Sun and Wang *Journal of Inequalities and Applications* (2020) 2020:75
https://doi.org/10.1186/s13660-020-02324-w

Journal of Inequalities and Applications
a SpringerOpen Journal

RESEARCH Open Access

L_1-Estimation for covariate-adjusted regression

Yaodong Sun[1,2] and Dehui Wang[1*]

*Correspondence:
wangdh@jlu.edu.cn
[1] Institute of Mathematics, Jilin University, Changchun, P.R. China
Full list of author information is available at the end of the article

Abstract

We study a covariate-adjusted regression(CAR) model that is proposed for such situations where both predictors and response in a regression model are not directly observable but are distorted by a multiplicative factor that is determined by an unknown function of some observable covariate. By establishing a connection to varying-coefficient models, we present the local linear L_1-estimation method when the underlying error distribution deviates from a normal distribution. The robust estimators of parameters are proposed in the underlying regression model. The consistency and asymptotic normality of the robust estimators are investigated. Since the limit distribution depends on the unknown components of the errors, an empirical likelihood ratio method based on L_1 estimator is proposed. The confidence intervals for the regression coefficients are constructed. Simulation results demonstrate the superiority of the proposed estimators over other classical estimators when the underlying errors have heavy tails. Pima Indian diabetes data set is conducted to illustrate the performance of the proposed method, where the response and predictors are potentially contaminated by body mass index.

MSC: Covariate-adjusted regression; Least absolute deviation estimation; Asymptotic normality; Local linear estimate

1 Introduction

Covariate-adjusted regression(CAR) was initially proposed for regression analysis by Sentürk and Müller [18], where both the response and predictors are not directly observed. The available data are distorted by unknown functions of some common observable covariate. An example is the fibrinogen data collected on 69 hemodialysis patients, where the regression of fibrinogen level on serum transferrin level is of interest in Kaysen et al. [11]. Both response and predictor are known to be influenced in a multiplicative fashion by body mass index, defined as weight/height^2. Based on the observation, Sentürk and Müller [18] suggested that the confounding variable affects the primary variables through a flexible multiplicative unknown function. Such way of adjustment may reduce non-negligible bias and lead to consistent estimators of the parameters of interest, which is through dividing by the body mass index identified as a common confounder. For the simple case of two variables of interest, the underlying variables are

$$Y = \frac{\tilde{Y}}{\psi(U)}, \qquad X = \frac{\tilde{X}}{\phi(U)}, \tag{1.1}$$

$U \perp\!\!\!\perp (Y, X)$, where "$\perp\!\!\!\perp$"indicates independence, Y and X are the unobservable continuous variables of interest, while $\tilde{Y}$ and $\tilde{X}$ are available distorted variables. U is an observed continuous scalar confounding variable, $\psi(\cdot)$ and $\phi(\cdot)$ denote unknown smooth contaminating functions of U. The main goal is to uncover the relationship between response Y and covariate X, based on the confounding variable U, and on the contaminate variables $\tilde{Y}$ and $\tilde{X}$. Sentürk and Müller [18] considered the simple linear regression model

$$Y = \gamma_0 + \gamma_1 X + e, \tag{1.2}$$

where γ_0 and γ_1 are unknown parameters, e is the error term, $e \perp\!\!\!\perp (X, U)$. Reasonable assumption for $\psi(\cdot)$ and $\phi(\cdot)$ is that the mean distorting effect vanishes, that is,

$$E\bigl(\psi(U)\bigr) = 1, \qquad E\bigl(\phi(U)\bigr) = 1. \tag{1.3}$$

The central objective, based on the observation of the confounding variable U and the distorted observations $(\tilde{Y}, \tilde{X})$ in (1.1) is to estimate the unknown parameters γ_0 and γ_1.

To eliminate the effect caused by distortions, Sentürk and Müller [19] proposed covariate-adjusted varying coefficient model(CAVCM) to target the covariate-adjusted relationship between longitudinal variables. Sentürk and Nguyen [20] proposed the estimation procedures based on local polynomial smoothing technique (*LP*) for the model. Cui et al. [3] considered the covariate-adjusted nonlinear regression and proposed a direct plug-in estimation procedure for the model. Li et al. [12] studied covariate-adjusted partially linear regression models and obtained confidence intervals for the regression coefficients.

According to model (1.1)–(1.3), the regression of $\tilde{Y}$ on $\tilde{X}$ can be expressed as

$$\tilde{Y} = \beta_0(U) + \beta_1(U)\tilde{X} + e(U), \tag{1.4}$$

where $\beta_0(U) = \psi(U)\gamma_0$, $\beta_1(U) = \gamma_1\psi(U)/\phi(U)$, $e(U) = \psi(U)e$.

This is a varying coefficient model with heteroscedasticity, that is, a useful extension of classical linear models. Varying coefficient models are widely used in diverse areas as the modeling bias can significantly be reduced and the "curse of dimensionality"problem can also be avoided. See, for example, Hastie and Tibshirani [8], Fan and Zhang [5–7]. Least-squares (*LS*) method is the popular approach in the vast literature on model (1.4), as *LS* method has favorable properties for a large class of error distributions. However, this method will break down when the random error is adversely affected by outliers and heavy-tail distributions. The robust estimation method is desired. In this article, we propose robust coefficient estimation motivated by Tang et al. [21]. We use a two-step estimation procedure to estimate the unknown parameters. Firstly, we use L_1-estimation to estimate varying coefficients based on local linear fit. Because model (1.4) is heteroscedastic, the inferring methods are not same. Secondly, the estimates of unknown parameters are constructed based on weighted averages of these functions. However, the limiting variance has a complex structure with several unknown components. An estimated empirical log-likelihood approach to construct the confidence region of the regression parameter is developed. An empirical log-likelihood ratio is proved to be asymptotically standard chi-square.

The rest of this article is organized as follows. In Sect. 2, we describe the L_1 estimation procedure and propose the estimation of both nonparametric and parametric components. We obtain the asymptotic results and discuss the efficiency of the estimators. In Sect. 3, we construct empirical likelihood based confidence regions for the parameters. Section 4 presents the hypothesis testing procedure. In Sects. 5 and 6, some simulations and empirical study are carried out to assess the performance of the proposed estimators and confidence regions. Section 7 concludes the paper with discussion. The proofs of theorems are deferred to Appendix.

2 Estimation and asymptotic behavior

Consider a covariate-adjusted regression model in the following general form:

$$\begin{cases} Y = \boldsymbol{X}^\tau \boldsymbol{\gamma} + e, \\ \tilde{Y} = \psi(U)Y, \\ \tilde{X}_r = \phi_r(U)X_r, \quad r = 1,\dots,p, \end{cases} \tag{2.1}$$

where $\boldsymbol{\gamma} = (\gamma_1,\dots,\gamma_p)^\tau$. Y and X_r, $r = 1,\dots,p$, are unobservable variables distorted by smooth function $\psi(U)$ and $\phi_r(U)$. $\tilde{Y}$, $\tilde{X}_r$, $r = 1,\dots,p$, and univariate confounder U are observable variables. e is random error with 0.5 quantile being zero, $E(\psi(U)) = 1$, $E(\phi_r(U)) = 1$, $r = 1,\dots,p$. Our goal is to estimate the unknown parameter $\boldsymbol{\gamma}$ consistently based the observed data, and to further establish asymptotic normality for the proposed estimators. The estimation of the regression coefficient $\boldsymbol{\gamma}$ is a two-step estimation procedure similar to that in Sentürk and Müller [18]. From model (2.1), we rewrite CAR into the following CAVCM:

$$\tilde{Y} = \tilde{\boldsymbol{X}}^\tau \boldsymbol{\beta}(U) + e(U), \tag{2.2}$$

where $\boldsymbol{\beta}(U) = (\beta_1(U),\dots,\beta_p(U))^\tau$, $\beta_r(U) = \gamma_r \frac{\psi(U)}{\phi_r(U)}$, $r = 1,\dots,p$, $e(U) = \psi(U)e$.

In the first step, we employ L_1-estimation to estimate varying coefficients $\boldsymbol{\beta}(U)$ based on local linear fit. For U in the neighborhood of u, we use a local linear approximation

$$\beta_r(U) \approx a_r(u) + a_r'(u)(U-u) \stackrel{\triangle}{=} a_r + b_r(U-u)$$

for $r = 1,\dots,p$.

Suppose that $\{U_i, \tilde{\boldsymbol{X}}_i, \tilde{Y}_i\}$, $i = 1,\dots,n$, are independent and identically distributed samples from model (2.1), $\tilde{\boldsymbol{X}}_i = (\tilde{X}_{i1},\dots,\tilde{X}_{ip})^\tau$. Let $(\hat{\boldsymbol{a}}^\tau, \hat{\boldsymbol{b}}^\tau)^\tau$ be the local linear L_1-estimate of $(\boldsymbol{a}^\tau, \boldsymbol{b}^\tau)^\tau$ by minimizing

$$\sum_{i=1}^n \bigl|\tilde{Y}_i - \tilde{\boldsymbol{X}}_i^\tau \bigl(\boldsymbol{a} + (U_i - u)\boldsymbol{b}\bigr)\bigr| K\bigl((U_i - u)/h\bigr), \tag{2.3}$$

where $\boldsymbol{a} = (a_1,\dots,a_p)^\tau$, $\boldsymbol{b} = (b_1,\dots,b_p)^\tau$.

In the second step, from (C1), (C2), and (2.1),

$$E(\tilde{X}_r) = E(X_r), \qquad E\big(\beta_r(U)\tilde{X}_r\big) = \gamma_r E(\tilde{X}_r), \quad r = 1,\dots,p.$$

The unknown regression parameters γ_r, $r = 1,\dots,p$, are obtained as averages of raw estimates $\hat{\beta}_r(U_i)$. The estimates are given by

$$\hat{\gamma}_r = \frac{1}{\bar{\tilde{X}}_r}\sum_{i=1}^{n}\frac{1}{n}\hat{\beta}_r(U_i)\tilde{X}_{ir}, \quad r = 1,\dots,p, \tag{2.4}$$

where $\bar{\tilde{X}}_r = \frac{1}{n}\sum_{i=1}^{n}\tilde{X}_{ir}$.

In this section, we establish the asymptotic properties of $\hat{\gamma}_r$.

Theorem 1 *Under the regularity conditions in Appendix, if $h \to 0$ and $nh \to \infty$ as $n \to \infty$, then*

$$\hat{\gamma}_r = \gamma_r + O_p\big(n^{-1/2}\big) + O_p(C_n), \quad r = 1,\dots,p, \tag{2.5}$$

where $C_n = O_p(h^2 + \log^{1/2}(1/h)/(nh))$.

Theorem 2 *Under the regularity conditions in Appendix, if $nh^2/\log(1/h) \to \infty$ and $nh^4 \to 0$ as $n \to \infty$, then the asymptotic distribution of $\hat{\gamma}_r$ is given by*

$$\sqrt{n}(\hat{\gamma}_r - \gamma_r) \xrightarrow{D} N\big(0,\sigma_r^2\big), \quad r = 1,\dots,p, \tag{2.6}$$

where

$$\begin{aligned}\sigma_r^2 = {} & \big\{\gamma_r^2 E\big(X_r^2\big)\operatorname{var}\big(\psi(U)\big) + \gamma_r^2\operatorname{var}(X_r) \\ & + 2\gamma_r^2\big[E\big(\phi_r(U)\psi(U)\big)E\big(X_r^2\big) - \big(E(X_r)\big)^2\big] + \gamma_r^2\operatorname{var}\big(\phi_r(U)X_r\big)\big\}/\big\{E(X_r)\big\}^2.\end{aligned}$$

The optimal bandwidth for $\hat{\beta}_r(\cdot)$ is $h \sim n^{-1/5}$. This bandwidth does not satisfy the condition in Theorem 2. In order to obtain the asymptotic normality for $\hat{\gamma}_r$, undersmoothing for $\hat{\beta}_r(\cdot)$ is necessary. The requirement has also been used in the literature for semiparametric model; see Carroll et al. [2] for a detailed discussion.

3 Empirical likelihood

Although we have obtained the asymptotic distribution of γ_r, the σ_r^2 is complex and includes several unknown components to be estimated. To resolve this difficulty, we propose an empirical likelihood method to construct a confidence interval for γ_r. For more information on the empirical likelihood estimation, we refer to Owen [16].

Note that $E((\beta_r(U_i) - \gamma_r)\tilde{X}_{ir}) = 0$ for $i = 1,2,\dots,n$, $r = 1,\dots,p$ if γ_r is the true parameter. Hence, the problem of testing whether γ_r is the true parameter is equivalent to testing whether $E((\beta_r(U_i) - \gamma_r)\tilde{X}_{ir}) = 0$. By Owen [15], to construct an empirical likelihood ratio function for γ_r, we denote $V_i(\gamma_r) = (\beta_r(U_i) - \gamma_r)\tilde{X}_{ir}$. That is, we can define the profile empirical likelihood ratio function

$$L_n(\gamma_r) = -2\max\left\{\sum_{i=1}^{n}\log(np_i)\,\middle|\, p_i \ge 0, \sum_{i=1}^{n}p_i = 1, \sum_{i=1}^{n}p_i V_i(\gamma_r) = 0\right\}.$$

It can be shown that $L_n(\gamma_r)$ is asymptotically chi-squared with 1 degree of freedom. However, $L_n(\gamma_r)$ cannot be directly used to make statistical inference on γ_r because $L_n(\gamma_r)$ contains the unknown $\beta_r(\cdot)$. A natural way is to replace $\beta_r(\cdot)$ by L_1-estimator $\hat{\beta}_r(\cdot)$ and to replace $V_i(\gamma_r)$ by $\hat{V}_i(\gamma_r)$. Then an estimated empirical likelihood ratio function is defined by

$$\hat{L}_n(\gamma_r) = -2\max\left\{\sum_{i=1}^n \log(np_i)\middle| p_i \geq 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i \hat{V}_i(\gamma_r) = 0\right\}.$$

By the Lagrange multiplier method, $\hat{L}_n(\gamma_r)$ can be represented as

$$\hat{L}_n(\gamma_r) = 2\sum_{i=1}^n \log\bigl(1 + \lambda \hat{V}_i(\gamma_r)\bigr), \tag{3.1}$$

where λ is determined by

$$\frac{1}{n}\sum_{i=1}^n \frac{\hat{V}_i(\gamma_r)}{1 + \lambda \hat{V}_i(\gamma_r)} = 0. \tag{3.2}$$

In the following, we show that $\log \hat{L}_n(\gamma_r)$ converges to the standard chi-square distribution with degree 1.

Theorem 3 *Under conditions of Theorem 2, we have*

$$\hat{L}_n(\gamma_r) \xrightarrow{D} \chi_1^2. \tag{3.3}$$

According to Theorem 3, we construct a $(1-\alpha)$-level confidence region of γ_r:

$$\mathrm{CR}_\alpha = \bigl\{\gamma_r : \hat{L}_n(\gamma_r) \leq c_\alpha\bigr\},$$

where c_α satisfies $P(\chi_1^2 \leq c_\alpha) = 1 - \alpha$.

4 Bootstrap test

It is often of practical interest to test for the significance of the regression coefficients. We consider the null hypothesis

$$H_0 : \beta_r(u) = c_r, \quad r = 1, \ldots, p, \tag{4.1}$$

where c_r is an unknown constant. Under the null hypothesis, the smooth estimator $\hat{\beta}_r(u)$ of $\beta_r(u)$ is expected to be close to a horizontal line. We average $\{\hat{\beta}_r(U_i)\}$ to obtain the estimator of parameter c_r. Similar to the statistics proposed by Cai, Fan, and Yao [1], the residual sum of squares under null hypothesis is

$$\mathrm{RSS}_0 = n^{-1}\sum_{i=1}^n \left|\tilde{Y}_i - \sum_{j\neq r}^p \hat{\beta}_j(U_i)\tilde{X}_{ij} - \hat{c}_r\tilde{X}_{ir}\right|.$$

Analogously, the the residual sum of squares corresponding to model (2.2) is

$$\mathrm{RSS}_1 = n^{-1}\sum_{i=1}^{n}\left|\tilde{Y}_i - \sum_{j=1}^{p}\hat{\beta}_j(U_i)\tilde{X}_{ij}\right|.$$

The goodness-of-fit test statistic is defined as

$$T_n = (\mathrm{RSS}_0 - \mathrm{RSS}_1)/\mathrm{RSS}_1 = \mathrm{RSS}_0/\mathrm{RSS}_1 - 1,$$

and we reject the null hypothesis (4.1) for large values of T_n. The distribution of T_n^* computed from the bootstrap samples is used as an approximation to the distribution of T_n. The p-value of the test is the relative frequency of the event $T_n^* \geq T_n$.

5 Simulation study

In this section, we carry out simulations to investigate the performance of our proposed methods as outlined in Sects. 2 and 3. We shall compare the finite sample performance of the *LP* procedure with our approach. The underlying (unobserved) multiple regression model considered is as follows:

$$Y = 3 + 0.1X_1 + 2X_2 - 0.2X_3 + e, \tag{5.1}$$

where the predictors X_1, X_2, and X_3 are distributed as $N(2, 1.5^2)$, $N(0.5, 0.25^2)$, $N(1, 1)$. For the distribution of the confounding variable U, it is generated from a uniform $[0, 1]$ distribution. The distorting functions considered are

$$\psi(U) = (U + 3)^2/a, \qquad \phi_1(U) = (U + 10)/b, \qquad \phi_3(U) = (U + 3)/c,$$

where (a, b, c) are (12.339, 10.5, 3.5) for $U \sim U[0, 1]$. The constants a, b, c are chosen such that the distorting functions satisfy the constraints in (1.3). In order to show the robustness of our estimators, the following different error distributions are considered: $N(0, 0.25)$, $t(3)$, and Cauchy$(0, 0.2)$. For the weight function, we use the Epanechnikov kernel, and the asymptotic optimal bandwidth h for *LP* has been considered in Senturk et al. [19]. We can produce simple formulas for the asymptotic optimal bandwidth h for L_1: $h_{L_1} = h_{LP}/\{4f(F^{-1}(0.5))\}^{1/5}$ motivated by Kai et al. [10]. We repeat the simulation 1000 times with sample sizes of 50, 100, and 200, respectively. The corresponding results are summarized in Table 1. As we can see from the table, when the error is normally distributed, the proposed L_1 estimators perform nearly as well as the *LP* estimators although they have slightly larger biases and standard deviations. However, for the other two non-normal errors, *LP* estimators are not as good as expected. And L_1 estimators have a significant improvement.

For the sample sizes, $n = 100$ and 200 samples are generated from the above simulated data, and for each sample, the 95% confidence intervals are computed using the empirical likelihood, which is reported in Table 2. When n increases, we see that the coverage probabilities increase.

Table 1 Summary of bias and standard deviation over 1000 simulations

n	Dist	Method	Bias (SD)			
			γ_0	γ_1	γ_2	γ_3
50	Normal	LP	−0.0071 (0.0730)	0.0003 (0.0093)	0.0129 (0.0920)	0.0003 (0.0130)
		L_1	−0.0083 (0.0767)	0.0002 (0.0072)	0.0134 (0.0939)	0.0008 (0.0132)
	t_3	LP	0.0114 (0.3651)	−0.0065 (0.1927)	0.0027 (0.4083)	0.0080 (0.3240)
		L_1	0.0213 (0.3566)	−0.0067 (0.1720)	0.0070 (0.3969)	0.0029 (0.3008)
	Cauchy	LP	0.0946 (32.7370)	0.8215 (35.8373)	−1.0257 (41.5098)	1.3761 (32.1225)
		L_1	−0.0861 (6.2061)	0.4566 (7.0982)	−0.6856 (8.31108)	−0.0929 (3.1853)
100	Normal	LP	−0.0063 (0.0830)	1.03e–05 (0.0078)	0.0116 (0.0685)	0.0004 (0.0091)
		L_1	−0.0079 (0.0798)	5.75e–05 (0.0083)	0.0080 (0.0762)	−0.0006 (0.0074)
	t_3	LP	−0.0249 (0.3126)	−0.0078 (0.0991)	0.0078 (0.2443)	0.0017 (0.1907)
		L_1	−0.0297 (0.2964)	0.0079 (0.0734)	0.0093 (0.1890)	0.0039 (0.1553)
	Cauchy	LP	1.2813 (26.0130)	0.4512 (20.6582)	−1.3174 (28.6553)	0.6500 (14.6341)
		L_1	0.0733 (0.5163)	0.0083 (0.1183)	0.043 (0.7523)	−0.0147 (0.3702)
200	Normal	LP	−0.0007 (0.0376)	3.87e–05 (0.0027)	0.0066 (0.0423)	−0.0002 (0.0043)
		L_1	−0.0024 (0.0383)	4.56e–05 (0.0024)	0.0046 (0.0427)	0.0005 (0.0041)
	t_3	LP	0.0139 (0.2856)	−0.0029 (0.0648)	−0.0059 (0.1711)	0.0015 (0.1347)
		L_1	−0.0123 (0.1844)	−0.0035 (0.0502)	0.0025 (0.1287)	0.0023 (0.1195)
	Cauchy	LP	1.7790 (23.2778)	−0.0714 (7.6870)	−1.1413 (20.6915)	−0.6269 (16.2315)
		L_1	0.0393 (0.2638)	−0.0030 (0.0678)	0.0012 (0.1530)	0.0059 (0.1407)

Table 2 Coverage probabilities of confidence regions when the nominal level is 0.95

n	Dist	γ_0	γ_1	γ_2	γ_3
100	Normal	0.8923	0.8265	0.9181	0.8370
	t_3	0.8437	0.7887	0.8396	0.7741
200	Normal	0.9182	0.8806	0.9273	0.8827
	t_3	0.8581	0.8140	0.8622	0.8029

6 Application

We illustrate the methodology via an application to the diabetes data set which contains eight-dimensional patterns to understand the prevalence of diabetes and other cardiovascular risk factors. The 131 subjects analyzed here are females at least 35 years old of Pima Indian heritage who were actually screened for diabetes. The female patients may have abnormal insulin action that prevents the body from normal utilization of glucose. Obesity is a risk factor in both diabetes and hypertension. One of the purposes of the study is to identify risk factors for diabetes, among which is hypertension. In this study, we investigate the relationship between plasma glucose (GLU) concentration and hypertensive measure, diastolic blood pressure (DBP). We analyze the simple linear regression relationship between GLU and DBP, $GLU = \gamma_0 + \gamma_1 DBP + e$. Body mass index ($BMI$) is identified to be a major factor significantly associated with elevated prevalence of hypertension and diabetes. Both the response and the predictor are potentially affected by body mass index (BMI), BMI = weight/height^2. The varying coefficient model has the following form $GLU = \beta_0(BMI) + \beta_1(BMI)DBP + e(BMI)$, based on the confounding variable BMI and the contaminate variables, GLU and DBP.

The parameters γ_0 and γ_1 are estimated by the covariate-adjusted regression algorithm. Three outliers are removed before the analysis. The p-values for covariate-adjusted regression estimates are obtained from 1000 bootstrap samples. For least squares regression, DBP was close to being significant, $p = 0.056$, while with covariate-adjusted regression it became highly significant, $p = 0.029$. Thus, the covariate-adjusted regression model is more appropriate for the data than least-squares regression. We shall compare the performance of the LP procedure with our approach. The LP estimates are $(\hat{\gamma}_0, \hat{\gamma}_1) = (72.1511, 0.5972)$. The L_1 estimates are $(\hat{\gamma}_0, \hat{\gamma}_1) = (74.2114, 0.6035)$. We estimate the standard deviation of γ_0 and γ_1 based on 1000 bootstrap samples for both these two methods. The corresponding standard deviation estimates of LP estimators are $\widehat{s.d.}(\hat{\gamma}_0) = 0.3129$ and $\widehat{s.d.}(\hat{\gamma}_1) = 0.0819$. The corresponding standard deviation estimates of L_1 estimators are $\widehat{s.d.}(\hat{\gamma}_0) = 0.2994$ and $\widehat{s.d.}(\hat{\gamma}_1) = 0.0611$. From the above, we can see that the difference between the estimated parameters based on LP modeling and L_1 modeling is relatively small; however, L_1 estimators have smaller standard errors than LP approach, which means that L_1 modeling has better performance. It is believed that the distortion effect of the obesity index on blood pressure is different from its effect on plasma glucose, and the distortion effect of the obesity index on GLU can be assessed directly from the estimated intercept function.

7 Discussion

In this paper, we propose a robust and efficient procedure for CAR, which has improved on the earlier proposed LP estimation when the underlying error distribution deviates from normal distribution, and the asymptotic normality has been established under some regular conditions. We propose a two-step estimation procedure considered for CAR. Firstly, we use L_1-estimation motivated by Tang et al. [21] to estimate varying coefficients based on local linear fit. The performance of the smoothing technique chosen for estimation of the varying coefficient functions in the first step does affect the overall performance of the CAR estimates in the second step. When the data contain outliers or come from population with heavy-tailed distributions, L_1-estimation should yield better estimators. Secondly, the estimates of unknown parameters are constructed based on weighted averages of these functions. In addition, an estimated empirical log-likelihood approach to construct the confidence region of the regression parameter is developed, and the confidence intervals for the regression coefficients are constructed. Finally, it is interesting to develop a robust and efficient variable selection procedure for the CAR in high dimension setting.

Appendix

To establish the main results given in Sects. 2 and 3, the following regularity conditions are imposed:

(C1) Contaminating functions $\psi(\cdot)$ and $\phi_r(\cdot)$ are twice continuously differentiable, satisfying $E(\psi(U_i)) = 1$, $E(\phi_r(U_i)) = 1$, $\phi_r(U_i) > 0$, $\psi(U_i) > 0$, $r = 1, \ldots, p$, $i = 1, \ldots, n$.

(C2) The variables $\boldsymbol{X}$, U, e are mutually independent, and the variables Y, U are mutually independent, $E(Y^2) < \infty$, $E(X_r^2) < \infty$, $r = 1, \ldots, p$.

(C3) The random variable U has bounded support Ω, $f_U(\cdot)$ is the density function of covariate U.

(C4) The kernel function $K(\cdot)$ is symmetric with a compact support and satisfies a Lipschitz condition.

(C5) Denoted by $f(\cdot)$ and $F(\cdot)$ are the density function and cumulative distribution of the error e, respectively. $f(\cdot)$ is bounded away from zero and has a continuous and uniformly bounded derivative. $F(0) = 0.5$.

(C6) $E(\tilde{\boldsymbol{X}}\tilde{\boldsymbol{X}}^{\tau}|U=u)$ is nonsingular for all $u \in \Omega$.

Remark These conditions are mild. Conditions (C1)–(C3) are assumed in Cui et al. [3]. Conditions (C4)–(C5) can be found in Tang et al. [21]. Condition (C6) can be found in Kai et al. [9].

In order to obtain our results, we first prove some lemmas.

Lemma 1 *Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be i.i.d. random vectors, where the Y_is are scalar random variables. Assume further that $E|Y|^r < \infty$ and that $\sup_X \int |y|^r f(x,y)\,dy < \infty$, where $f(x,y)$ denotes the joint density of (X, Y). Let $K(\cdot)$ be a bounded positive function with bounded support, satisfying a Lipschitz condition. Then*

$$\sup_{\boldsymbol{X}\in D}\left|n^{-1}\sum_{i=1}^{n}\left\{K_h(X_i-x)Y_i-E\left[K_h(X_i-x)Y_i\right]\right\}\right|=O_p\left(\frac{\log^{1/2}(1/h)}{\sqrt{nh}}\right),$$

provided that $n^{2\epsilon-1}h \to \infty$ for some $\epsilon < 1 - r^{-1}$, $K_h(\cdot) = K(\cdot/h)/h$.

The proof of Lemma 1 can be found in Mack and Silverman [13].

Lemma 2 *Under the regularity conditions* (C1)–(C6), *if $h \to 0$ and $nh \to \infty$ as $n \to \infty$, then*

$$\hat{\boldsymbol{\beta}}(U)-\boldsymbol{\beta}(U)=\frac{1}{\sqrt{nh}}O_p\left(h^2+\log^{1/2}(1/h)/\sqrt{nh}\right). \tag{A.1}$$

Proof Let $\boldsymbol{\eta} = (nh)^{1/2}(\boldsymbol{a} - \boldsymbol{\beta}(u))$, $\boldsymbol{\zeta} = (nh)^{1/2}h(\boldsymbol{b} - \boldsymbol{\beta}'(u))$, $\Delta_i = (nh)^{-1/2}\tilde{\boldsymbol{X}}_i^{\tau}[\boldsymbol{\eta} + h^{-1}(U_i-u)\boldsymbol{\zeta}]$, $r_i(u) = \tilde{\boldsymbol{X}}_i^{\tau}\{\boldsymbol{\beta}(U_i) - \boldsymbol{\beta}(u) - \boldsymbol{\beta}'(U_i-u)\}$. We recall $\{\hat{\boldsymbol{a}}^{\tau}, \hat{\boldsymbol{b}}^{\tau}\}^{\tau}$ minimizes

$$\begin{aligned}
&\sum_{i=1}^{n}\left|\tilde{Y}_i-\tilde{\boldsymbol{X}}_i^{\tau}\left(\boldsymbol{a}+(U_i-u)\boldsymbol{b}\right)\right|K\left((U_i-u)/h\right),\\
&\left(\hat{\boldsymbol{a}}^{\tau},\hat{\boldsymbol{b}}^{\tau}\right)^{\tau}=\operatorname*{argmin}_{\boldsymbol{a},\boldsymbol{b}}\sum_{i=1}^{n}\left|\tilde{Y}_i-\tilde{\boldsymbol{X}}_i^{\tau}\left(\boldsymbol{a}+(U_i-u)\boldsymbol{b}\right)\right|K\left((U_i-u)/h\right)\\
&\qquad=\operatorname*{argmin}_{\boldsymbol{a},\boldsymbol{b}}\sum_{i=1}^{n}\Big\{\Big|(nh)^{-1/2}\tilde{\boldsymbol{X}}_i^{\tau}\Big[(nh)^{1/2}\left(\boldsymbol{a}-\boldsymbol{\beta}(u)\right)\\
&\qquad\quad+h^{-1}(U_i-u)(nh)^{1/2}h\left(\boldsymbol{b}-\boldsymbol{\beta}'(u)\right)\Big]\\
&\qquad\quad-\left[\tilde{\boldsymbol{X}}_i^{\tau}\left\{\boldsymbol{\beta}(U_i)-\boldsymbol{\beta}(u)-\boldsymbol{\beta}'(U_i-u)\right\}+e(U_i)\right]\Big|\\
&\qquad\quad-\left|\tilde{\boldsymbol{X}}_i^{\tau}\left\{\boldsymbol{\beta}(U_i)-\boldsymbol{\beta}(u)-\boldsymbol{\beta}'(U_i-u)\right\}+e(U_i)\right|\Big\}K\left((U_i-u)/h\right),\\
&\qquad=\operatorname*{argmin}_{\boldsymbol{a},\boldsymbol{b}}\sum_{i=1}^{n}\left(\left|\Delta_i-r_i(u)-e(U_i)\right|-\left|r_i(u)+e(U_i)\right|\right)K\left((U_i-u)/h\right).
\end{aligned}$$

The last equality holds because the last term is free of the optimization variables $\boldsymbol{a}$ and $\boldsymbol{b}$. By applying the following identity:

$$|x-y|-|x|=y\left(2I(x\le 0)-1\right)+2\int_0^y\left\{I(x\le s)-I(x\le 0)\right\}ds,$$

$$
\begin{aligned}
&\sum_{i=1}^{n}\bigl(\bigl|r_i(u)+e(U_i)-\Delta_i\bigr|-\bigl|r_i(u)+e(U_i)\bigr|\bigr)K\bigl((U_i-u)/h\bigr)\\
&\quad=2\sum_{i=1}^{n}\Delta_i\bigl[I\bigl(e(U_i)\le -r_i(u)\bigr)-1/2\bigr]K\bigl((U_i-u)/h\bigr)\\
&\qquad+2\sum_{i=1}^{n}K\bigl((U_i-u)/h\bigr)\int_0^{\Delta_i}\bigl[I\bigl(e(U_i)+r_i(u)\le s\bigr)-I\bigl(e(U_i)+r_i(u)\le 0\bigr)\bigr]\,ds\\
&\quad=\bigl[2\boldsymbol{W}_n^{\tau}\bigl(\boldsymbol{\eta}^{\tau},\boldsymbol{\zeta}^{\tau}\bigr)^{\tau}+2B_n\bigr],
\end{aligned}
$$

where

$$
\begin{aligned}
\boldsymbol{W}_n^{\tau}&=(nh)^{-1/2}\sum_{i=1}^{n}K\bigl((U_i-u)/h\bigr)\bigl[I\bigl(e_i\le -r_i(u)/\psi(U_i)\bigr)-1/2\bigr]\tilde{\boldsymbol{X}}_i^{\tau}\left(I_p,\frac{U_i-u}{h}I_p\right),\\
B_n&=\sum_{i=1}^{n}K\bigl((U_i-u_0)/h\bigr)\int_0^{\Delta_i}\bigl[I\bigl(e(U_i)+r_i(u)\le s\bigr)-I\bigl(e(U_i)+r_i(u)\le 0\bigr)\bigr]\,ds.
\end{aligned}
$$

Since L_1-loss is a special case of quantile loss function at 0.5, the next proof is similar to that of Theorem 3.1 of Kai et al. [10]. By Lemma 1, we have

$$
B_n=E(B_n)+O_p\bigl(\log^{1/2}(1/h)/\sqrt{nh}\bigr). \tag{A.2}
$$

The conditional expectation of B_n can be calculated as follows:

$$
\begin{aligned}
&E(B_n|U,\boldsymbol{X})\\
&\quad=\sum_{i=1}^{n}K\bigl((U_i-u)/h\bigr)\int_0^{\Delta_i}\bigl\{F\bigl(s/\psi(U_i)-r_i(u)/\psi(U_i)\bigr)\bigr]-F\bigl(-r_i(u)/\psi(U_i)\bigr)\bigr\}\,ds\\
&\quad=\frac{1}{2}\bigl(\boldsymbol{\eta}^{\tau},\boldsymbol{\zeta}^{\tau}\bigr)\left(\frac{1}{nh}\sum_{i=1}^{n}\frac{1}{\psi(U_i)}K\bigl((U_i-u)/h\bigr)f\bigl(-r_i(u)/\psi(U_i)\bigr)\bigl(\tilde{\boldsymbol{X}}_i^{\tau},\tilde{\boldsymbol{X}}_i^{\tau}(U_i-u)/h\bigr)^{\tau}\right.\\
&\qquad\left.\times\bigl(\tilde{\boldsymbol{X}}_i^{\tau},\tilde{\boldsymbol{X}}_i^{\tau}(U_i-u)/h\bigr)\right)\bigl(\boldsymbol{\eta}^{\tau},\boldsymbol{\zeta}^{\tau}\bigr)^{\tau}+O_p\bigl(\log^{1/2}(1/h)/\sqrt{nh}\bigr)\\
&\quad\triangleq\frac{1}{2}\bigl(\boldsymbol{\eta}^{\tau},\boldsymbol{\zeta}^{\tau}\bigr)S_n\bigl(\boldsymbol{\eta}^{\tau},\boldsymbol{\zeta}^{\tau}\bigr)^{\tau}+O_p\bigl(\log^{1/2}(1/h)/\sqrt{nh}\bigr),
\end{aligned}
$$

where

$$
\begin{aligned}
S_n=\frac{1}{nh}\sum_{i=1}^{n}\biggl\{&\frac{1}{\psi(U_i)}K\left(\frac{U_i-u}{h}\right)f\bigl(-r_i(u)/\psi(U_i)\bigr)\\
&\times\bigl(\tilde{\boldsymbol{X}}_i^{\tau},\tilde{\boldsymbol{X}}_i^{\tau}(U_i-u)/h\bigr)^{\tau}\bigl(\tilde{\boldsymbol{X}}_i^{\tau},\tilde{\boldsymbol{X}}_i^{\tau}(U_i-u)/h\bigr)\biggr\}.
\end{aligned}
$$

It can be shown that

$$
E(S_n)=\frac{f_U(u)}{\psi(u)}S+O_p\bigl(h^2\bigr),
$$

where $S = \operatorname{diag}(f(0), \mu_2 f(0)) \otimes E(\tilde{\boldsymbol{X}}\tilde{\boldsymbol{X}}^\tau | U = u)$, $\mu_2 = \int u^2 K(u)\,du$. Then

$$\begin{aligned}
&\boldsymbol{W}_n^\tau(\boldsymbol{\eta}^\tau, \boldsymbol{\zeta}^\tau)^\tau + B_n \\
&\quad = \boldsymbol{W}_n^\tau(\boldsymbol{\eta}^\tau, \boldsymbol{\zeta}^\tau)^\tau + E(B_n) + O_p\bigl(\log^{1/2}(1/h)/\sqrt{nh}\bigr) \\
&\quad = \boldsymbol{W}_n^\tau(\boldsymbol{\eta}^\tau, \boldsymbol{\zeta}^\tau)^\tau + \frac{f_U(u)}{2\psi(u)}(\boldsymbol{\eta}^\tau, \boldsymbol{\zeta}^\tau) S (\boldsymbol{\eta}^\tau, \boldsymbol{\zeta}^\tau)^\tau + O_p\bigl(h^2 + \log^{1/2}(1/h)/\sqrt{nh}\bigr).
\end{aligned}$$

Similar to the proof procedures of Theorem 3.1 of Kai et al. [10], by applying the convexity lemma of Pollard [17] and the quadratic approximation lemma of Fan et al. [4], the minimizer can be expressed as

$$(\hat{\boldsymbol{\eta}}^\tau, \hat{\boldsymbol{\zeta}}^\tau)^\tau = -\frac{\psi(u)}{2 f_U(u)} S^{-1} \boldsymbol{W}_n + O_p\bigl(h^2 + \log^{1/2}(1/h)/\sqrt{nh}\bigr). \tag{A.3}$$

That is,

$$\begin{aligned}
\hat{\boldsymbol{\eta}} = {} & -\frac{\psi(u)}{f_U(u) f(0)} \bigl(E(\tilde{\boldsymbol{X}}\tilde{\boldsymbol{X}}^\tau | U = u)\bigr)^{-1} \sum_{i=1}^n \biggl\{ \frac{1}{\sqrt{nh}} K\bigl((U_i - u)/h\bigr) \\
& \times \bigl[I\bigl(e_i \le -r_i(u)/\psi(U_i)\bigr) - 1/2\bigr] \tilde{\boldsymbol{X}}_i \biggr\} + O_p\bigl(h^2 + \log^{1/2}(1/h)/\sqrt{nh}\bigr).
\end{aligned}$$

Hence, we get

$$\begin{aligned}
\hat{\boldsymbol{\beta}}(u) - \boldsymbol{\beta}(u) = {} & -\frac{\psi(u)}{f_U(u) f(0)} \bigl(E(\tilde{\boldsymbol{X}}\tilde{\boldsymbol{X}}^\tau | U = u)\bigr)^{-1} \sum_{i=1}^n \biggl\{ \frac{1}{nh} K\bigl((U_i - u)/h\bigr) \\
& \times \bigl[I\bigl(e_i \le -r_i(u)/\psi(U_i)\bigr) - 1/2\bigr] \tilde{\boldsymbol{X}}_i \biggr\} + \frac{1}{\sqrt{nh}} O_p\bigl(h^2 + \log^{1/2}(1/h)/\sqrt{nh}\bigr).
\end{aligned}$$

Obviously, the asymptotic expression of $\boldsymbol{\beta}(U_k)$ is

$$\begin{aligned}
\hat{\boldsymbol{\beta}}(U_k) = {} & \boldsymbol{\beta}(U_k) - \frac{\psi(U_k)}{f_U(U_k) f(0)} \bigl(E(\tilde{\boldsymbol{X}}_k \tilde{\boldsymbol{X}}_k^\tau | U_k)\bigr)^{-1} \sum_{i=1}^n \biggl\{ \frac{1}{nh} K((U_i - U_k)/h \\
& \times \bigl[I\bigl(e_i \le -r_i(U_k)/\psi(U_i)\bigr) - 1/2\bigr] \tilde{\boldsymbol{X}}_i \biggr\} + \frac{1}{\sqrt{nh}} O_p\bigl(h^2 + \log^{1/2}(1/h)/\sqrt{nh}\bigr)
\end{aligned}$$

for $k = 1, \ldots, n$.

We split the second term in the previous expression into two parts $R_{1k} + R_{2k}$, where

$$\begin{aligned}
R_{1k} = {} & -\frac{\psi(U_k)}{f_U(U_k) f(0)} \bigl(E(\tilde{\boldsymbol{X}}_k \tilde{\boldsymbol{X}}_k^\tau | U_k)\bigr)^{-1} \sum_{i=1}^n \biggl\{ \frac{1}{nh} K\bigl((U_i - U_k)/h\bigr) \\
& \times \biggl[I\biggl(e_i \le -\frac{r_i(U_k)}{\psi(U_i)}\biggr) - I(e_i \le 0) \biggr] \tilde{\boldsymbol{X}}_i \biggr\}
\end{aligned}$$

and

$$R_{2k} = -\frac{\psi(U_k)}{f_U(U_k)f(0)}\big(E(\tilde{\boldsymbol{X}}_k\tilde{\boldsymbol{X}}_k^\tau|U_k)\big)^{-1}\sum_{i=1}^n\left\{\frac{1}{nh}K\big((U_i-U_k)/h\big)\right.$$
$$\left.\times\left[I(e_i\le 0)-\frac{1}{2}\right]\tilde{\boldsymbol{X}}_i\right\}.$$

Write

$$E\big(\|R_{1k}\|^2\big) = \frac{1}{n^2h^2}\sum_{i=1}^n E(R_{ii,1k}) + \frac{2}{n^2h^2}\sum_{i\neq j}E(R_{ij,1k}),$$

where

$$R_{ii,1k} = \frac{\psi^2(U_k)}{f_U^2(U_k)f^2(0)}\tilde{\boldsymbol{X}}_i^\tau\big(E(\tilde{\boldsymbol{X}}_k\tilde{\boldsymbol{X}}_k^\tau|U_k)\big)^{-2}\tilde{\boldsymbol{X}}_i$$
$$\times\left[I\left(e_i\le-\frac{r_i(U_k)}{\psi(U_i)}\right)-I(e_i\le 0)\right]^2K^2\big((U_i-U_k)/h\big),$$
$$R_{ij,1k} = \frac{\psi^2(U_k)}{f_U^2(U_k)f^2(0)}\tilde{\boldsymbol{X}}_i^\tau\big(E(\tilde{\boldsymbol{X}}_k\tilde{\boldsymbol{X}}_k^\tau|U_k)\big)^{-2}\tilde{\boldsymbol{X}}_jK\big((U_i-U_k)/h\big)K\big((U_j-U_k)/h\big)$$
$$\times\big[I\big(e_i\le-r_i(U_k)/\psi(U_i)\big)-I(e_i\le 0)\big]\big[I\big(e_j\le-r_j(U_k)/\psi(U_j)\big)-I(e_j\le 0)\big].$$

By the fact

$$\big[I\big(e_i\le-r_i(U_k)/\psi(U_i)\big)-I(e_i\le 0)\big]^2 = \big|I\big(e_i\le-r_i(U_k)/\psi(U_i)\big)-I(e_i\le 0)\big|,$$

without loss of generality, assuming $-\frac{r_i(U_k)}{\psi(U_i)}>0$, we have

$$\begin{aligned}
&E(R_{ii,1k})\\
&\quad= E\big\{E(R_{ii,1k})|U_k,U_i\big\}\\
&\quad= E\left\{\frac{\psi^2(U_k)}{f_U^2(U_k)f^2(0)}E\big(\tilde{\boldsymbol{X}}_i^\tau\big[E\big(\tilde{\boldsymbol{X}}_k\tilde{\boldsymbol{X}}_k^\tau|U_k\big)\big]^{-2}\tilde{\boldsymbol{X}}_i|U_i\big)\right.\\
&\qquad\left.\times\big[F\big(-r_i(U_k)/\psi(U_i)\big)-F(0)\big]K^2\big((U_i-U_k)/h\big)\right\}\\
&\quad= E\left\{\frac{\psi^2(U_k)}{f_U^2(U_k)f^2(0)\psi(U_i)}E\big(\tilde{\boldsymbol{X}}_i^\tau\big[E\big(\tilde{\boldsymbol{X}}_k\tilde{\boldsymbol{X}}_k^\tau|U_k\big)\big]^{-2}\tilde{\boldsymbol{X}}_i|U_i\big)\right.\\
&\qquad\left.\times f(\xi)\big(-r_i(U_k)\big)K^2\big((U_i-U_k)/h\big)\right\}\\
&\quad\le M\int\int\frac{\psi^2(U_k)f_U(U_i)}{f_U(U_k)\psi(U_i)}E\big(\tilde{\boldsymbol{X}}_i^\tau\big[E\big(\tilde{\boldsymbol{X}}_k\tilde{\boldsymbol{X}}_k^\tau|U_k\big)\big]^{-2}\tilde{\boldsymbol{X}}_i|U_i\big)\\
&\qquad\times(U_i-U_k)^2K^2\big((U_i-U_k)/h\big)\,dU_k\,dU_i,
\end{aligned}$$

where ξ between 0 and $-\frac{r_i(U_k)}{\psi(U_i)}$. Noting that $K(\cdot)$ is a symmetric function, we have $E(R_{ii,1k}) = O(h^3)$ uniformly for k. In the same spirit, we can prove $E(R_{ij,1k}) = O(h^6)$ uniformly for k. It follows that

$$E\big(\|R_{1k}\|\big)^2 = \frac{1}{n^2h^2}nO\big(h^3\big) + \frac{2}{n^2h^2}n(n-1)O\big(h^6\big) = O\big(h^4\big)$$

uniformly for k.

For R_{2k}, noting that

$$E\left\{\sum_{i=1}^{n} K\big((U_i - U_k)/h\big)\big[I(e_i \le 0) - 1/2\big]\tilde{X}_{ir}\right\}^2 = O(nh),$$

we have $R_{2k} = O_p(\frac{1}{\sqrt{nh}})$.

Therefore, we can obtain (A.1). □

Lemma 3 *Under assumptions* (C1)–(C6), *if* $nh^2/\log(1/h) \to \infty$ *and* $nh^4 \to 0$ *as* $n \to \infty$, *we have*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{V}_i(\gamma_r) \xrightarrow{D} N\big(0, A(\gamma_r)\big), \tag{A.4}$$

where $A(\gamma_r) = \gamma_r^2 E(\psi(U_i) - \phi_r(U_i))^2 E(X_{ir}^2)$.

Proof We use some elementary calculation to obtain

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{V}_i(\gamma_r) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\big(\beta_r(U_i) - \gamma_r\big)\tilde{X}_{ir} + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\big(\hat{\beta}_r(U_i) - \beta_r(U_i)\big)\tilde{X}_{ir}. \tag{A.5}$$

By central theorems for the sum of independent and identically distributed random variables, we can obtain that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\big(\beta_r(U_i) - \gamma_r\big)\tilde{X}_{ir} \xrightarrow{D} N\big(0, A(\gamma_r)\big). \tag{A.6}$$

By Lemma 2, we can show that $n^{-1/2}\sum_{i=1}^{n}(\hat{\beta}_r(U_i) - \beta_r(U_i))\tilde{X}_{ir} \xrightarrow{P} 0$. This together with (A.5) and (A.6) proves Lemma 3. □

Lemma 4 *Under conditions of Lemma* 2, *we have*

$$\frac{1}{n}\sum_{i=1}^{n}\hat{V}_i^2(\gamma_r) \xrightarrow{P} A(\gamma_r). \tag{A.7}$$

Proof

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\hat{V}_i^2(\gamma_r) &= \frac{1}{n}\sum_{i=1}^{n}\big(\beta_r(U_i) - \gamma_r\big)^2\tilde{X}_{ir}^2 + \frac{1}{n}\sum_{i=1}^{n}\big(\hat{\beta}_r(U_i) - \beta_r(U_i)\big)^2\tilde{X}_{ir}^2 \\
&\quad + \frac{2}{n}\sum_{i=1}^{n}\big(\hat{\beta}_r(U_i) - \beta_r(U_i)\big)\big(\beta_r(U_i) - \gamma_r\big)\tilde{X}_{ir}^2 \\
&=: M_1 + M_2 + M_3.
\end{aligned}$$

By the law of large numbers, we obtain $M_1 \xrightarrow{P} A(\gamma_r)$. Hence, to prove Lemma 4, we only need to show that $M_l \xrightarrow{P} 0$, $l = 2, 3$.

By conditions (C1)–(C3) and Lemma 2, we obtain

$$|M_2| \leq \frac{1}{n} \max_{1\leq i\leq n} \bigl(\hat{\beta}_r(U_i) - \beta_r(U_i)\bigr)^2 \sum_{i=1}^{n} \tilde{X}_{ir}^2 = o_p(1).$$

By the similar argument for M_3, we have

$$|M_3| \leq \frac{2}{n} \max_{1\leq i\leq n} \bigl|\hat{\beta}_r(U_i) - \beta_r(U_i)\bigr| \sum_{i=1}^{n} \bigl(\beta_r(U_i) - \gamma_r\bigr)\tilde{X}_{ir}^2 = o_p(1).$$

The proof is completed. □

Lemma 5 *Under the assumptions of Theorem 3, we have*

$$\max_{1\leq i\leq n} \bigl|\hat{V}_i(\gamma_r)\bigr| = o_p\bigl(n^{1/2}\bigr).$$

Proof Some elementary calculation yields that

$$\max_{1\leq i\leq n} \bigl|\hat{V}_i(\gamma_r)\bigr| \leq \max_{1\leq i\leq n} \bigl|\bigl(\beta_r(U_i) - \gamma_r\bigr)\tilde{X}_{ir}\bigr| + \max_{1\leq i\leq n} \bigl|\bigl(\hat{\beta}_r(U_i) - \beta_r(U_i)\bigr)\tilde{X}_{ir}\bigr|.$$

By conditions (C1) and (C2), we have $\frac{1}{n}\sum_{i=1}^{n}(\beta_r(U_i) - \gamma_r)^2\tilde{X}_{ir}^2 \overset{a.s.}{=} A(\gamma_r) < \infty$. This implies that $\max_{1\leq i\leq n} |(\beta_r(U_i) - \gamma_r)\tilde{X}_{ir}| = o_p(n^{1/2})$. By Markov's inequality, for any $\kappa > 0$,

$$\begin{aligned} P\Bigl\{n^{1/2} \max_{1\leq i\leq n} \bigl|\bigl(\hat{\beta}_r(U_i) - \beta_r(U_i)\bigr)\tilde{X}_{ir}\bigr| > \kappa\Bigr\} &\leq \sum_{i=1}^{n} P\bigl\{\bigl|\hat{\beta}_r(U_i) - \beta_r(U_i))\tilde{X}_{ir}\bigr| > \kappa\sqrt{n}\bigr\} \\ &\leq \frac{1}{n\kappa^2} \sum_{i=1}^{n} E\bigl\{\bigl[\hat{\beta}_r(U_i) - \beta_r(U_i))\bigr]^2\tilde{X}_{ir}^2\bigr\} \to 0. \end{aligned}$$

This is $\max_{1\leq i\leq n} |(\hat{\beta}_r(U_i) - \beta_r(U_i))\tilde{X}_{ir}| = o_p(n^{1/2})$. The proof is completed. □

Lemma 6 *Under the conditions of Theorem 3, we have*

$$\lambda = O_p\bigl(n^{-1/2}\bigr).$$

Proof By using Lemmas 3–5 and the same method in Owen [14], we can prove this lemma. Here, we omit the process of the proof. □

Proof of Theorem 1 The proposed estimates $\hat{\gamma}_r$ can be denoted by

$$\begin{aligned} \hat{\gamma}_r &= \frac{1}{\bar{\tilde{X}}_r} \sum_{i=1}^{n} \frac{1}{n} \hat{\beta}_r(U_i)\tilde{X}_{ir} \\ &= \frac{1}{\bar{\tilde{X}}_r} \frac{1}{n} \sum_{i=1}^{n} \bigl[\hat{\beta}_r(U_i) - \beta_r(U_i)\bigr]\tilde{X}_{ir} + \frac{1}{\bar{\tilde{X}}_r} \frac{1}{n} \sum_{i=1}^{n} \beta_r(U_i)\tilde{X}_{ir} \\ &= Q_1 + Q_2. \end{aligned}$$

Applying Lemma 2, we have

$$Q_1 = O_p(C_n)\frac{1}{\bar{\tilde{X}}_r}\frac{1}{n}\sum_{i=1}^{n}\tilde{X}_{ir} = O_p(C_n),$$

$$\begin{aligned} Q_2 &= \frac{1}{\bar{\tilde{X}}_r}\frac{1}{n}\sum_{i=1}^{n}\left[\gamma_r\frac{\psi(U_i)}{\phi_r(U_i)}\phi_r(U_i)\tilde{X}_{ir}\right] \\ &= \gamma_r\frac{1}{\bar{\tilde{X}}_r}\frac{1}{n}\sum_{i=1}^{n}\psi_r(U_i)X_{ir}. \end{aligned}$$

It is obvious that the following result holds: $\bar{\tilde{X}}_r = E(X_r) + O_p(n^{-1/2})$. Thereafter, we get $Q_2 = \gamma_r + O_p(n^{-1/2})$. The law of large numbers is used in the previous equations, and Theorem 1 holds. □

Proof of Theorem 2 Motivated by methodology in Sentürk and Müller [19], we show

$$\sqrt{n}\begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n}\hat{\beta}_r(U_i)\tilde{X}_{ir} - \gamma_r E(X_r) \\ \bar{\tilde{X}}_r - E(X_r)\end{pmatrix} \xrightarrow{D} N\left(\mathbf{0}, \begin{pmatrix} \Sigma_{r,11}, & \Sigma_{r,12} \\ \Sigma_{r,21}, & \Sigma_{r,22}\end{pmatrix}\right), \tag{A.8}$$

where

$$\begin{aligned} &\Sigma_{r,11} = \gamma_r^2 E(X_r^2)\operatorname{var}(\psi(U)) + \gamma_r^2\operatorname{var}(X_r), \\ &\Sigma_{r,12} = \Sigma_{r,21} = \gamma_r\left[E(\phi_r(U)\psi(U))E(X_r^2) - (E(X_r))^2\right], \qquad \Sigma_{r,22} = \operatorname{var}(\phi_r(U)X_r). \end{aligned}$$

The asymptotic normality of $\sqrt{n}(\hat{\gamma}_r - \gamma_r)$ for $r = 0, \dots, n$ will follow (A.8) with a simple application of the δ-method, since $\hat{\gamma}_r = \sum_{i=1}^{n} n^{-1}\hat{\beta}_r(U_i)\tilde{X}_{ir}/\bar{\tilde{X}}_r$ as defined in (2.4). In view of the Cramér–Wald device, we need only verify that, for any real a, b,

$$\sqrt{n}\left\{a\left[\sum_{i=1}^{n}(\hat{\beta}_r(U_i)\tilde{X}_{ir})/n - \gamma_r E(X_r)\right] + b\left[\bar{\tilde{X}}_r - E(X_r)\right]\right\} \xrightarrow{D} N(0, \sigma_r^{*2}), \tag{A.9}$$

where

$$\begin{aligned} \sigma_r^{*2} = {} & a^2\gamma_r^2 E(X_r^2)\operatorname{var}(\psi(U)) + a^2\gamma_r^2\operatorname{var}(X_r) \\ & + 2ab\gamma_r\left[E(\phi_r(U)\psi(U))E(X_r^2) - (E(X_r))^2\right] + b^2\operatorname{var}(\phi_r(U)X_r). \end{aligned}$$

Write

$$\begin{aligned} &\sqrt{n}\left\{a\left[\sum_{i=1}^{n}(\hat{\beta}_r(U_i)\tilde{X}_{ir})/n - \gamma_r E(X_r)\right] + b\left[\bar{\tilde{X}}_r - E(X_r)\right]\right\} \\ &\quad = \sqrt{n}\left\{\frac{a}{n}\sum_{i=1}^{n}\left[(\hat{\beta}_r(U_i) - \beta_r(U_i))\tilde{X}_{ir}\right] + \frac{a}{n}\sum_{i=1}^{n}\left[\beta_r(U_i)\tilde{X}_{ir}\right] - a\gamma_r E(X_r) \right. \\ &\qquad \left. + \frac{b}{n}\sum_{i=1}^{n}\tilde{X}_{ir} - bE(X_r)\right\} \end{aligned}$$

$$
\begin{aligned}
&= \sqrt{n}\left\{\frac{a}{n}\sum_{i=1}^{n}\left[\left(\hat{\beta}_r(U_i)-\beta_r(U_i)\right)\tilde{X}_{ir}\right]\right\} + \sqrt{n}\left\{\frac{a}{n}\sum_{i=1}^{n}\left[\beta_r(U_i)\tilde{X}_{ir}\right] - a\gamma_r E(X_r)\right. \\
&\quad \left. + \frac{b}{n}\sum_{i=1}^{n}\tilde{X}_{ir} - bE(X_r)\right\} \\
&= I_1 + I_2.
\end{aligned}
$$

For I_1, using Lemma 2 and the conditions in (C1)–(C6), we have

$$
\begin{aligned}
&\sqrt{n}\sum_{i=1}^{n}\frac{1}{n}\left(\left(\hat{\beta}_r(U_i)-\beta_r(U_i)\right)\tilde{X}_{ir}\right) \\
&= \sqrt{n}\sum_{i=1}^{n}\frac{1}{n}\left\{-\frac{\psi(U_i)}{f_U(U_i)f(0)}\sum_{j=1}^{n}\left(\frac{1}{nh}K\left((U_j-U_i)/h\right)\right]\left[I(e_j\le 0)-\frac{1}{2}\right]\right. \\
&\quad \left.\times\left(E\left(\tilde{\boldsymbol{X}}_i\tilde{\boldsymbol{X}}_i^{\tau}|U_i\right)^{-1}\tilde{\boldsymbol{X}}_j\right)_r\right) + \frac{1}{\sqrt{nh}}O_p\left(h^2+\log^{1/2}(1/h)/\sqrt{nh}\right)\right\}\tilde{X}_{ir} \\
&= \sqrt{n}\sum_{i=1}^{n}\frac{1}{n}\left\{-\frac{\psi(U_i)}{f_U(U_i)f(0)}\sum_{j=1}^{n}\left(\frac{1}{nh}K\left((U_j-U_i)/h\right)\right]\left[I(e_j\le 0)-\frac{1}{2}\right]\right. \\
&\quad \left.\times\left(E\left(\tilde{\boldsymbol{X}}_i\tilde{\boldsymbol{X}}_i^{\tau}|U_i\right)^{-1}\tilde{\boldsymbol{X}}_j\right)_r\right)\right\}\tilde{X}_{ir} + o_p(1) \\
&= G_n + o_p(1).
\end{aligned}
$$

Now, let us deal with G_n. Notice that $E(I(e_j \le 0) - 1/2) = 0$. Hence, we get

$$
\begin{aligned}
E\left(G_n^2\right) &\le \frac{C}{n}\sum_{j=1}^{n}E\left\{\sum_{i=1}^{n}\frac{1}{nh}K\left((U_j-U_i)/h\right)\left[I(e_j\le 0)-\frac{1}{2}\right]\tilde{X}_{ir}\right)\right\}^2 \\
&= \frac{1}{(nh)^2}O(nh) = o(1).
\end{aligned}
$$

This implies $I_1 = o_p(1)$. By the central limit theorem,

$$
I_2 \xrightarrow{D} N\left(0, \sigma_r^{*2}\right).
$$

This completes the proof of Theorem 2. □

Proof of Theorem 3 We use a Taylor expansion to (3.1), and by Lemmas 3–6 we can obtain

$$
\hat{L}_n(\gamma_r) = 2\sum_{i=1}^{n}\left\{\lambda\hat{V}_i(\gamma_r) - \left[\lambda\hat{V}_i(\gamma_r)\right]^2/2\right\} + o_p(1). \tag{A.10}
$$

By Eq. (3.2), we have

$$
\begin{aligned}
0 &= \frac{1}{n}\sum_{i=1}^{n}\frac{\hat{V}_i(\gamma_r)}{1+\lambda\hat{V}_i(\gamma_r)} \\
&= \frac{1}{n}\sum_{i=1}^{n}\hat{V}_i(\gamma_r) - \frac{1}{n}\sum_{i=1}^{n}\hat{V}_i^2(\gamma_r)\lambda + \frac{1}{n}\sum_{i=1}^{n}\frac{\hat{V}_i^3(\gamma_r)\lambda^2}{1+\lambda\hat{V}_i(\gamma_r)}.
\end{aligned} \tag{A.11}
$$

By Lemmas 3–6, the final term of (A.11) has norm bounded by

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\left|\hat{V}_i^3(\gamma_r)\right|\lambda^2\left|1+\lambda\hat{V}_i(\gamma_r)\right|^{-1} &\le O_p\left(n^{-1}\right)\max_{1\le i\le n}\left|\hat{V}_i(\gamma_r)\right|\frac{1}{n}\sum_{i=1}^{n}\left|\hat{V}_i(\gamma_r)\right|^2\\ &= O_p\left(n^{-1}\right)o_p\left(n^{1/2}\right)O_p(1)\\ &= o_p\left(n^{-1/2}\right). \end{aligned}\tag{A.12}$$

This, together with (A.11), yields

$$\sum_{i=1}^{n}\left[\hat{V}_i(\gamma_r)\lambda\right]^2=\sum_{i=1}^{n}\hat{V}_i(\gamma_r)\lambda+o_p(1),$$

$$\lambda=\left[\sum_{i=1}^{n}\hat{V}_i^2(\gamma_r)\right]^{-1}\sum_{i=1}^{n}\hat{V}_i(\gamma_r)+o_p\left(n^{-1/2}\right).$$

Then, by (A.10), we have

$$\hat{L}_n(\gamma_r)=\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{V}_i(\gamma_r)\right]^2\left[\frac{1}{n}\sum_{i=1}^{n}\hat{V}_i^2(\gamma_r)\right]^{-1}+o_p(1).$$

This, together with Lemmas 3 and 4, completes the proof. □

Acknowledgements
We sincerely thank referees and editor for their helpful comments and suggestions, which have improved this version of the manuscript.

Funding
This work is supported by the National Natural Science Foundation of China (NO. 1187028, 11731015, 11571051, 11501241), Natural Science Foundation of Jilin Province (No. 20180101216JC, 20170101057JC, 20150520053JH), and Program for Changbaishan Scholars of Jilin Province (2015010).

Availability of data and materials
Pima Indian diabetes data set is used for empirical study. This dataset is originally from the National Institute of Diabetes and Digestive and Kidney Diseases. The data set is available on the web site: https://www.kaggle.com/uciml/pima-indians-diabetes-database.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
YS and DW conceived the idea of the study; YS analyzed the data; DW interpreted the results; YS wrote the paper; all the authors discussed the results and revised the manuscript. All authors read and approved the final manuscript.

Author details
[1]Institute of Mathematics, Jilin University, Changchun, P.R. China. [2]College of Mathematics, Jilin Normal University, Siping, P.R. China.

Publisher's Note

Received: 15 May 2019 Accepted: 26 February 2020 Published online: 19 March 2020

References

1. Cai, Z., Fan, J., Yao, Q.: Functional-coefficient regression models for nonlinear time series. J. Am. Stat. Assoc. **95**, 941–956 (2000)
2. Carroll, R., Fan, J., Gijbels, I., Wand, M.: Generalized partially linear single-index models. J. Am. Stat. Assoc. **92**, 477–489 (1997)

3. Cui, X., Guo, W.S., Zhu, L.X.: Covariate-adjusted nonlinear regression. Ann. Stat. **37**, 1839–1870 (2009)
4. Fan, J., Gijbels, I.: Local Polynomial Modelling and Its Applications. Chapman & Hall, London (1996)
5. Fan, J., Zhang, W.: Statistical estimation in varying coefficient model. Ann. Stat. **27**, 1491–1518 (1999)
6. Fan, J., Zhang, W.: Simultaneous confidence bands and hypothesis testing in varying-coefficient models. Scand. J. Stat. **27**, 715–731 (2000)
7. Fan, J., Zhang, W.: Statistical methods with varying coefficient models. Stat. Interface **1**, 179–195 (2008)
8. Hastie, T.J., Tibshirani, R.J.: Varying-coefficient models. J. R. Stat. Soc., Ser. B **55**, 757–796 (1993)
9. Kai, B., Li, R., Zou, H.: Local composite quantile regression smoothing: an efficient and safe alternative to local polynomial regression. J. R. Stat. Soc., Ser. B **72**, 49–69 (2010)
10. Kai, B., Li, R., Zou, H.: New efficient estimation and variable selection methods for semiparametric varying-coefficient partially linear models. Ann. Stat. **39**, 305–332 (2011)
11. Kaysen, G.A., Dubin, J.A., Müller, H.G., Mitch, W.E., Rosales, L.M., Levin, N.W., the Hemo Study Group: Relationships among inflammation nutrition and physiologic mechanisms establishing albumin levels in hemodialysis patients. Kidney Int. **61**, 2240–2249 (2002)
12. Li, F., Lin, L., Cui, X.: Covariate-adjusted partially linear regression models. Commun. Stat., Theory Methods **39**, 1054–1074 (2010)
13. Mack, Y., Silverman, B.: Weak and strong uniform consistency of kernel regression estimates. Probab. Theory Relat. Fields **61**, 405–415 (1982)
14. Owen, A.B.: Empirical likelihood ratio confidence regions. Ann. Stat. **18**, 90–120 (1990)
15. Owen, A.B.: Empirical likelihood for linear models. Ann. Stat. **19**, 1725–1747 (1991)
16. Owen, A.B.: Empirical Likelihood. Chapman & Hall, New York (2001)
17. Pollard, D.: Asymptotics for least absolute deviation regression estimators. Econom. Theory **7**, 186–199 (1991)
18. Sentürk, D., Müller, H.G.: Covariate-adjusted regression. Biometrika **92**, 75–89 (2005)
19. Sentürk, D., Müller, H.G.: Inference for covariate-adjusted regression via varying coefficient models. Ann. Stat. **34**, 654–679 (2006)
20. Senturk, D., Nguyen, D.V.: Estimation in covariate-adjusted regression. Comput. Stat. Data Anal. **50**, 3294–3310 (2006)
21. Tang, Q.G., Wang, J.D.: L_1-Estimation for varying coefficient models. Statistics **39**, 389–404 (2005)