}^- \hat{B} \hat{A}^-\right\}^{-1/2} \sqrt{n}(\hat{\beta}^{(r)} - \beta_0^{(r)}) \xrightarrow{D} N(0, I_{p-1})$$

and

$$\left\{J_{\hat{\beta}^{(r)}} \hat{A}^- \hat{B} \hat{A}^- J^T_{\hat{\beta}^{(r)}}\right\}^{-1/2} \sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{D} N(0, I_p).$$

Using Theorem 10.2d in Arnold (1981), we obtain

$$(\hat{\beta}^{(r)} - \beta_0^{(r)})^T \left\{n^{-1} \hat{A}^- \hat{B} \hat{A}^-\right\}^{-1} (\hat{\beta}^{(r)} - \beta_0^{(r)}) \xrightarrow{D} \chi^2_{p-1}$$

and

$$(\hat{\beta} - \beta_0)^T \left\{n^{-1} J_{\hat{\beta}^{(r)}} \hat{A}^- \hat{B} \hat{A}^- J^T_{\hat{\beta}^{(r)}}\right\}^{-1} (\hat{\beta} - \beta_0) \xrightarrow{D} \chi^2_p.$$

The above results can be used to construct the confidence regions/intervals of $\beta_0^{(r)}$ and β_0, and to make the hypothesis test.

The following theorem shows that the estimator $\hat{\beta}^*$ has asymptotic normality.

Theorem 4. Suppose that conditions (C1)–(C7) hold. Then

$$\sqrt{n}\left(\hat{\beta}^* - \beta_0\right) \xrightarrow{D} N\left(0, C^- D C^-\right),$$

where

$$\begin{aligned}
C &= E\left\{w(\beta_0^T X)\{\dot{\boldsymbol{g}}^T(\beta_0^T X)Z\}^2\{X - \mu(\beta_0^T X, Z)\}^{\otimes 2}\right\}, \\
D &= E\left[\frac{1}{\overline{G}(Y)}\{\varphi^*(W, Y; \beta_0^{(r)})\}^{\otimes 2}\right] - E\left[\frac{1}{\overline{F}(C)\overline{G}^2(C)}\{\tilde{\varphi}^*(C; \beta_0^{(r)})\}^{\otimes 2}\right], \\
\varphi^*(W, Y; \beta_0^{(r)}) &= w(\beta_0^T X)\{Y - \boldsymbol{g}^T(\beta_0^T X)Z\}\dot{\boldsymbol{g}}^T(\beta_0^T X)Z\{X - \mu(\beta_0^T X, Z)\}
\end{aligned}$$

and $\tilde{\varphi}^*(y; \beta_0^{(r)}) = E\left\{\varphi^*(W, Y; \beta_0^{(r)}) T(Y \geq y)\right\}$.

Because $\hat{\beta}$ is root-n consistent, we immediately obtain the following result.

Theorem 5. Suppose that conditions (C1)–(C7) hold. Then

$$\sqrt{nh}\left\{\hat{\boldsymbol{g}}^*(u_0) - \boldsymbol{g}(u_0) - \boldsymbol{b}(u_0)\right\} \xrightarrow{D} N\left(0, \Sigma(u_0; \beta_0)\right),$$

where $\boldsymbol{b}(u_0) = \frac{1}{2}h^2\ddot{\boldsymbol{g}}(u_0)\int_{-1}^{1} t^2K(t)dt$, $\ddot{\boldsymbol{g}}(u_0) = (g_1''(u_0), \ldots, g_q''(u_0))^T$,

$$\begin{aligned}
\Sigma(u_0; \beta_0) &= \Psi^{-1}(u_0)\Omega(u_0)\Psi^{-1}(u_0), \\
\Psi(u_0; \beta_0) &= f_{\beta_0}(u_0)E(ZZ^T|\beta_0^TX = u_0)
\end{aligned}$$

and

$$\Omega(u_0; \beta_0) = f_{\beta_0}(u_0)E\left[\{V_{1G} - Y + \varepsilon\}^2ZZ^T|\,\beta_0^TX = u_0\right]\int_{-1}^{1} K^2(u)du.$$

In Theorem 5, if the condition (C6) is replaced by $nh^2/\log n \to \infty$ and $nh^5 \to 0$, then

$$\sqrt{nh}\{\hat{\boldsymbol{g}}^*(u_0) - \boldsymbol{g}(u_0)\} \xrightarrow{D} N(0, \Sigma(u_0)). \tag{3.3}$$

Applying Theorem 5, we can construct the pointwise confidence intervals for a component of $\boldsymbol{g}(u_0)$. However, we need to use the plugin estimations for the asymptotic bias and covariance of $\hat{\boldsymbol{g}}(u_0)$. Obviously, the asymptotic bias and covariance of $\hat{\boldsymbol{g}}(u_0)$ are dependent on $\boldsymbol{b}(u_0)$, $\Psi(u_0; \beta_0)$ and $\Omega(u_0; \beta_0)$, we need to estimate them. The estimators $\Psi(u_0; \beta_0)$ and $\Omega(u_0; \beta_0)$ are defined as

$$\widehat{\Psi}(u_0) = \frac{1}{n}\sum_{i=1}^{n} Z_iZ_i^TK_h(\beta^TX_i - u_0)$$

and

$$\widehat{\Omega}(u_0) = \frac{1}{nh}\sum_{i=1}^{n} \hat{\zeta}_i(u_0)\hat{\zeta}_i^T(u_0),$$

respectively, where $\hat{\zeta}_i(u_0) = \{V_{iG_n} - \hat{\boldsymbol{g}}^{*T}(\hat{\beta}^TX_i)Z_i\}Z_iK((\hat{\beta}^TX_i - u_0)/h)$, $\hat{\boldsymbol{g}}^*(\cdot) = \hat{\boldsymbol{g}}(\cdot; \hat{\beta})$, $\hat{\boldsymbol{g}}(\cdot; \hat{\beta})$ is given by (2.9), $\hat{\beta}$ is given by the estimating equations (2.18), and $K(\cdot)$ and h are defined in (2.7). We now consider the estimator of $\boldsymbol{b}(u_0)$. Note that

$$E\left[\{\boldsymbol{g}(\beta_0^TX) - \boldsymbol{g}(u_0)\}K_h(\beta_0^TX - u_0)\right] = \boldsymbol{b}(u_0) + o_P(h^2).$$

Hence, an estimator of $\boldsymbol{b}(u_0)$ is defined as

$$\hat{\boldsymbol{b}}(u_0) = \frac{1}{n}\sum_{i=1}^{n}\{\hat{\boldsymbol{g}}^*(\hat{\beta}^TX_i) - \hat{\boldsymbol{g}}^*(u_0)\}K_h(\beta_0^TX_i - u_0).$$

It is easy to prove that $\widehat{\Psi}(u_0)$, $\widehat{\Omega}(u_0)$ and $\hat{\boldsymbol{b}}(u_0)$ are the consistent estimators of $\Psi(u_0; \beta_0)$, $\Omega(u_0; \beta_0)$ and $\boldsymbol{b}(u_0)$, respectively. Assume that $\Psi(u_0; \beta_0)$ is invertible. Then $\Psi^{-1}(u_0; \beta_0)$ can be consistently estimated by $\widehat{\Psi}^{-1}(u_0)$. Finally, we can obtain a consistent estimator $\widehat{\Sigma}(u_0)$ of $\Sigma(u_0;\beta_0)$ by substituting $\Omega(u_0;\beta_0)$ and $\Psi^{-1}(u_0;\beta_0)$ of $\Sigma(u_0;\beta_0)$ with $\widehat{\Omega}(u_0)$ and $\widehat{\Psi}^{-1}(u_0)$. From Theorem 5, we have

$$\{\widehat{\Sigma}(u_0)\}^{-1/2}\sqrt{nh}\left\{\hat{\boldsymbol{g}}^*(u_0)-\boldsymbol{g}(u_0)\}-\hat{\boldsymbol{b}}(u_0)\right\} \xrightarrow{D} N(0,\boldsymbol{I}_q), \tag{3.4}$$

where $\boldsymbol{I}_q$ is the unit matrix of order q. Using (3.4), a pointwise confidence interval for each component $g_j(u_0)$ of $\boldsymbol{g}(u_0)$ can be given by

$$\hat{g}_j^*(u_0)-\hat{b}_j(u_0)\pm z_{1-\alpha/2}(nh)^{-1/2}\hat{\sigma}_j(u_0), \quad j=1,\ldots,q, \tag{3.5}$$

where $\hat{g}_j^*(u_0)$ and $\hat{b}_j(u_0)$ are the jth components of $\hat{\boldsymbol{g}}^*(u_0)$ and $\hat{\boldsymbol{b}}(u_0)$, $\hat{\sigma}_j(u_0)$ is the (j,j)th element of $\widehat{\Sigma}(u_0)$, and $z_{1-\alpha/2}$ is the $1-\alpha/2$ quantile value of the standard normal distribution.

If the condition (C6) in Theorem 5 is replaced by $nh^2/\log n\to\infty$ and $nh^5\to 0$, then by (3.3), we have

$$\hat{g}_j^*(u_0)\pm z_{1-\alpha/2}(nh)^{-1/2}\hat{\sigma}_j(u_0), \quad j=1,\ldots,q. \tag{3.6}$$

Formulas (3.5) and (3.6) can be also used to construct a pointwise confidence interval of $g_j(u_0)$.

4 Numerical properties

In this section, we first carry out a simulation study to demonstrate the performance of our method in finite data set. We then apply the single-index varying-coefficient model and the estimation method to a real data set.

4.1 Simulation study

Consider the single-index varying-coefficient model (2.1), where X_i are the multivariate standard normal distribution with $p=3$, $Z_i=(1,Z_{i1},,Z_{i2})^T$, $(Z_{i1},,Z_{i2})^T$ are independent random vectors uniformly distributed on $[0,1]^2$, $\varepsilon_i\sim N(0,0.4^2)$, the regression parameter is $\beta_0=(\beta_{01},\beta_{02},\beta_{03})^T=(1/3,2/3,2/3)^T$, and the coefficient functions are $g_0(t)=2\exp(-t)$, $g_1(t)=6t^3$ and $g_2(t)=5\cos(\pi t)$, respectively. The response Y is right censored. The censored variable C follows an exponential distribution with the parameter $\lambda = 0.07$. The average censoring rate is about 20%.

We used the Epanechnikov kernel function $K(x) = 0.75(1 - x^2)I(|x| \le 1)$ and the kernel function $\mathcal{K}(t, z_1, z_2) = K_0(t)K_0(z_1)K_0(z_2)$, where $K_0(x) = (3/8)(3 - 5x^2)I(|x| \le 1)$, and used the cross-validation (CV) and modified multi-fold cross-validation (MCV) methods (see, Cai et al. 2000b) to select the optimal bandwidths b and h, whose calculation formulas are similar, only use V_{iG_n} instead of the response variable Y_i in the formulas. Since the simulation results are not sensitive to the choice of this weight function $w(\cdot)$, we take $w(\cdot) = 1$, and the calculation is stable.

Two methods were used to make the simulation for the estimate of β_0. That is, the RELM and NREEM given in Section 2. The simulations were considered in the following three situations.

(I) The bias, standard deviation (SD) and mean squared error (MSE) for the estimators $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\beta}_3$ were computed by 500 runs with sample sizes 60, 100 and 150. The simulated results are presented in Table 1.

Table 1. The bias, SD and MSE for the estimators $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\beta}_3$.

		$\hat{\beta}_1$			$\hat{\beta}_2$			$\hat{\beta}_3$		
n	Method	Bias	SD	MSE	Bias	SD	MSE	Bias	SD	MSE
60	RELM	−0.1283	0.1326	0.0340	0.0265	0.1140	0.0137	0.0185	0.1160	0.0183
	NREEM	0.1263	0.2359	0.0716	−0.0996	0.2449	0.0699	−0.1099	0.2392	0.0693
100	RELM	−0.1241	0.0946	0.0244	0.0180	0.0811	0.0069	0.0177	0.0805	0.0068
	NREEM	0.0876	0.2467	0.0704	−0.0885	0.2442	0.0675	−0.1088	0.2385	0.0687
150	RELM	−0.1242	0.0944	0.0243	−0.0170	0.0810	0.0068	−0.0167	0.0802	0.0067
	NREEM	0.0729	0.2454	0.0655	−0.0837	0.2406	0.0649	−0.0989	0.2347	0.0649

From Table 1 we can see that the estimators $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\beta}_3$ based on RELM have smaller MSE than the estimators based on NREEM. This shows that RELM improved estimated accuracy. In addition, all the bias, SD and RMSE decrease as the sample size n increases.

(II) The performances of the estimators $\hat{g}_0^*(t)$, $\hat{g}_1^*(t)$ and $\hat{g}_2^*(t)$ were considered when sample size is 60. Figure 1 (a)–(c) give the real coefficient function curves, the estimated coefficient function curves and approximate 95% pointwise confidence intervals. The estimators $\hat{g}^*_\nu(\cdot)$ are assessed by using the root mean squared errors (RMSE). That is,

$$\mathrm{RMSE}_\nu = \left[n_{\mathrm{grid}}^{-1} \sum_{k=1}^{n_{\mathrm{grid}}} \{\hat{g}^*_\nu(t_k) - g(t_k)\}^2 \right]^{1/2}, \quad \nu = 0, 1, 2, \tag{4.1}$$

where $\{u_k, k = 1, \ldots, n_{\mathrm{grid}}\}$ are regular grid points. The boxplot for the 500 RMSEs is given in Figure 1 (d).

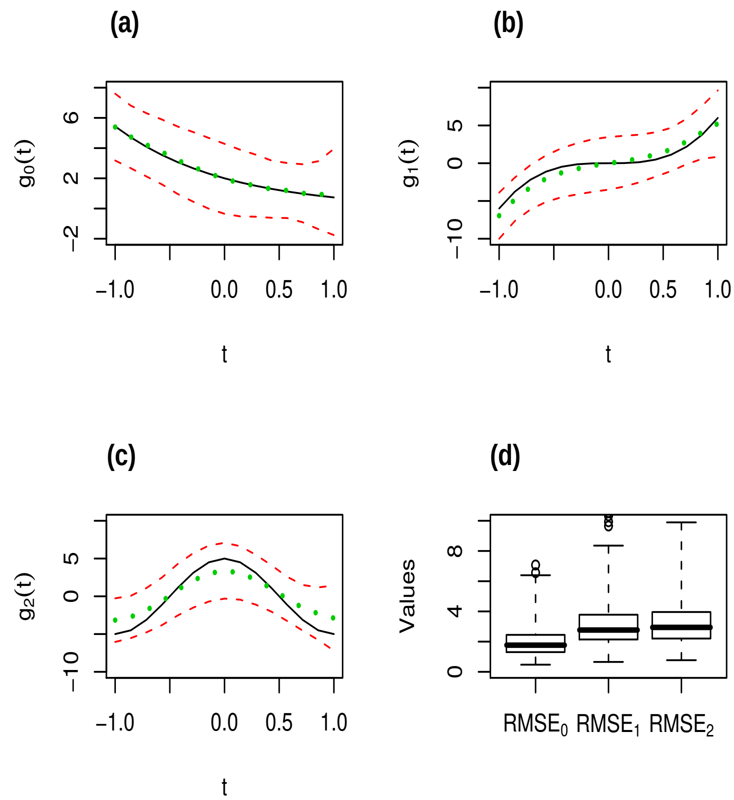

Figure 1: (a)–(c) are the real cures (solid curves), estimated curves (dotted curves) and approximate 95% pointwise confidence intervals (dashed curves) for $g_0(t)$, $g_1(t)$ and $g_2(t)$, respectively. (d) is the boxplots of the 500 RMSE values for estimate of $g_0(t)$, $g_1(t)$ and $g_2(t)$ when $n = 60$.

From Figure 1 (a)–(c) we see that each estimated curves are close to the real coefficient function curves. Figure 1 (d) shows that the RMSEs of estimates for coefficient functions are small.

(III) The confidence regions for (β_{01}, β_{02}) and (β_{01}, β_{03}), and their coverage probabilities were also computed from 200 simulation runs, which were based on EL and the normal approximation (NA) when the sample size was 100. The simulation results are presented in Figure 2.

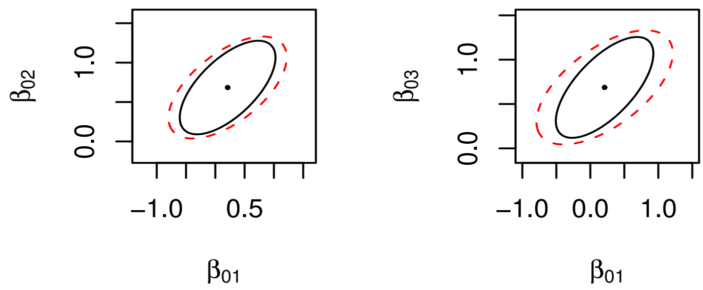

Figure 2: Approximate 95% confidence regions for (β_{01}, β_{02}) and (β_{01}, β_{03}), based on EL (solid curve) and NA (dashed curve) when $n = 100$.

Figure 2 shows that EL gives smaller confidence regions than NA. For (a), the empirical coverage probability for EL and NA are 0.935 and 0.930, respectively. For (b), the empirical coverage probability for EL and NA are 0.930 and 0.925, respectively.

4.2 A real example

We now illustrate the proposed method through the application of an environmental data set. The data set used here consists of daily measurements of pollutants and other environmental factors in New Territories East in Hong Kong between January 1, 2000 and June 30, 2000. For this data set, there are three pollutants and an environmental factors. That is, sulphur dioxide (in g/m^3)X_1, nitrogen dioxide (in g/m^3)X_2, ozone (in g/m^3)X_3, and temperature (in Celsius) Z. Wong, Ip and Zhang (2008) investigated this data set using a partially varying-coefficient single-index model. Our main interest is to study the relationship between the levels of chemical pollutants and the number of daily total hospital admissions (Y) for respiratory diseases in New Territories East in Hong Kong. We uae the single-index varying-coefficient model to fit the data set. That is,

$$Y = g_0(\beta_0^T X) + g_1(\beta_0^T X)Z + \varepsilon, \tag{4.2}$$

where $\beta_0^T X = \beta_{01}X_1 + \beta_{02}X_2 + \beta_{03}X_3$.

To use the data set to illustrate our method, we assume that the response variable Y is censored. The distribution of the censored variable C was taken as $G(t) = 1 - \exp(-0.01t)$, and hence about 33% of the Y values are censored. We used the Epanechnikov kernel function $K(x) = 0.75(1 - x^2)I(|x| \leq 1)$ and the kernel function $\mathcal{K}(t, z) = K_0(t)K_0(z)$, where

$K_0(x) = (3/8)(3-5x^2)I(|x| \le 1)$, We also used the modified multi-fold cross-validation (MCV) method proposed by Cai et al. (2000b) to select the optimal bandwidths. Since Δ was randomly generated, the estimates of β_0, $g_0(t)$ and $g_1(t)$ and the confidence regions of β_0 were computed from 50 simulation runs, which are based on EL and NA. By calculation, we obtain that the estimate of β_0 is $(0.3009, 0.7181, 0.5280)^T$, corresponding standard error is $\{0.1364, 0.0865, 0.0621\}$.

By calculation, we obtain that the bandwidth of the estimators $\hat{g}_0^*(t)$ and $\hat{g}_1^*(t)$ is 26.474. The estimated curves are shown in Figure 3.

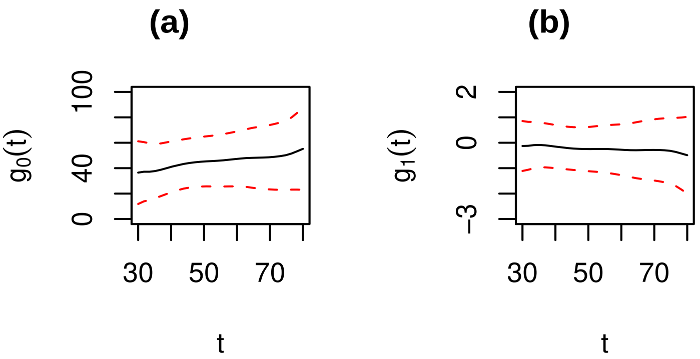

Figure 3: The estimated curves (solid curve)and approximate 95% pointwise confidence intervals (dashed curves) for the coefficient functions $g_0(t)$ and $g_1(t)$.

From Figure 3 (a) it is clear that the number of daily total hospital admissions for respiratory disease patients in New Territories East in Hong Kong is increasing with the air pollutants index—a linear combination $t = \hat{\beta}^T x$ of the chemical pollutants—increasing, which implies that increasing air pollution increases the onset of respiratory diseases. This suggests that the government should take air pollution seriously. However, if we consider the interaction between three contaminants and an environmental factor, then the number of daily total hospital admissions for respiratory disease patients is decreasing with the air pollutants index $t = \hat{\beta}^T x$ increasing. Moreover, we can also see from Figure 3 that the estimated curves of the coefficient functions are not linear, that is, the linear model is not suitable to fit this data set.

The confidence regions for (β_{01}, β_{02}) and (β_{01}, β_{03}) were also computed from 50 simulation runs, which were based on EL and NA. The simulation results are presented in Figure 4. Figure 4 shows that EL gives smaller confidence regions than NA.

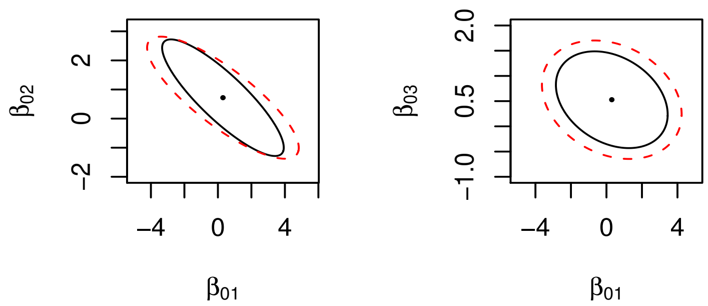

Figure 4: Approximate 95% confidence regions for (β_{01}, β_{02}) and (β_{01}, β_{03}), based on EL (solid curve) and NA (dashed curve).

The coefficient of determination, R^2_{new}, is used to evaluate the goodness of fit for nonlinear regression models. That is,

$$R^2_{\text{new}} = 1 - \sum_{i-1}^{n} \frac{\Delta_i}{\overline{G}_n(V_i-)}(V_i - \widehat{Y}_i)^2 \Big/ \sum_{i=1}^{n} \frac{\Delta_i}{\overline{G}_n(V_i-)} V_i^2, \tag{4.3}$$

where $\widehat{Y}_i = \hat{g}_0^*(\hat{\beta}^T X_i) + \hat{g}_1^*(\hat{\beta}^T X_i) Z_i$ is a prediction value of Y_i, $1 \le i \le n$, and $G_n(\cdot)$ is defined in (2.6). R^2_{new} and the coefficient of determination in the linear model, R^2, are identical in the sense of the goodness of fit. For model (4.2), $R^2_{\text{new}} = 0.9443$. The value of R^2_{new} is very hight, indicating that the model (4.2) is reasonable to model this environmental data set.

5 Concluding remarks

We in this paper proposed a bias-corrected empirical log-likelihood ratio statistic for the regression parameter in model (1.1), and showed that the ratio is asymptotically standard chi-squared. We constructed the estimators for both the regression parameter and the coefficient functions. The asymptotic properties of these estimators were proved, and the consistent estimators of asymptotic variances were also constructed. The obtained results can be directly used to construct the confidence regions/intervals of the regression parameter and the pointwise confidence intervals of the coefficient functions. Our study achieved two aims. One is the bias correction directly inside the EL ratio, rather than multiplying an adjustment factor externally, which reflects the essential characteristics of EL, and the constructed EL ratio is asymptotically standard chi-squared. The second is a bias correction method to avoid undersmoothing the coefficient functions, and the optimal bandwidth can be selected by data-driven method. Simulation study and real data analysis demonstrate the superiority and utility of the proposed method. Our approach can also be used to study other semiparametric regression models with right censored data, such as partially linear single-index model, partially linear varying coefficient model, etc.

Acknowledgments: The research was supported by the National Natural Science Foundation of China (11971001). The data set used was provided by professor Riquan Zhang of the East China Normal University.

Appendix A: Proofs of Theorems 1–4

In this appendix, we prove Theorems 1–3 and 5. The proof of Theorem 4 is similar to the proof of Theorem 3, and therefore omit its proof. The following Lemmas 1–4 are useful for proving these Theorems. Lemma 1 is the lemma A.1 of Wang et al. (2010), and it can also be used when the variable t is removed. The proof of Lemma 2 is similar to the proof of Theorem 3.1 in He et al. (2016), and hence the details are omitted. The proofs of Lemmas 3 and 4 can be found in the supplementary material.

Lemma 1. Suppose that $\{\xi_i(u,\beta), 1 \le i \le n\}$ are random variables, and satisfy the following two conditions:

$$\frac{1}{n}\sum_{i=1}^{n} |\xi_i(u,\beta) - \xi_i(u_0,\beta_0)| \le cn^a(|u-u_0| + \|\beta-\beta_0\|) \tag{A.1}$$

for some constants $c > 0$, $a \ge 0$, u_0 and β_0;

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i(u,\beta)\right| > \varepsilon_n\right) \le \frac{1}{2} \tag{A.2}$$

for $\beta \in \mathcal{B}_n$ and $\varepsilon_n > 0$ depend only on n. Then

$$P\left(\sup_{(u,\beta)\in\mathcal{U}\times\mathcal{B}_n}\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i(u,\beta)\right| > \varepsilon_n\right) \le cn^{2pa}\varepsilon_n^{-2p}E\left\{\sup_{\beta\in\mathcal{B}_n}\exp\left(\frac{-n^2\varepsilon_n^2/128}{\sum_{i=1}^{n}\xi_i^2(\beta)}\right)\wedge 1\right\},$$

where $\mathcal{U}$ is a compact support set of the distribution of U, $\mathcal{B}_n = \{\beta \mid \beta \in R^p, \|\beta-\beta_0\| \le c_1 n^{-1/2}\}$, and c_1 is a positive constant.

Lemma 2. Suppose that the sample $\{(W_i, V_i, \Delta_i), 1 \le i \le n\}$ from (W, V, Δ) are independent and identically distributed, and $F(y) = P(Y \le y)$ and $G(t) = P(C \le t)$ are continuous, where Y and C are the random variables, $V = \min\{Y, C\}$ and $\Delta = I(Y \le C)$. If for each fixed $\theta \in \Theta$, $\zeta(w, v; \theta)$ is a measurable function of (w, v), and the condition

$$\int \frac{\zeta^2(w, v; \theta)}{\overline{G}(v)} F(\mathrm{d}w, \mathrm{d}v) < \infty$$

hold, then uniformly on $\theta \in \Theta$,

$$\int \zeta(w, v; \theta) F_n(\mathrm{d}w, \mathrm{d}v) = \frac{1}{n} \sum_{i=1}^{n} \frac{\Delta_i}{\overline{G}(V_i)} \zeta(W_i, V_i; \theta) + o_P(n^{-1/2}),$$

where $F(w, y)$ is the joint distribution function of (W, Y) and $F_n(w, v)$ is defined in (2.13).

Lemma 3. Suppose that conditions (C1)–(C7) hold. Then

$$\sup_{\beta^{(r)} \in \mathcal{B}_n^*} \|\widehat{Q}^*(\beta^{(r)}) - Q_n(\beta^{(r)})\| = o_P(n^{-1/2}),$$

where $\mathcal{B}_n^* = \{\beta^{(r)} | \, \|\beta^{(r)} - \beta_0^{(r)}\| \le c_2 n^{-1/2}\}$ for a positive constant c_2, $\widehat{Q}^*(\beta^{(r)})$ is defined in (2.17) and

$$\begin{aligned} Q_n(\beta^{(r)}) = {} & \frac{1}{n} \sum_{i=1}^{n} \frac{\Delta_i}{\overline{G}_n(V_i-)} w(\beta^T X_i) \{V_i - \boldsymbol{g}^T(\beta^T X_i) Z_i\} \\ & \times \dot{\boldsymbol{g}}^T(\beta^T X_i) Z_i J_{\beta^{(r)}}^T \{X_i - \mu(\beta^T X_i, Z_i)\}. \end{aligned}$$

Lemma 4. Suppose that conditions (C1)–(C7) hold. Then

$$\sqrt{n} Q_n(\beta_0^{(r)}) \xrightarrow{D} N(0, B), \tag{A.3}$$

$$\frac{1}{n} \sum_{i=1}^{n} \hat{\eta}_i(\beta_0^{(r)}) \hat{\eta}_i^T(\beta_0^{(r)}) \xrightarrow{P} B \tag{A.4}$$

and

$$\max_{1 \le i \le n} |\hat{\eta}_i(\beta_0^{(r)})| = o_P(n^{1/2}). \tag{A.5}$$

where $\hat{\eta}_i(\beta_0^{(r)})$, $Q_n(\beta_0^{(r)})$ and B are defined in (2.15), Lemma 3 and Theorem 3, respectively.

Proof of Theorem 1. Let

$$S_{n,k}(u; \beta) = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{\beta^T X_i - u}{h} \right)^k Z_i Z_i^T K_h(\beta^T X_i - u), \quad k = 0, 1, 2, 3.$$

Then, we have

$$S_n(u; \beta) \equiv D^T(u; \beta) \Omega(u; \beta) D(u; \beta) = \begin{pmatrix} S_{n,0}(u; \beta) & S_{n,1}(u; \beta) \\ S_{n,1}(u; \beta) & S_{n,2}(u; \beta) \end{pmatrix}, \tag{A.6}$$

where $D(u;\beta)$ and $\Omega(u;\beta)$ are defined in (2.7). By Lemma 1, we can prove

$$S_{n,k}(u;\beta) - E\{S_{n,k}(u;\beta)\} = O_P((nh/\log n)^{-1/2}), \quad k = 0,1,2,3,4.$$

uniformly for $u \in \mathcal{U}_w$ and $\beta \in \mathcal{B}_n$. A simple calculation yields

$$E\{S_{n,k}(u;\beta)\} = \kappa_k \Psi(u;\beta) + O_P(h), \quad k = 0,1,2,3,$$

uniformly for $u \in \mathcal{U}_w$ and $\beta \in \mathcal{B}_n$, where $\Psi(u;\beta) = f_\beta(u) E(ZZ^T|\beta^T X = u)$, $\kappa_k = \int_{-1}^{1} t^k K(t) dt$, $k = 0,1,2,3$, and $f_\beta(u)$ is a probability density function of $\beta^T X$. Therefore, we have

$$S_{n,k}(u;\beta) = \kappa_k \Psi(u;\beta) + O_P(c_n), \quad k = 0,1,2,3,4,$$

uniformly for $u \in \mathcal{U}_w$ and $\beta \in \mathcal{B}_n$, where $c_n = (nh/\log n)^{-1/2} + h$. Note that $\kappa_0 = 1$ and $\kappa_1 = 0$. By (A.6), we can obtain

$$S_n(u;\beta) = \Psi(u;\beta) \otimes \begin{pmatrix} 1 & 0 \\ 0 & \kappa_2 \end{pmatrix} + O_P(c_n),$$

uniformly for $u \in \mathcal{U}_w$ and $\beta \in \mathcal{B}_n$, where $\otimes$ is the Kronecker product. Using the fact

$$(A + hB)^{-1} = A^{-1} - hA^{-1}BA^{-1} + O(h^2)$$

for any matrices A and B, we have

$$S_n^{-1}(u;\beta) = \Psi^{-1}(u;\beta) \otimes \begin{pmatrix} 1 & 0 \\ 0 & \kappa_2^{-1} \end{pmatrix} + O_P(c_n), \tag{A.7}$$

uniformly for $u \in \mathcal{U}_w$ and $\beta \in \mathcal{B}_n$. Let

$$\xi_n(u;\beta) = \begin{pmatrix} \xi_{n,0}(u;\beta) \\ \xi_{n,1}(u;\beta) \end{pmatrix} \quad \text{and} \quad \xi_n^*(u;\beta) = \begin{pmatrix} \xi_{n,0}^*(u;\beta) \\ \xi_{n,1}^*(u;\beta) \end{pmatrix}.$$

where

$$\xi_{n,l}(u;\beta) = \frac{1}{n}\sum_{i=1}^{n} \left(\frac{\beta^T X_i - u}{h}\right)^l K_h(\beta^T X_i - u) Z_i V_{iG_n}, \quad l = 0,1$$

and

$$\xi_{n,l}^*(u;\beta) = \frac{1}{n}\sum_{i=1}^{n} \left(\frac{\beta^T X_i - u}{h}\right)^l K_h(\beta^T X_i - u) Z_i \{V_{iG_n} - Z_i^T \boldsymbol{g}(\beta^T X_i)\}, \quad l = 0,1.$$

Using Taylor's expansion for $\boldsymbol{g}(\cdot)$ at u, we can derive

$$\xi_{n,l}(u;\beta) - \xi_{n,l}^*(u;\beta) = \boldsymbol{g}(u)S_{n,l}(u;\beta) + h\dot{\boldsymbol{g}}(u)S_{n,l+1}(u;\beta) \\ + \frac{1}{2}h^2\ddot{\boldsymbol{g}}(u)S_{n,l+2}(u;\beta) + o_P(h^2), \quad l = 0, 1,$$

uniformly for $u \in \mathcal{U}_w$ and $\beta \in \mathcal{B}_n$. So that

$$\xi_n(u;\beta) - \xi_n^*(u;\beta) = S_n(u;\beta)\begin{pmatrix} \boldsymbol{g}(u) \\ h\dot{\boldsymbol{g}}(u) \end{pmatrix} + \frac{1}{2}h^2\ddot{\boldsymbol{g}}(u)\begin{pmatrix} S_{n,2}(u;\beta) \\ S_{n,3}(u;\beta) \end{pmatrix} + o_P(h^2),$$

uniformly for $u \in \mathcal{U}_w$ and $\beta \in \mathcal{B}_n$, where $S_n(u;\beta)$ is defined by (A.6). Thus it follows, from (2.7) and (A.7), that

$$\begin{pmatrix} \hat{\boldsymbol{g}}(u;\beta) - \boldsymbol{g}(u) \\ h\{\hat{\dot{\boldsymbol{g}}}(u;\beta) - \dot{\boldsymbol{g}}(u)\} \end{pmatrix} = \Psi^{-1}(u;\beta)\xi_n^*(u;\beta) + \frac{1}{2}h^2\ddot{\boldsymbol{g}}(u)\begin{pmatrix} \kappa_2 \\ \frac{\kappa_3}{\kappa_2} \end{pmatrix} + o_P(h^2). \qquad \text{(A.8)}$$

uniformly for $u \in \mathcal{U}_w$ and $\beta \in \mathcal{B}_n$. If we prove

$$\xi_n^*(u;\beta) = O_P((nh/\log n)^{-1/2}) + O_P(n^{-1/2}), \qquad \text{(A.9)}$$

uniformly for $u \in \mathcal{U}_w$ and $\beta \in \mathcal{B}_n$, combining (A.8), (A.9) and condition (C3), we can complete the proof of Theorem 1. We now prove (A.9). For the K-M estimator in (2.6), we have

$$\frac{1}{\overline{G}_n(V_i-)} = \frac{1}{\overline{G}(V_i)} + \frac{G_n(V_i-) - G(V_i)}{\overline{G}^2(V_i)} + \frac{\{G_n(V_i-) - G(V_i)\}^2}{\overline{G}_n(V_i-)\overline{G}^2(V_i)}. \qquad \text{(A.10)}$$

By the results of Gill (1983) and Zhou (1992), we have

$$\sup_{v \le V_{(n)}} |G_n(v) - G(v)| = O_P(n^{-1/2}) \qquad \text{(A.11)}$$

and

$$W_n = \sup_{v \le V_{(n)}} \left| \frac{G_n(v-) - G(v)}{1 - G_n(v-)} \right| = O_P(1), \qquad \text{(A.12)}$$

where $V_{(n)} = \max\{V_1, \ldots, V_n\}$. Using (A.10)–(A.12) and Lemma 1, and similar to the above proof, we can prove (A.9). This completes the proof of Theorem 1. $\square$

Proof of Theorem 2. By the Lagrange multiplier method, $\hat{l}(\beta_0^{(r)})$ can be represented as

$$\hat{l}(\beta_0^{(r)}) = 2\sum_{i=1}^{n} \log\left(1 + \lambda^T \hat{\eta}_i(\beta_0^{(r)})\right), \qquad \text{(A.13)}$$

where $\lambda = \lambda(\beta_0^{(r)})$ is a $p \times 1$ vector given as the solution to

$$h(\lambda) = \frac{1}{n}\sum_{i=1}^{n} \frac{\hat{\eta}_i(\beta_0^{(r)})}{1 + \lambda^T \hat{\eta}_i(\beta_0^{(r)})} = 0.$$

It can be shown that for large n, $h(\lambda) = 0$ has a unique solution λ_n such that $\lambda_n^T \hat{\eta}_i(\beta_0^{(r)}) > -1$ for all i. Its proof is similar to the proof of Theorem 4.2 in He et al. (2016), and hence the details are omitted. By (2.17), Lemma 3 and (A.3) of Lemma 4, we have

$$\sqrt{n}\widehat{Q}(\beta_0^{(r)}) \xrightarrow{D} N(0, B(\beta_0^{(r)})). \tag{A.14}$$

Therefore, From (A.4), (A.5) and (A.14), and using the same arguments as are used in the proof of (2.14) in Owen (1990), we can show that

$$\lambda = O_P(n^{-1/2}). \tag{A.15}$$

Applying the Taylor formula to (A.13), and invoking (A.4), (A.5), (A.14) and (A.15), we get

$$\hat{l}(\beta_0^{(r)}) = 2\sum_{i=1}^{n} \left[\lambda^T \hat{\eta}_i(\beta_0^{(r)}) - \{\lambda^T \hat{\eta}_i(\beta_0^{(r)})\}^2/2\right] + o_P(1). \tag{A.16}$$

Note that $h(\lambda) = 0$. It follows that

$$\begin{aligned} 0 &= \sum_{i=1}^{n} \frac{\hat{\eta}_i(\beta_0^{(r)})}{1 + \lambda^T \hat{\eta}_i(\beta_0^{(r)})} \\ &= \sum_{i=1}^{n} \hat{\eta}_i(\beta_0^{(r)}) - \sum_{i=1}^{n} \hat{\eta}_i(\beta_0^{(r)}) \hat{\eta}_i^T(\beta_0^{(r)}) \lambda + \sum_{i=1}^{n} \frac{\hat{\eta}_i(\beta_0^{(r)}) \{\lambda^T \hat{\eta}_i(\beta_0^{(r)})\}^2}{1 + \lambda^T \hat{\eta}_i(\beta_0^{(r)})}. \end{aligned}$$

This, together with (A.4), (A.5), (A.14) and (A.15), proves that

$$\sum_{i=1}^{n} \{\lambda^T \hat{\eta}_i(\beta_0^{(r)})\}^2 = \sum_{i=1}^{n} \lambda^T \hat{\eta}_i(\beta_0^{(r)}) + o_P(1) \tag{A.17}$$

and

$$\lambda = \left(\sum_{i=1}^{n} \hat{\eta}_i(\beta_0^{(r)}) \hat{\eta}_i^T(\beta_0^{(r)})\right)^{-1} \sum_{i=1}^{n} \hat{\eta}_i(\beta_0^{(r)}) + o_P(n^{-1/2}). \tag{A.18}$$

Therefore, from (A.16)–(A.18) we have

$$\hat{l}(\beta_0^{(r)}) = \{\sqrt{n}\widehat{Q}^T(\beta_0^{(r)})\} \left(\frac{1}{n}\sum_{i=1}^{n} \hat{\eta}_i(\beta_0^{(r)}) \hat{\eta}_i^T(\beta_0^{(r)})\right)^{-1} \{\sqrt{n}\widehat{Q}(\beta_0^{(r)})\} + o_P(1). \tag{A.19}$$

This, together with (A.19), (A.4), (A.14) and Slutsky's theorem, proves Theorem 2. $\square$

Proof of Theorem 3. We now prove the asymptotic normality of $\hat{\beta}$. The proof is divided into two steps: step (I) proves the existence of the estimator $\hat{\beta}$, and step (II) proves the asymptotic normality of $\hat{\beta}$.

(I) **Existence.** We prove the following fact: Under conditions (C1)–(C7) and with probability one there exists an estimator of β_0 solving the estimating equations (2.18) in $\mathcal{B}_n^{**}$, where $\mathcal{B}_n^{**} = \{\beta^{(r)} | \, \|\beta^{(r)} - \beta_0^{(r)}\| = Mn^{-1/2}\}$ for some constant M such that $0 < M < \infty$. From Lemma 3 and (A.10)–(A.12), we obtain

$$\begin{aligned}\widehat{Q}^*(\beta^{(r)}) &= Q_n(\beta^{(r)}) + o_P(n^{-1/2}) \\ &= Q_n(\beta_0^{(r)}) - A_n(\beta^{(r)} - \beta_0^{(r)}) + o_P(n^{-1/2}) \qquad \text{(A.20)}\end{aligned}$$

uniformly for $\beta^{(r)} \in \mathcal{B}_n^{**}$, where

$$A_n = \frac{1}{n}\sum_{i=1}^n \frac{\Delta_i w(\beta_0^T X_i)}{\overline{G}_n(V_i-)} \{\dot{\boldsymbol{g}}^T(\beta_0^T X_i) Z_i\}^2 J_{\beta_0^{(r)}}^T \{X_i - \mu(\beta_0^T X_i, Z_i)\}^{\otimes 2} J_{\beta_0^{(r)}}.$$

By Lemma 2 and the law of large numbers, we have

$$A_n = A + o_P(1), \qquad \text{(A.21)}$$

where A is defined in Theorem 3. From (A.20) and (A.21) we get

$$\widehat{Q}^*(\beta^{(r)}) = Q_n(\beta_0^{(r)}) - A(\beta^{(r)} - \beta_0^{(r)}) + o_P(n^{-1/2}). \qquad \text{(A.22)}$$

uniformly for $\beta^{(r)} \in \mathcal{B}_n^{**}$. Therefore, we have

$$n(\beta^{(r)} - \beta_0^{(r)})\widehat{Q}^*(\beta^{(r)}) = n(\beta^{(r)} - \beta_0^{(r)})Q_n(\beta_0^{(r)}) - n(\beta^{(r)} - \beta_0^{(r)})A(\beta^{(r)} - \beta_0^{(r)}) + o_P(1).$$

We note that the above formula is dominated by the term $\sim M^2$ because $\sqrt{n}\|\beta^{(r)} - \beta_0^{(r)}\| = M$, whereas $|n(\beta^{(r)} - \beta_0^{(r)})^T Q_n(\beta_0^{(r)})| = MO_P(1)$, and $n(\beta^{(r)} - \beta_0^{(r)})A(\beta^{(r)} - \beta_0^{(r)}) \sim M^2$. So, for any given $\eta > 0$, if M is chosen large enough, then it will follows that $n(\beta^{(r)} - \beta_0^{(r)})\widehat{Q}^*(\beta^{(r)}) < 0$ on an event with probability $1 - \eta$. From the arbitrariness of η, we can prove the existence of the estimator of β_0 in $\mathcal{B}_n^{**}$ as in the proof of Theorem 5.1 of Welsh (1989). The details are omitted.

(II) **Asymptotic normality.** From step (I) we find that $\hat{\beta}^{(r)}$ is a solution in $\mathcal{B}_n^{**}$ to the equation $\widehat{Q}^*(\beta^{(r)}) = 0$, namely $\widehat{Q}^*(\hat{\beta}^{(r)}) = 0$, where $\widehat{Q}^*(\beta^{(r)})$ is defined in (2.17). From (A.22) we have

$$0 = Q_n(\beta_0^{(r)}) - A(\hat{\beta}^{(r)} - \beta_0^{(r)}) + o_P(n^{-1/2}),$$

and hence

$$\sqrt{n}(\hat{\beta}^{(r)} - \beta_0^{(r)}) = A^{-1}\sqrt{n}Q_n(\beta_0^{(r)}) + o_P(1). \tag{A.23}$$

For the estimator $\hat{\beta}$, according to the calculation in Wang at el. (2010), we known

$$\hat{\beta} - \beta_0 = J_{\beta_0^{(r)}}(\hat{\beta}^{(r)} - \beta_0^{(r)}) + O_P(n^{-1}).$$

Therefore, we have

$$\sqrt{n}(\hat{\beta} - \beta_0) = J_{\beta_0^{(r)}}A^{-1}\sqrt{n}Q_n(\beta_0^{(r)}) + o_P(1). \tag{A.24}$$

This, together with (A.23), (A.24), (A.3) of Lemma 4 and Slutsky's theorem, proves Theorem 3. □

Proof of Theorem 5 By (A.8) and (2.9), we can obtain

$$\begin{aligned}
&\sqrt{nh}\left\{\hat{\boldsymbol{g}}^*(u_0) - \boldsymbol{g}(u_0) - 0.5\kappa_2 h^2 \ddot{\boldsymbol{g}}(u_0)\right\} \\
&= \Psi^{-1}(u_0)(nh)^{-1/2}\sum_{i=1}^{n}\{V_{iG} - Y_i + \varepsilon_i\}Z_i K\left(\frac{\beta_0^T X_i - u_0}{h}\right) + o_P(1) \\
&\equiv \Psi^{-1}(u_0;\beta_0)R_n(u_0;\beta_0) + o_P(1), && \text{(A.25)}
\end{aligned}$$

where $\Psi(u_0;\beta_0) = f_{\beta_0}(u_0)E(ZZ^T|\beta_0^T X = u_0)$. It is not difficult to prove $E\{R_n(u_0;\beta_0)\} = 0$ and $\mathrm{var}((R_n(u_0;\beta_0)) = \Omega(u_0;\beta_0) + o(1)$. We can check that $R_n(u_0;\beta_0)$ satisfies the conditions of the Cramér-Wold theorem and the Lindeberg condition (Serfling, 1980). Therefore, we get that

$$R_n(u_0;\beta_0) \xrightarrow{D} N(0, \Omega(u_0;\beta_0)). \tag{A.26}$$

This, together with (A.25), (A.26) and Slutsky's theorem, proves Theorem 5. □

References

Arnold, S. f. (1981). *The Theory of Linear Models and Multivariate Analysis.* John Wiley & Sons, New York.

Cai, Z. W., Fan, J. Q. and Li, R. Z. (2000a). Efficient estimation and inferences for varying-coefficient models. *J. Amer. Statist. Assoc.* **95**, 888–902.

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

Chen, R. and Tsay, S. (1993). Functional-coefficient autoregressive models. *J. Amer. Statist. Assoc.* **88**, 298–308.

Chiou, J. M. and Müller, H. G. (1998). Quasi-likelihood regression with unknown link and variance functions. *J. Amer. Statist. Assoc.* **93**, 1376–1387.

Cui, X., Härdle, W. and Zhu, L. X. (2011). The EFM approach for single-index models. *Ann. Statist.*, 39(3): 1658–1688.

Diggle, P. J. (1990). *Time Series: A Biostatistical Introduction.* Oxford University Press, Oxford. 387.

Fan, J. Q., Yao, Q. W. and Cai, Z. W. (2003). Adaptive varying-coefficient linear models. *J. Roy. Statist. Soc. ser. B* **65**, 57–80.

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

Gill, R. D. (1983). Large sample behavior of the product-limit estimator on the whole line. *Ann. Statist.* **11**, 49–58.

Härdle, W., Hall, P. and Ichimura, H. (1993). Optimal smoothing in single-index models. *Ann. Statist.* **21**, 157–178.

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

He, S. Y., Liang, W., Shen, J. S. and Yang, G. (2016). Empirical likelihood for right censored lifetime data. *J. Amer. Statist. Assoc.*, **111**, 646–655.

Huang, J. Z. and Shen, H. P. (2004). Functional Coefficient Regression Models for Nonlinear Time series: A Polynomial Spline Approach. *Scandinavian J. Statist.* **31**, 515–535.

Huang, Z. S. and Zhang, R. Q. (2013). Profile empirical-likelihood inferences for the single-index-coefficient regression model. *Statist. Comput.*, **23**, 455–465.

Hristache, M., Juditsky, A. and Spokoiny, V. (2001). Direct estimation of the index coefficient in a single-index model. *Ann. Statist.* **29**, 595–623.

Ichimura, H. (1993). Semiparametric least squares(SLS) and weighted SLS estimation of single-index models. *J. Econometrics* **58**, 71–120.

Li, K. C. (1991). Sliced inverse regression for dimension reduction (with discussion). *J. Amer. Statist. Assoc.* **86**, 316–342.

Lu, Z. D., Tjøstheim, D. and Yao, Q. W. (2007). Adaptive varying-coefficient linear models for stochastic processes: Asymptotic theory. *Statist. Sinica* **17**, 177–197.

Owen, A. B. (1990). Empirical likelihood ratio confidence regions. *Ann. Statist.* **18**, 90–120.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* New York: John Wiley & Sons.

Stute, W. (1982). A law for the logarithm of kernel density estimators. *Ann. Prob.* **10**, 414–422.

Wang, J. L., Xue, L. G., Zhu, L. X. and Chong, Y. S. (2010). Estimation for a partial-linear single-index model. *Ann. Statist.* **38**, 246–274.

Wang, Q. H. and Xue L. G. (2011) Statistical inference partially-varying-coefficient single-index model. *J. Mult. Anal.*, **102**, 1–19.

Weisberg, S. and Welsh, A. H. (1994). Adapting for the Missing Linear Link. *Ann. Statist.* **22**, 1674–1700.

Welsh, A. H. (1989). On M-processes and M-estimation. *Ann. Statist.* **17** 337–361. [Correction(1990) **18** 1500.]

Wu, C. O., Chiang, C. T. and Hoover, D. R. (1998). Asymptotic confidence regions for kernel smoothing of a varying-coefficient model with longitudinal data. *J. Amer. Statist. Assoc.* **93**, 1388–1402.

Xia, Y. C. and Li, W. K. (1999). On single-index coefficient regression models. *J. Amer. Statist. Assoc.* **94**, 1275–1285.

Xia, Y., Tong, H. Li, W. K. and Zhu L. X. (2002). An adaptive estimation of dimension reduction space. *J. R. Statist. Soc.* B **64**, 363–410.

Xue, L. G. and Pamg Z. (2013). Statistical inference for a single-index varying-coefficient model. *Statist. Comput.*, **23** 589–599.

Xue, L. G. and Wang, Q. H. (2012). Empirical likelihood for single-index varying-coefficient models. *Bernoulli*, **18**, 836–856.

Xue, L. G. and Zhu, L. X. (2006). Empirical likelihood for single-index model. *J. Mult. Anal.* **97**, 1295–1312.

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

Yu, Y. and Ruppert, D. (2002). Penalized spline estimation for partially linear single-index models. *J. Amer. Statist. Assoc.* **97**, 1042–1054.

Zhou, M. (1992). Asymptotic normality of the 'synthetic data' regression estimator for censored survival data. *Ann. Statist.* **20**, 1002–1021.

Zhu, L. X. and Fang, K. T. (1996). Asymptotics for the kernel estimates of sliced inverse regression. *Ann. Statist.* **24**, 1053-1067.