Chin. Ann. Math. Ser. B
47(3), 2026, 449–474
DOI: 10.1007/s11401-026-0032-9

Chinese Annals of Mathematics, Series B

Jump-Preserving Estimation for the Discontinuous Link Function in a Single-Index Multiplicative Model*

Yinjun CHEN[1] Hu YANG[2]

Abstract This study explores a broader variety of single-index multiplicative models (SIMM for short) with an unknown, discontinuous link function. Relaxing the continuity assumption in nonparametric functions enhances the applicability to positive data. However, the authors find an issue with the existing least product relative error (LPRE for short) technique at jump points, posing a challenge in estimating the link function accurately in SIMMs. They propose an automated method that combines the LPRE technique with jump-preserving methods to simultaneously estimate the unknown parameter vector and the discontinuous link function. Their approach is flexible and practical, not requiring prior knowledge of jump point details. Furthermore, they establish the asymptotic properties of the estimators for the parametric vector and the discontinuous function components under reasonable conditions. To validate their approach, they conduct numerical simulations evaluating the performance with finite samples. Additionally, they demonstrate the effectiveness of the approach through real data analysis.

Keywords Discontinuous link function, Jump-Preserving techniques, Least product relative error, Single-index multiplicative model
2020 MR Subject Classification 62G05, 62G08, 62G20

1 Introduction

Nonparametric estimation of discontinuous functions has gained significant attention in recent decades and has found widespread application in various scientific disciplines. For instance, a jump discontinuity in the sea-level pressure was discovered in the early 1960s in oceanography, as documented in [19–20]. In finance, jumps are commonly observed in various financial time series such as stock price, gold price, exchange rate, and stock price index series, as reported in [11, 15]. Different approaches have been proposed to address the estimation of such functions. Some methods adopt a two-step process, first identifying the locations of the discontinuities and then estimating the regression function in separate intervals determined by the identified jumps (see [12, 21, 24]). Other methods propose jump-preserving estimation techniques that automatically account for potential jumps in the regression function without prior knowledge of the number and locations of the jump points, as seen in [7, 17, 19]. The above-mentioned

Manuscript received February 24, 2023. Revised March 14, 2024.
[1]College of Mathematics and Statistics, Chongqing University, Chongqing 401331, China.
E-mail: lcee@cqu.edu.cn
[2]Corresponding author. College of Mathematics and Statistics, Chongqing University, Chongqing 401331, China. E-mail: yh@cqu.edu.cn
*This work was supported by the National Natural Science Foundation of China (No. 12371281).

method primarily addresses the issue of discontinuous functions in univariate nonparametric functions. However, in some applications, it may be necessary to consider multiple predictors simultaneously to adequately capture complicated data structures. As noted by [13], it can be challenging to apply nonparametric jump regression analysis to situations with multiple variables for two reasons: First, the jump structures in spaces of different dimensions are substantially different, and second, most jump regression analysis methods rely on local regression, which is less effective in higher dimensions. As is well known, semiparametric models combine the advantages of parametric and nonparametric models, making them suitable for situations where the data distribution is uncertain. Additionally, they are capable of handling nonlinear relationships. Therefore, it is crucial to develop a semiparametric jump-preserving method that can account for multiple predictors.

Fortunately, there have been a few studies that have considered semiparametric models with discontinuous functions. For varying coefficient models with discontinuous coefficient functions, [27–28] proposed an adaptive jump-preserving (AJP for short) estimation procedure to estimate the coefficient functions with jumps. Then, [26] developed a nonparametric method to diagnose jump discontinuities based on Nadaraya-Watson estimation and least squares fitting. In the context of jump additive models, [13] proposed a jump-preserving backfitting algorithm that achieves the jump-preserving property by adaptively choosing at each design point in each iteration. For single-index models, they can be primarily categorized into two types of structural break based on the characteristics of their model structure. One type involves potential structural break in the index coefficients, often referred to as breakpoints. The other type pertains to discontinuities in the link function. In the case of single-index models with a structural break in the index coefficient, [1, 10] assumed a stable link function and proposed a statistic for testing the presence of a structural break and focused on estimating the location of the structural break, respectively. Meanwhile, for single-index models with discontinuous link function, [14] proposed a semiparametric jump-preserving estimation method and demonstrated that the resulting estimators could effectively preserve jumps for normal error distribution. Additionally, when there are outliers in the observations or the error distribution is heavy-tailed, [8–9] developed a modified EM algorithm to implement the adaptive semiparametric estimation and showed that the resulting estimators are robust and efficient for different error distributions. [18] investigated the test size and power of a cumulative-sum statistic based on weighted residuals for detecting changes in the link function. It is worth noting that nonparametric structural breaks pose greater challenges compared to parametric structural breaks, requiring more complex handling and analytical techniques. Additionally, we also observe that in domains where wide-ranging jumps exist, the collected sample data often have positive measurements, such as financial asset prices, patient lifetimes, and body fat indexes. Therefore, the multiplicative regression model is a natural choice for analyzing data with a positive response. Consider the following multiplicative regression model

$$Y_i = \exp(X_i^{\mathrm{T}}\beta_0)\varepsilon_i, \quad i = 1, \cdots, n, \tag{1.1}$$

where Y_i is the response variable, X_i is the p-vector of explanatory variables, β_0 is the corresponding p-vector of regression parameters to be estimated and ε_i is the error term with positive value. When a logarithmic transformation is applied to this model, it simplifies to a linear model, which allows for common estimation methods, such as least squares (LS for short) estimation and least absolute deviation (LAD for short) estimation, to be applied. However, it is important to note that a linear relationship in the transformed model is not necessarily linear in the original one, and this comes at the cost of losing an intuitive interpretation for the transformed models. To estimate the parameters in model (1.1), [2] proposed an alternative to traditional estimation methods called least absolute relative error (LARE for short), which minimizes

$$\sum_{i=1}^{n}\Big\{\frac{|Y_i - \exp(X_i^{\mathrm{T}}\beta)|}{Y_i} + \frac{|Y_i - \exp(X_i^{\mathrm{T}}\beta)|}{\exp(X_i^{\mathrm{T}}\beta)}\Big\}.$$

Compared to the commonly used mean absolute error, the suggested approach relies on two types of relative errors that are symmetric in the actual value and its predicted value. A direct calculation yields $\sum_{i=1}^{n}(|\varepsilon_i^{-1} - 1| + |\varepsilon_i - 1|)$, which was non-differentiable and more complicated than linear programming. Due to this, [3] proposed the least product relative error (LPRE for short) estimation by minimizing

$$\sum_{i=1}^{n}\Big\{\Big|\frac{Y_i - \exp(X_i^{\mathrm{T}}\beta)}{Y_i}\Big| \times \Big|\frac{Y_i - \exp(X_i^{\mathrm{T}}\beta)}{\exp(X_i^{\mathrm{T}}\beta)}\Big|\Big\},$$

which is equivalent to minimizing $\sum_{i=1}^{n}(\varepsilon_i^{-1} + \varepsilon_i)$. The LPRE objective function has the appealing property of being infinitely differentiable and strictly convex, which not only ensures the numerical simplicity of the approach, but also guarantees the uniqueness of the solution. Thanks to these advantages, [23] employed the LPRE objective function to propose a testing procedure for detecting and estimating change points in multiplicative regression models. As the objective phenomena under study become increasingly complex, there is a growing demand for greater precision. Simple linear structures are no longer sufficient for effectively analyzing and processing data. [16, 22, 25] extended the multiplicative regression model to a single-index multiplicative model: $Y_i = \exp(g(X_i^{\mathrm{T}}\beta_0))\varepsilon_i, i = 1, \cdots, n$, and studied the estimation method of the unknown index parameters β_0 and link function $g(\cdot)$. However, a common assumption in these studies is that the link function $g(\cdot)$ is smooth.

As mentioned earlier, discontinuous link functions are encountered in many fields. Ignoring the discontinuity of the link function may lead to inconsistent estimators at points where the link function has jump discontinuities. Therefore, it is necessary to develop consistent estimators that take the discontinuity of the link function into account. The primary aim of our paper is to investigate the introduction of an unknown discontinuous link function in the SIMM and study the impact of this discontinuity on estimations. Estimating the unknown link function becomes challenging due to the unknown number and locations of the jump points. In response to this challenge, we propose a semiparametric jump-preserving local LPRE (JPLLPRE for short) estimator for the SIMM. However, we also encountered another challenge in constructing appropriate measurement criteria for local linear estimators, as the loss function of our method differed from that of [9, 14]. To address this challenge, we introduced a novel criterion called the weighted relative error mean squares (WREMS for short) criterion, which combines certain properties of relative error. This criterion served as a valuable tool for evaluating the performance of our proposed estimator. Our main contribution lies in filling this research gap by introducing the JPLLPRE method, which effectively detects the existing jump in the link function of SIMM. Additionally, we provide empirical evidence of the consistency and asymptotic normality of the resulting estimators, even at the discontinuous points, under certain technical conditions.

The structure of this paper is as follows. Section 2 presents an iterative algorithm for estimating the parameter vector β_0 and the discontinuous link function $g(\cdot)$ in SIMM, based on the local LPRE method and the jump-preserving technique. We also discuss the parameter selection procedure. In Section 3, we provide the theoretical properties of the proposed estimators. In Section 4, we conduct numerical simulations. Section 5 provides an analysis of real data. Lastly, in Section 6, we summarize our findings and offer concluding remarks.

2 Estimation Procedures

The proposed estimation procedure is discussed in the following three subsections. In Subsection 2.1, we briefly introduce the jump-preserving method for discontinuous nonparametric regression models proposed by Gijbels, Lambert and Qiu [7]. In Subsection 2.2, we present the JPLLPRE procedure based on three local linear kernel estimates. In Subsection 2.3, we discuss the parameter selection procedure, which is also a key component of the estimation.

2.1 Preliminary knowledge on jump-preserving estimation

[7] considered the nonparametric regression model $y = g(u) + \varepsilon$ with $g(\cdot)$ the unknown regression function. Suppose that $g(u)$ has support in $\mathbb{U} = [a, b]$ and $g(\cdot)$ has jumps at points $\underline{u}_j$ with jump magnitudes d_j for $j = 1, \cdots, s$. Without loss of generality, we assume that $g(\cdot)$ is right-continuous at each jump location. Note that the number of jumps s, locations $\underline{u}_j$ of jump points and the magnitudes d_j are all unknown. However, we can automatically obtain this information from the definition of jump points. For a given point u_0, denote $g(u_0^-)$ and $g(u_0^+)$ as the left and right limit of $g(u_0)$, respectively. Furthermore, by utilizing corresponding single-side observations, the estimates for $g(u_0^-)$ and $g(u_0^+)$ can be derived. If $g(u_0^-) \neq g(u_0^+)$,

then u_0 is considered a jump point, with the corresponding magnitude being $g(u_0^+) - g(u_0^-)$.

Let K_c be a bounded symmetric density kernel function with support $[-\delta, \delta]$ and let $h > 0$ be the bandwidth parameter. Further assume the distance of two adjacent jumps must be greater than $2\delta h$. More precisely, we define two one-sided kernels from the kernel K_c by putting $K_l(u) = K_c(u)I\{u \in [-\delta, 0)\}$ and $K_r(u) = K_c(u)I\{u \in [0, \delta]\}$. We further separate the support $\mathbb{U}$ into two parts: (i) the continuous region D_1 and (ii) the discontinuity region D_2 containing jump points and their neighborhoods, that is, $D_2 = \bigcup_{j=1}^{s} (\underline{u}_j - \delta h, \underline{u}_j + \delta h)$; $D_1 = [a + \delta h, b - \delta h] \backslash D_2$. Then the region D_2 can be further separated into two parts: $D_{2,l} = \bigcup_{j=1}^{s} (\underline{u}_j - \delta h, \underline{u}_j)$ and $D_{2,r} = \bigcup_{j=1}^{s} [\underline{u}_j, \underline{u}_j + \delta h)$. The subregions $D_{2,l}$ and $D_{2,r}$ contain only the left and the right neighborhoods of the jump points respectively. It is obvious that the defined regions are mutually exclusive. Note that the definitions of $D_{2,l}$ and $D_{2,r}$ are correct when the function to be estimated is right continuous. If the function is left continuous, then the left region contains $\underline{u}_j$ instead.

Suppose that $\{(u_i, y_i), i = 1, \cdots, n\}$ is a sample from (u, y). By Taylor's expansion, the three local linear estimators at u can be obtained by

$$(\widehat{a}_v, \widehat{b}_v) = \arg\min_{a,b} \sum_{i=1}^{n} [y_i - a - b(u_i - u)]^2 K_{h,v}(u_i - u), \quad v = c, l, r, \tag{2.1}$$

where $K_{h,v}(\cdot) = \frac{K_v(\cdot/h)}{h}$. $(\widehat{a}_v, \widehat{b}_v)$ are called the left, right and centered local linear estimates for $(g(u), g'(u))$.

For the three estimates, the weighted residual mean square (WRMS for short) is defined as

$$\mathrm{WRMS}_v(u) = \frac{\sum_{i=1}^{n} [y_i - \widehat{a}_v - \widehat{b}_v(u_i - u)]^2 K_{h,v}(u_i - u)}{\sum_{i=1}^{n} K_{h,v}(u_i - u)}, \quad v = c, l, r.$$

Then an important diagnostic is based on the maximum of the pairwise differences, which is defined as

$$\mathrm{diff}(u) = \max(\mathrm{WRMS}_c(u) - \mathrm{WRMS}_l(u), \mathrm{WRMS}_c(u) - \mathrm{WRMS}_r(u)).$$

Clearly, the diagnostic quantity allows for making a data-driven choice between the three estimators is given by

$$\widehat{g}(u) = \begin{cases} \widehat{a}_c(u), & \text{if } \mathrm{diff}(u) \le \lambda, \\ \widehat{a}_l(u), & \text{if } \mathrm{diff}(u) > \lambda \text{ and } \mathrm{WRMS}_l(u) < \mathrm{WRMS}_r(u), \\ \widehat{a}_r(u), & \text{if } \mathrm{diff}(u) > \lambda \text{ and } \mathrm{WRMS}_l(u) > \mathrm{WRMS}_r(u), \\ \dfrac{\widehat{a}_l(u) + \widehat{a}_r(u)}{2}, & \text{if } \mathrm{diff}(u) > \lambda \text{ and } \mathrm{WRMS}_l(u) = \mathrm{WRMS}_r(u), \end{cases} \tag{2.2}$$

where $\lambda > 0$ is a threshold parameter, which plays the role of regulating sensitivity.

2.2 Jump-preserving local least product relative error estimation

This section presents the relative error estimation method for the single-index multiplicative model with jumps (SIMMJ for short), utilizing the jump-preserving estimation approach and the LPRE criterion. We study the SIMMJ of the following form:

$$Y = \exp(g(X^{\mathrm{T}}\beta_0))\varepsilon = \begin{cases} \exp(g_1(X^{\mathrm{T}}\beta_0))\varepsilon, & X^{\mathrm{T}}\beta_0 < \underline{u}_1, \\ \quad\vdots & \\ \exp(g_s(X^{\mathrm{T}}\beta_0))\varepsilon, & X^{\mathrm{T}}\beta_0 \geq \underline{u}_s, \end{cases} \tag{2.3}$$

where $g(\cdot)$ is a discontinuous unknown link function, $g_j(\cdot)$ is unspecified for $j = 1, \cdots, s$ and $\underline{u}_j$ is called a jump point. Y is the response variable, X is a p-dimensional vector, β_0 is an unknown index parameter and ε is a random error. Both Y and ε considered in model (2.3) are positive variables. The model error ε satisfies $E\{\varepsilon - \varepsilon^{-1}\} = 0$, which is used for the least product relative error estimation (see [3]). To guarantee the identification of the model, we require $\beta_0^{\mathrm{T}}\beta_0 = 1$ and the first element of β_0 is positive.

The local linear smoothing method with a symmetric kernel function is often employed to estimate the link function and index coefficient in model (2.3). From [16], the local LPRE method can estimate $g(\cdot)$ and β_0 simultaneously under the assumption that $g(\cdot)$ is smooth; otherwise, local LPRE will blur possible jumps. As mentioned above, the smoothing part and jumps of the discontinuous regression function can be estimated simultaneously by the jump-preserving local linear method. Therefore, by introducing a one-sided kernel function, we combine the local LPRE method to propose an adaptive semiparametric estimation for the SIMM.

The estimator of the model (2.3) can be obtained by solving the following optimization:

$$\begin{aligned} &\min_{\beta} \sum_{j=1}^{n} \min_{v=c,l,r} \Big\{ \min_{a_{v,j}, b_{v,j}} \sum_{i=1}^{n} [Y_i \exp(-a_{v,j} - b_{v,j}\mathbf{x}_{ij}^{\mathrm{T}}\beta) \\ &+ Y_i^{-1} \exp(a_{v,j} + b_{v,j}\mathbf{x}_{ij}^{\mathrm{T}}\beta)] K_{h,v}(\mathbf{x}_{ij}^{\mathrm{T}}\beta) \Big\}, \end{aligned} \tag{2.4}$$

where $\mathbf{x}_{ij} = X_i - X_j$. Since the objective function of (2.4) is a complex nonlinear function, directly solving such a problem is nontrivial. We decompose (2.4) into two parts.

P1 Given β,

$$\begin{aligned} &(\widehat{a}_{v,j}, \widehat{b}_{v,j})_{j=1}^{n} \\ = &\operatorname*{arg\,min}_{(a_{v,j}, b_{v,j})_{j=1}^{n}} \sum_{j=1}^{n}\sum_{i=1}^{n} [Y_i \exp(-a_{v,j} - b_{v,j}\mathbf{x}_{ij}^{\mathrm{T}}\beta) + Y_i^{-1} \exp(a_{v,j} + b_{v,j}\mathbf{x}_{ij}^{\mathrm{T}}\beta)] K_{h,v}(\mathbf{x}_{ij}^{\mathrm{T}}\beta). \end{aligned}$$

So for any $j \in \{1, 2, \cdots, n\}$,

$$(\widehat{a}_{v,j}, \widehat{b}_{v,j})$$

$$= \underset{(a_{v,j},b_{v,j})}{\arg\min} \sum_{i=1}^{n} [Y_i \exp(-a_{v,j} - b_{v,j}\mathbf{x}_{ij}^{\mathrm{T}}\beta) + Y_i^{-1}\exp(a_{v,j} + b_{v,j}\mathbf{x}_{ij}^{\mathrm{T}}\beta)] K_{h,v}(\mathbf{x}_{ij}^{\mathrm{T}}\beta). \tag{2.5}$$

P2 Given $\widehat{a}_j$ and $\widehat{b}_j$ by comparing the diagnostic quantity, then

$$\widehat{\beta} = \underset{\beta}{\arg\min} \sum_{j=1}^{n}\sum_{i=1}^{n} [Y_i \exp(-\widehat{a}_j - \widehat{b}_j\mathbf{x}_{ij}^{\mathrm{T}}\beta) + Y_i^{-1}\exp(\widehat{a}_j + \widehat{b}_j\mathbf{x}_{ij}^{\mathrm{T}}\beta)]. \tag{2.6}$$

The subproblem P1 deals with estimating $(a_{v,j}, b_{v,j})$, $j = 1, \cdots, n, v = c, l, r$, as if β is known. While in P2, β is estimated through the least product relative error criterion.

As a result, we have formulated an iterative JPLLPRE estimation algorithm. Initially, considering the β derived from (2.5), we compute three estimators, namely $(\widehat{a}_{c,j}, \widehat{b}_{c,j})$, $(\widehat{a}_{l,j}, \widehat{b}_{l,j})$ and $(\widehat{a}_{r,j}, \widehat{b}_{r,j})$, utilizing three distinct kernel functions for $j = 1, \cdots, n$. Subsequently, for each individual sample point, following a methodology akin to [16], we select one estimator from the three by comparing the differences in weighted relative error mean squares, and denote the resulting estimators as $(\widehat{a}_j, \widehat{b}_j)$. Finally, utilizing the chosen estimators, we recalibrate β by solving (2.6). The complete algorithmic outline is presented below.

Step 0 Initialize by obtaining initial $\beta_{(0)}$ through the least square estimation using the log-transformed dataset $\{X_i, \log(Y_i), i = 1, \cdots, n\}$.

Step 1 Given $\beta = \beta_{(k)}$, we can obtain $\{\widehat{a}_{v,j}^{(k)}, \widehat{b}_{v,j}^{(k)}\}_{j=1}^{n}$, $v = c, l, r$ by minimizing the following objective function

$$\sum_{i=1}^{n} [Y_i \exp(-a_{v,j} - b_{v,j}\mathbf{x}_{ij}^{\mathrm{T}}\beta_{(k)}) + Y_i^{-1}\exp(a_{v,j} + b_{v,j}\mathbf{x}_{ij}^{\mathrm{T}}\beta_{(k)})] K_{h,v}(\mathbf{x}_{ij}^{\mathrm{T}}\beta_{(k)}) \tag{2.7}$$

with respect to $(a_{v,j}, b_{v,j})$.

Step 2 Evaluate the quality of the three estimators using their weighted relative error mean squares, defined as

$$\mathrm{WREMS}_v(u_j) = \frac{\sum_{i=1}^{n} (\widehat{\varepsilon}_{i,v} - \widehat{\varepsilon}_{i,v}^{-1})^2 K_{h,v}(\mathbf{x}_{ij}^{\mathrm{T}}\beta_{(k)})}{\sum_{i=1}^{n} K_{h,v}(\mathbf{x}_{ij}^{\mathrm{T}}\beta_{(k)})},$$

where $\widehat{\varepsilon}_{i,v} = Y_i \exp(-\widehat{a}_{v,j}^{(k)} - \widehat{b}_{v,j}^{(k)}\mathbf{x}_{ij}^{\mathrm{T}}\beta_{(k)})$, $\widehat{\varepsilon}_{i,v}^{-1} = Y_i^{-1}\exp(\widehat{a}_{v,j}^{(k)} + \widehat{b}_{v,j}^{(k)}\mathbf{x}_{ij}^{\mathrm{T}}\beta_{(k)})$ and $u_j = X_j^{\mathrm{T}}\beta_{(k)}$. The diagnostic quantity is defined as

$$\mathrm{diff}(u_j) = \max(\mathrm{WREMS}_c(u_j) - \mathrm{WREMS}_l(u_j), \mathrm{WREMS}_c(u_j) - \mathrm{WREMS}_r(u_j)). \tag{2.8}$$

The jump-preserving local LPRE estimator is given by

$$\widehat{g}(u_j) = \begin{cases} \widehat{a}_{c,j}^{(k)}(u_j), & \text{if } \mathrm{diff}(u_j) \le \lambda, \\ \widehat{a}_{l,j}^{(k)}(u_j), & \text{if } \mathrm{diff}(u_j) > \lambda \text{ and } \mathrm{WREMS}_l(u_j) < \mathrm{WREMS}_r(u_j), \\ \widehat{a}_{r,j}^{(k)}(u_j), & \text{if } \mathrm{diff}(u_j) > \lambda \text{ and } \mathrm{WREMS}_l(u_j) > \mathrm{WREMS}_r(u_j), \\ \dfrac{\widehat{a}_{l,j}^{(k)}(u_j) + \widehat{a}_{r,j}^{(k)}(u_j)}{2}, & \text{if } \mathrm{diff}(u_j) > \lambda \text{ and } \mathrm{WREMS}_l(u_j) = \mathrm{WREMS}_r(u_j), \end{cases} \tag{2.9}$$

where λ is an appropriate threshold value. Represent $\widehat{g}(u_j)$ as $\widehat{a}_j^{(k)}$ and $\widehat{g}'(u_j)$ as $\widehat{b}_j^{(k)}$ with $\widehat{b}_j^{(k)}$ being the estimate corresponding to $\widehat{a}_j^{(k)}$.

Step 3 Given $\{\widehat{a}_j^{(k)}, \widehat{b}_j^{(k)}\}_{j=1}^n$, obtain $\widehat{\beta}$ by minimizing

$$\sum_{j=1}^n \sum_{i=1}^n [Y_i \exp(-\widehat{a}_j^{(k)} - \widehat{b}_j^{(k)} \beta^{\mathrm{T}}(X_i - X_j)) + Y_i^{-1} \exp(\widehat{a}_j^{(k)} + \widehat{b}_j^{(k)} \beta^{\mathrm{T}}(X_i - X_j))] \tag{2.10}$$

with respect to β. The estimator β_{k+1} of β is defined as a minimizer of (2.10).

Step 4 Standardize $\beta_{(k+1)} := \frac{\mathrm{sgn}(\beta_{(k+1),1})\beta_{(k+1)}}{\|\beta_{(k+1)}\|_2}$, where $\beta_{(k+1),1}$ is the first component of $\beta_{(k+1)}$ and $\beta_{(k+1),1} > 0$. Repeat Steps 1–3 until a tolerance error of 10^{-4} is achieved. Denote the final estimate of β by $\widehat{\beta}$.

Step 5 Using $\widehat{\beta}$ obtained from Step 4, compute the final estimate $\widehat{g}(\cdot)$ of $g(\cdot)$ by performing the previous Steps 1–2.

Remark 2.1 It is noted that $\sqrt{n}$-consistency of the initial estimator for β in Step 0 is required in the proposed method. The initial estimator methodology in Step 0 has been verified to be $\sqrt{n}$-consistent (see [4]).

2.3 Selection of parameters

In the construction of our estimator $\widehat{g}(\cdot)$, two parameters are needed to be chosen: The bandwidth h and the threshold λ. We choose h and λ by the leave-one-out least product cross-validation (LPCV for short) (see [16]). The LPCV criterion is given by

$$\begin{aligned}\mathrm{LPCV}(h, \lambda) = \frac{1}{n} \sum_{i=1}^n \Big\{ \Big| \frac{Y_i - \exp(\widehat{g}^{(-i)}(X_{-i}^{\mathrm{T}} \widehat{\beta}^{(-i)}; h, \lambda))}{Y_i} \Big| \\ \times \Big| \frac{Y_i - \exp(\widehat{g}^{(-i)}(X_{-i}^{\mathrm{T}} \widehat{\beta}^{(-i)}; h, \lambda))}{\exp(\widehat{g}^{(-i)}(X_{-i}^{\mathrm{T}} \widehat{\beta}^{(-i)}; h, \lambda))} \Big| \Big\},\end{aligned} \tag{2.11}$$

where $\widehat{g}^{(-i)}(\cdot)$ and $\widehat{\beta}^{(-i)}$ are respectively $\hat{g}(\cdot)$ and $\widehat{\beta}$ with the ith observation removed. The LPCV bandwidth and threshold are selected by minimizing (2.11) with respect to $h > 0$ and $\lambda > 0$. As in [5–6], a suitable range of threshold value λ can be defined by a maximum value $\lambda_{\max}$, e.g., defined as $\lambda_{\text{grid}} = (0.001, 0.01, 0.1, 1)\lambda_{\max}$ and an equispaced grid of bandwidth values, e.g., $h_{\text{grid}} = (0.01, 0.02, \cdots, 0.15)$ can be used. Here, we summarize the practical estimation algorithm, which iterates over the bandwidth parameter h and the threshold parameter λ in a nested procedure.

Consider a grid of bandwidths $h_{\text{grid}} := (h_1, \cdots, h_m)$.

Iterate over these bandwidths and put $h := h_q,\ q = 1, \cdots, m$.

For this bandwidth,

- Calculate the estimates $\widehat{a}_v(u_j)$ and $\widehat{b}_v(u_j)$ for $v = c, l, r,\ j = 1, \cdots, n$.
- Take $\lambda_{\max} := \max(|\widehat{a}_c(u_j) - \widehat{a}_r(u_j)|, |\widehat{a}_c(u_j) - \widehat{a}_l(u_j)|)$, $j = 1, \cdots, n$.
- Put $\lambda_{\text{grid}} := (0.001, 0.01, 0.1, 1)\lambda_{\max}$.

Now iterate over the threshold values and put $\lambda := \lambda_d,\ d = 1, 2, 3, 4$.

(1) For the combination of h and λ, calculate $\widehat{g}^{(-i)}(X^{\mathrm{T}}_{-i}\widehat{\beta}; h, \lambda)$ as in (2.9), but removing the ith observation.

(2) Compute $\frac{1}{n}\sum\limits_{i=1}^{n}\{Y_i \exp[-\widehat{g}^{(-i)}(X^{\mathrm{T}}_{-i}\widehat{\beta}; h, \lambda)] + Y_i^{-1}\exp[\widehat{g}^{(-i)}(X^{\mathrm{T}}_{-i}\widehat{\beta}; h, \lambda)] - 2\}$, which is equivalent to (2.11).

• Retain the value of λ that yields the minimum for the sum in the former step and associate with h_q by putting $\widetilde{\lambda}_q$.

Repeat the above procedure for each bandwidth and look for the bandwidth h_q (and associated threshold $\widetilde{\lambda}_q$) that yields the lowest value for the sum.

Calculate the final estimate with (2.9) from the couple (h, λ) obtained as above.

3 Asymptotic Properties

The link function in a single-index model is typically assumed to be continuous. However, in this paper, we consider a more general link function that allows for possible jumps. This includes many nonparametric and single-index models with continuous regression functions as exceptional cases. In order to establish asymptotic normality for the proposed estimates, we require the following technical conditions. While some of these conditions may not be the weakest possible, they are convenient for our proofs.

(C1) The density function $f(u)$ of $U = X^{\mathrm{T}}\beta$ and its derivatives up to third order are bounded on $\mathbb{R}$ for all $\beta \in \{\beta : |\beta - \beta_0| < \delta_\beta\}$ where $\delta_\beta > 0$ is a small constant. In addition, we assume that X is bounded.

(C2) The link function $g(\cdot)$ is bounded, and has finite jump points at $\underline{u}_j,\ j = 1, 2, \cdots, s$. $g(u)$ is right continuous at jump points. There exist constants $L > 0$ and $\alpha > 0$ such that $|g(u) - g(v)| \le L|u - v|^\alpha$ when $|u - v|$ is small for $u, v \in \mathbb{U}$. When $u \in D_1$, the derivatives up to third order of $g(u)$ are bounded. Further, $g(u)$ has left- and right- bounded second-order derivatives at the jump points.

(C3) $K_c(\cdot)$ is a symmetric density function with a bounded support with finite moments of all orders and is Lipschitz continuous.

(C4) $\lim\limits_{n\to\infty} h = 0,\ \lim\limits_{n\to\infty} nh^4 = \infty$.

(C5) $E\{\varepsilon - \varepsilon^{-1}|U = u\} = 0$, $\sigma_*^2(u) = E\{(\varepsilon - \varepsilon^{-1})^2|U = u\}$, $F(u) = E\{\varepsilon + \varepsilon^{-1}|U = u\}$, $G(u) = E\{\varepsilon^2 - \varepsilon^{-2}|U = u\}$ and $H(u) = E\{(\varepsilon + \varepsilon^{-1})^2|U = u\}$ are bounded and continuous with respect to u.

Remark 3.1 The Conditions (C1)–(C4) are used in the literature of the single-index model with jump points (see [9, 14]). Condition (C5) is often applied in the literature for relative error estimators, which is needed for asymptotic variance of estimators. In addition, the first moment condition $E\{\varepsilon - \varepsilon^{-1}|U = u\} = 0$ ensures consistency and the second moment condition $\sigma_*^2(u) = E\{(\varepsilon - \varepsilon^{-1})^2|U = u\} < \infty$ ensures the asymptotic normality of the LPRE estimator, which plays the same role as the assumption of zero mean and the finite second moment for the LS methods, respectively.

In order to obtain the main results, we need the following two lemmas, where $U_i = X_i^{\mathrm{T}}\beta$, $\widehat{U}_i = X_i^{\mathrm{T}}\widehat{\beta}$. For simplicity, we write

$$U_i^* = \begin{pmatrix} 1 \\ \frac{U_i-u}{h} \end{pmatrix}, \quad \theta = \begin{pmatrix} g(u) \\ hg'(u) \end{pmatrix}, \quad \widehat{\theta}_v = \begin{pmatrix} \widehat{a}_v(u) \\ h\widehat{b}_v(u) \end{pmatrix}.$$

Denote that $R_i = g(U_i) - g(u) - g'(u)(U_i - u)$ and $\gamma_{i,v} = g(U_i) - \widehat{a}_v(u) - \widehat{b}_v(u)(U_i - u) = R_i - (\widehat{\theta}_v - \theta)^{\mathrm{T}}U_i^*$. Then

$$Y_i \exp(-\widehat{a}_v(u) - \widehat{b}_v(u)(U_i - u)) = \varepsilon_i \exp(\gamma_{i,v})(1 + o_p(1))$$

and

$$Y_i^{-1} \exp(\widehat{a}_v(u) + \widehat{b}_v(u)(U_i - u)) = \varepsilon_i^{-1} \exp(-\gamma_{i,v})(1 + o_p(1)).$$

Lemma 3.1 *Under Conditions* (C1)–(C5), *we can obtain*

$$\frac{1}{n}\sum_{i=1}^{n} K_{h,v}(U_i - u)(\varepsilon_i + \varepsilon_i^{-1})\Big(\frac{U_i - u}{h}\Big)^j = F(u)f(u)\mu_{j,v} + o_p(1) \tag{3.1}$$

and

$$\frac{1}{n}\sum_{i=1}^{n} K_{h,v}(U_i - u)(\varepsilon_i + \varepsilon_i^{-1})R_i\Big(\frac{U_i - u}{h}\Big)^j = \frac{1}{2}h^2 g''(u)F(u)f(u)\mu_{j+2,v} + o_p(h^2), \tag{3.2}$$

where $F(u) = E(\varepsilon + \varepsilon^{-1}|U = u)$.

Proof We shall prove (3.1) since the same arguments can show (3.2). Denote $T_n = n^{-1}\sum\limits_{i=1}^{n} K_{h,v}(U_i - u)(\varepsilon_i + \varepsilon_i^{-1})\big(\frac{U_i-u}{h}\big)^j$. Firstly, by the change of variable, we obtain

$$\begin{aligned} E(T_n) &= n^{-1}\sum_{i=1}^{n} E\Big\{K_{h,v}(U_i - u)E[\varepsilon_i + \varepsilon_i^{-1}|U_i = u]\Big(\frac{U_i - u}{h}\Big)^j\Big\} \\ &= \int F(t)\Big(\frac{t - u}{h}\Big)^j K_{h,v}(U_i - u)f(t)\mathrm{d}t \\ &= \mu_{j,v}F(u)f(u) + o(1) \end{aligned}$$

and

$$\begin{aligned} \mathrm{Var}(T_n) &\le n^{-2}\sum_{i=1}^{n} E\Big\{K_{h,v}^2\Big(\frac{U_i - u}{h}\Big)E[(\varepsilon_i + \varepsilon_i^{-1})^2 \mid U_i = u]\Big(\frac{U_i - u}{h}\Big)^{2j}\Big\} \\ &= (nh)^{-1}\nu_{2j,v}H(u)f(u)(1 + o(1)), \end{aligned}$$

where $H(u) = E[(\varepsilon + \varepsilon^{-1})^2 \mid U = u]$. Based on the result $T_n = E(T_n) + O_p(\sqrt{\mathrm{Var}(T_n)})$ and the assumption $nh \to \infty$, it follows that

$$T_n = \mu_{j,v}F(u)f(u) + o_p(1).$$

This completes the proof of Lemma 3.1.

Lemma 3.2 *Under the regularity Conditions* (C1)–(C5), *with probability approaching* 1, *there exists a consistent local minimizer* $\widehat{\theta}_v$ *such that*

$$\|\widehat{\theta}_v - \theta\| = O_p((nh)^{-\frac{1}{2}} + h^2).$$

Proof Denote $\alpha_n = (nh)^{-\frac{1}{2}} + h^2$ and $\theta = \theta^* + \alpha_n \mu$. Then $(\widehat{a}_v(u), \widehat{b}_v(u))$ minimize

$$\sum_{i=1}^{n} \{Y_i \exp[-a_v(u) - b_v(u)(U_i - u)] + Y_i^{-1} \exp[a_v(u) + b_v(u)(U_i - u)] - 2\} K_{h,v}(U_i - u)$$

with respect to $(a_v(u), b_v(u))$. Then, with the foregoing notation, we can see that θ^* minimizes the function

$$L_n(\theta) = \sum_{i=1}^{n} [Y_i \exp(-\theta^T U_i^*) + Y_i^{-1} \exp(\theta^T U_i^*)] K_{h,v}(U_i - u)$$

with respect to θ. Note that $L_n(\theta)$ is convex in θ. It is sufficient to verify that for any given $\eta > 0$, there exists a large constant C satisfying

$$P\Big\{ \inf_{\|\boldsymbol{\mu}\| = C} L_n(\theta^* + \alpha_n \boldsymbol{\mu}) > L_n(\theta^*) \Big\} \geq 1 - \eta, \tag{3.3}$$

which implies that with asymptotic probability one, there exists a local minimizer of $L_n(\theta)$ in the ball $\{\theta^* + \alpha_n \boldsymbol{\mu} : \|\boldsymbol{\mu}\| \leq C\}$.

Using Taylor expansion, it follows that

$$\begin{aligned}
& L_n(\theta^* + \alpha_n \boldsymbol{\mu}) - L_n(\theta^*) \\
= {} & n^{-1} \sum_{i=1}^{n} K_{h,v}(U_i - u) \{\varepsilon_i \exp(R_i + \alpha_n \boldsymbol{\mu}^{\mathrm{T}} U_i^*)(1 + o_p(1)) \\
& + \varepsilon_i^{-1} \exp(-R_i + \alpha_n \boldsymbol{\mu}^{\mathrm{T}} U_i^*)(1 + o_p(1)) \\
& - \varepsilon_i \exp(R_i)[1 + o_p(1) - \varepsilon_i^{-1} \exp(-R_i)](1 + o_p(1))\} \\
= {} & n^{-1} \sum_{i=1}^{n} K_{h,v}(U_i - u) \{[\varepsilon_i \exp(R_i) - \varepsilon_i^{-1} \exp(-R_i)] \alpha_n \boldsymbol{\mu}^{\mathrm{T}} U_i^*\} \\
& + n^{-1} \sum_{i=1}^{n} K_{h,v}(U_i - u) \Big\{ \frac{1}{2} [\varepsilon_i \exp(R_i) + \varepsilon_i^{-1} \exp(-R_i)] \alpha_n^2 (\boldsymbol{\mu}^{\mathrm{T}} U_i^*)^2 \Big\} \\
& + n^{-1} \sum_{i=1}^{n} K_{h,v}(U_i - u) \Big\{ \frac{1}{6} [\varepsilon_i \exp(\xi_{1i}) + \varepsilon_i^{-1} \exp(-\xi_{2i})] \alpha_n^3 (\boldsymbol{\mu}^T U_i^*)^3 \Big\} + o_p(1) \\
\triangleq {} & \mathrm{I}_1 + \mathrm{I}_2 + \mathrm{I}_3 + o_p(1),
\end{aligned} \tag{3.4}$$

where ξ_{1i} is between R_i and $R_i + \alpha_n \boldsymbol{\mu}^{\mathrm{T}} U_i^*$, ξ_{2i} is between $-R_i$ and $-R_i + \alpha_n \boldsymbol{\mu}^{\mathrm{T}} U_i^*$. Write $R_i = g(U_i) - g(u) - g'(u)(U_i - u)$, then $R_i = \frac{1}{2} g''(u)(U_i - u)^2 (1 + o(1)) = O(h^2)$ on the support of $K_{h,v}(\cdot)$. By Taylor expansion that $\mathrm{e}^x = 1 + x(1 + o(1))$ as $|x| \to 0$, we have

$$\mathrm{I}_1 = n^{-1} \sum_{i=1}^{n} K_{h,v}(U_i - u) \Big[(\varepsilon_i - \varepsilon_i^{-1}) + (\varepsilon_i + \varepsilon_i^{-1}) \frac{1}{2} g''(u)(U_i - u)^2 (1 + o(1)) \Big] \alpha_n \boldsymbol{\mu}^{\mathrm{T}} U_i^*.$$

By directly calculating the mean and variance, we derive

$$\begin{aligned} E(\mathrm{I}_1) &= \alpha_n h^2 E\Big\{K_{h,v}(U_i-u)E(\varepsilon_i+\varepsilon_i^{-1}|U_i)\frac{1}{2}g''(u)(U_i-u)^2\boldsymbol{\mu}^{\mathrm{T}}U_i^*\Big\}(1+o(1)) \\ &= O(C\alpha_n h^2), \\ \mathrm{Var}(\mathrm{I}_1) &= n^{-1}\alpha_n^2\mathrm{Var}\Big\{K_{h,v}(U_i-u)[(\varepsilon_i-\varepsilon_i^{-1})+(\varepsilon_i+\varepsilon_i^{-1})\frac{1}{2}g''(u)(U_i-u)^2]\boldsymbol{\mu}^{\mathrm{T}}U_i^*\Big\} \\ &= O(C^2\alpha_n^2(nh)^{-1}). \end{aligned}$$

Hence, $\mathrm{I}_1 = O(C\alpha_n h^2) + C\alpha_n O_p((nh)^{-\frac{1}{2}}) = O_p(\alpha_n^2)$. Similarly, $\mathrm{I}_3 = O_p(\alpha_n^3)$.

From Lemma 3.1, it follows that

$$\begin{aligned} \mathrm{I}_2 &= \frac{1}{2n}\sum_{i=1}^n K_{h,v}(U_i-u)(\varepsilon_i+\varepsilon_i^{-1})\alpha_n^2(\boldsymbol{\mu}^{\mathrm{T}}U_i^*)^2 \\ &\quad + \frac{1}{2n}\sum_{i=1}^n K_{h,v}(U_i-u)(\varepsilon_i-\varepsilon_i^{-1})R_i(1+o(1))\alpha_n^2(\boldsymbol{\mu}^{\mathrm{T}}U_i^*)^2 \\ &= \alpha_n^2 F(u)f(u)\boldsymbol{\mu}^{\mathrm{T}}S_v\boldsymbol{\mu}(1+o(1)), \end{aligned}$$

where S_v is a positive matrix, $\|\boldsymbol{\mu}\| = C$ and $F(u) > 0$, we can choose C large enough such that I_2 dominates both I_1 and I_3 with a probability if at least $1-\eta$. Thus, (3.3) holds. Hence, with the probability approaching 1, there exists a local minimizer $\widehat{\theta}_v$ such that $\|\widehat{\theta}_v - \theta\| \le \alpha_n C$. Based on the definition of θ, we can get $\|\widehat{\theta}_v - \theta\| = O_p(h^2 + (nh)^{-\frac{1}{2}})$. The proof is completed.

Let $\mu_{j,v} = \int t^j K_v(t)\mathrm{d}t$ and $\nu_{j,v} = \int t^j K_v^2(t)\mathrm{d}t$, $v = c, l, r$. When the link function is continuous, β can be estimated $\sqrt{n}$ consistently (see [16]). We consider that our estimator $\widehat{\beta}$ is in a neighborhood of β, $\widehat{\beta} \in \Delta_n = \{\beta : \|\beta-\beta_0\|_2 \le Cn^{-\frac{1}{2}+c_0}, 0 \le c_0 < \frac{1}{2}\}$, in which C is some positive constant.

Theorem 3.1 *Under Conditions* (C1)–(C5)*, the mean squared errors* (MSE *for short*) *of the three estimators of the link function for* $\beta \in \Delta_n$ *are as follows*:

(i) *For any* $u \in D_1$*,*

$$\mathrm{MSE}(\widehat{a}_v(u)) = \Big\{\frac{1}{2}h^2 g''(u)B_v\Big\}^2 + \frac{\sigma_*^2(u)V_v}{nhF^2(u)f(u)} + o\Big(h^4 + \frac{1}{nh}\Big), \quad v = c, l, r,$$

where $B_v = \frac{\mu_{2,v}^2 - \mu_{1,v}\mu_{3,v}}{\mu_{0,v}\mu_{2,v} - \mu_{1,v}^2}$ *and* $V_v = \frac{\mu_{2,v}^2\nu_{0,v} - 2\mu_{1,v}\mu_{2,v}\nu_{1,v} + \mu_{1,v}^2\nu_{2,v}}{(\mu_{0,v}\mu_{2,v} - \mu_{1,v}^2)^2}$.

(ii) *For any* $u \in D_{2,l}$*, that is,* $u = \underline{u}_j - \tau h$ *with* $\tau \in (0, \delta)$*, we have*

$$\begin{aligned} \mathrm{MSE}(\widehat{a}_l(u)) &= \Big\{\frac{1}{2}h^2 g''(\underline{u}_j-)B_l\Big\}^2 + \frac{\sigma_*^2(u)V_l}{nhF^2(u)f(u)} + o\Big(h^4 + \frac{1}{nh}\Big), \\ \mathrm{MSE}(\widehat{a}_r(u)) &= \Big\{d_j\int_\tau^\delta K_r(w)\frac{\mu_{2,r} - \mu_{1,r}w}{\mu_{0,r}\mu_{2,r} - \mu_{1,r}^2}\mathrm{d}w\Big\}^2 + \frac{\sigma_*^2(u)V_r}{nhF^2(u)f(u_x)} + o\Big(\frac{1}{nh}\Big), \\ \mathrm{MSE}(\widehat{a}_c(u)) &= \Big\{d_j\int_\tau^\delta K_c(w)\mathrm{d}w\Big\}^2 + \frac{\sigma_*^2(u)V_c}{nhF^2(u)f(u)} + o\Big(\frac{1}{nh}\Big). \end{aligned}$$

(iii) *For any $u \in D_{2,r}$, that is, $u = \underline{u}_j + \tau h$ with $\tau \in [0, \delta)$, we have*

$$\mathrm{MSE}(\widehat{a}_r(u)) = \Big\{\frac{1}{2}h^2 g''(\underline{u}_j+)B_r\Big\}^2 + \frac{\sigma_*^2(u)V_r}{nhF^2(u)f(u)} + o\Big(h^4 + \frac{1}{nh}\Big),$$

$$\mathrm{MSE}(\widehat{a}_l(u)) = \Big\{ - d_j \int_{-\delta}^{-\tau} K_l(w)\frac{\mu_{2,l} - \mu_{1,l}w}{\mu_{0,l}\mu_{2,l} - \mu_{1,l}^2}\mathrm{d}w\Big\}^2 + \frac{\sigma_*^2(u)V_l}{nhF^2(u)f(u)} + o\Big(\frac{1}{nh}\Big),$$

$$\mathrm{MSE}(\widehat{a}_c(u)) = \Big\{ - d_j \int_{-\delta}^{-\tau} K_c(w)\mathrm{d}w\Big\}^2 + \frac{\sigma_*^2(u)V_c}{nhF^2(u)f(u)} + o\Big(\frac{1}{nh}\Big).$$

Proof (i) Suppose $u \in D_1$. The solution $\widehat{\theta}_v$ satisfies the following equation:

$$0 = \sum_{i=1}^n K_{h,v}(U_i - u)\{Y_i \exp(-\widehat{a}_v(u) - \widehat{b}_v(u)(U_i - u)) - Y_i^{-1}\exp(\widehat{a}_v(u) + \widehat{b}_v(u)(U_i - u))\}U_i^*.$$

Applying Lemma 3.2, it is easy to obtain that $\gamma_{i,v} = O_p(h^2 + \|\widehat{\theta}_v - \theta\|) = O_p(\|\widehat{\theta}_v - \theta\|) = o_p(1)$. Then, by Taylor expansion, $\mathrm{e}^x = 1 + x(1 + o(1))$ as $|x| \to 0$, we have

$$\begin{aligned} 0 &= \sum_{i=1}^n K_{h,v}(U_i - u)\{\varepsilon_i \exp(\gamma_{i,v}) - \varepsilon_i^{-1}\exp(-\gamma_{i,v})\}U_i^* + o_p(1) \\ &= \sum_{i=1}^n K_{h,v}(U_i - u)(\varepsilon_i - \varepsilon_i^{-1})U_i^* + \sum_{i=1}^n K_{h,v}(U_i - u)(\varepsilon_i + \varepsilon_i^{-1})\gamma_{i,v}(1 + o_p(1))U_i^* + o_p(1) \\ &\triangleq W_{1,v} + W_{2,v} + o_p(1). \end{aligned} \tag{3.5}$$

Furthermore, the second term of the above expression can be expressed as follows:

$$\begin{aligned} W_{2,v} &= \sum_{i=1}^n K_{h,v}(U_i - u)(\varepsilon_i + \varepsilon_i^{-1})R_iU_i^* - \sum_{i=1}^n K_{h,v}(U_i - u)(\varepsilon_i + \varepsilon_i^{-1})U_i^*U_i^{*\mathrm{T}}(\widehat{\theta}_v - \theta) + o_p(1) \\ &\triangleq W_{21,v} + W_{22,v}. \end{aligned}$$

Using Lemma 3.1, we obtain

$$W_{21,v} = \frac{1}{2}nh^2F(u)f(u)\begin{pmatrix}\mu_{2,v}\\ \mu_{3,v}\end{pmatrix}g''(u) + o_p(nh^2)$$

and

$$\begin{aligned} W_{22,v} &= -nF(u)f(u)\begin{pmatrix}\mu_{0,v} & \mu_{1,v}\\ \mu_{1,v} & \mu_{2,v}\end{pmatrix}(\widehat{\theta}_v - \theta)(1 + o_p(1)) \\ &= -nF(u)f(u)S_v(\widehat{\theta}_v - \theta)(1 + o_p(1)), \end{aligned}$$

where S_v denotes a 2×2 matrix. Then, it follows from (3.5) that

$$(\widehat{\theta}_v - \theta) = \frac{1}{2}h^2 g''(u)S_v^{-1}\begin{pmatrix}\mu_{2,v}\\ \mu_{3,v}\end{pmatrix} + \frac{S_v^{-1}W_{1,v}}{nF(u)f(u)} + o_p(1). \tag{3.6}$$

Based on Condition (C5), we can easily get $E\{W_{1,v}\} = 0$. Similar to Lemma 3.1, we have

$$E\Big\{\sum_{i=1}^n K_{h,v}^2(U_i - u)(\varepsilon_i - \varepsilon_i^{-1})^2\Big(\frac{U_i - u}{h}\Big)^j\Big\} = \frac{1}{h}\nu_{j,v}\sigma_*^2(u)f(u)(1 + o_p(1)).$$

So

$$\begin{aligned}\mathrm{Cov}(W_{1,v}) &= nh^{-1}\sigma_*^2(u)f(u)\begin{pmatrix}\nu_{0,v} & \nu_{1,v}\\ \nu_{1,v} & \nu_{2,v}\end{pmatrix}(1+o_p(1))\\ &= nh^{-1}\sigma_*^2(u)f(u)S_v^*(1+o_p(1)),\end{aligned}$$

where S_v^* denotes a 2×2 matrix. Based on (3.6), the asymptotic bias and variance of $\widehat{a}_v(u)$ are naturally given by

$$\mathrm{bias}(\widehat{a}_v(u)) = \frac{1}{2}h^2 g''(u)B_v + o(h^2)$$

and

$$\mathrm{Var}(\widehat{a}_v(u)) = \frac{\sigma_*^2(u)V_v}{nhF^2(u)f(u)} + o\Big(\frac{1}{nh}\Big),$$

where $B_v = \frac{\mu_{2,v}^2-\mu_{1,v}\mu_{3,v}}{\mu_{0,v}\mu_{2,v}-\mu_{1,v}^2}$ and $V_v = \frac{\mu_{2,v}^2\nu_{0,v}-2\mu_{1,v}\mu_{2,v}\nu_{1,v}+\mu_{1,v}^2\nu_{2,v}}{(\mu_{0,v}\mu_{2,v}-\mu_{1,v}^2)^2}$. Thus the first part of Theorem 3.1 is proved.

(ii) Suppose $u\in D_{2,l}$. Let $u=\underline{u}_j-\tau h, \tau\in(0,\delta)$. Similar to (i), we can obtain the expression of $\mathrm{MSE}(\widehat{a}_l(u))$. The solution $\widehat{\theta}_r$ satisfies the equation

$$0=\sum_{i=1}^n K_{h,r}(U_i-u)(\varepsilon_i-\varepsilon_i^{-1})U_i^* + \sum_{i=1}^n K_{h,r}(U_i-u)(\varepsilon_i+\varepsilon_i^{-1})\gamma_{i,r}(1+o_p(1))U_i^*. \tag{3.7}$$

Note that the second term on the left side of (3.7) can be expressed as follows:

$$\sum_{i=1}^n K_{h,r}(U_i-u)(\varepsilon_i+\varepsilon_i^{-1})R_iU_i^* - \sum_{i=1}^n K_{h,r}(U_i-u)(\varepsilon_i+\varepsilon_i^{-1})U_i^*U_i^{*\mathrm{T}}(\widehat{\theta}_r-\theta) \triangleq J_{1,r}-J_{2,r}.$$

Applying Lemma 3.1, we can get

$$\begin{aligned}J_{1,r} &= \sum_{u_i\ge\underline{u}_j} K_{h,r}(U_i-u)(\varepsilon_i+\varepsilon_i^{-1})(d_j+o(1))U_i^* + \sum_{u_i<\underline{u}_j} K_{h,r}(U_i-u)(\varepsilon_i+\varepsilon_i^{-1})U_i^*o(1)\\ &= d_j\int_{u\ge\underline{u}_j} K_{h,r}(U_i-u)\begin{pmatrix}1\\ \frac{U_i-u}{h}\end{pmatrix}(\varepsilon_i+\varepsilon_i^{-1})f(u)\mathrm{d}u+o(1)\\ &= nd_jF(u)f(u)\begin{pmatrix}\int_\tau^\delta K_r(w)\mathrm{d}w\\ \int_\tau^\delta K_r(w)w\mathrm{d}w\end{pmatrix}+o(1)\end{aligned}$$

and

$$J_{2,r} = nF(u)f(u)S_r(1+o_p(1))(\widehat{\theta}_r-\theta).$$

Then the asymptotic bias and variance of $\widehat{a}_r(u)$ are naturally given by

$$\mathrm{bias}(\widehat{a}_r(u)) = d_j\int_\tau^\delta K_r(w)\frac{\mu_{2,r}-\mu_{1,r}w}{\mu_{0,r}\mu_{2,r}-\mu_{1,r}^2}\mathrm{d}w+o(1)$$

and

$$\mathrm{Var}(\widehat{a}_r(u)) = \frac{\sigma_*^2(u)V_r}{nhF^2(u)f(u)} + o\Big(\frac{1}{nh}\Big).$$

Similarly, we can prove the third equation of (ii).

(iii) The third part of Theorem 3.1 can be proved in the same way as (ii).

Remark 3.2 Since $\mathrm{MSE}(\widehat{a}_v(u)) = \mathrm{Bias}(\widehat{a}_v(u))^2 + \mathrm{Var}(\widehat{a}_v(u))$, from the first part of Theorem 3.1, the three estimators are consistent in the continuity regions of $g(u)$. However, when $u \in D_{2,l}$, only $\widehat{a}_l(u)$ is consistent. $\widehat{a}_r(u)$ and $\widehat{a}_c(u)$ are inconsistent in the left neighborhood of the jump point. Similarly, when $u \in D_{2,r}$, only $\widehat{a}_r(u)$ is consistent.

Theorem 3.2 *Under Conditions* (C1)–(C5), *we can obtain the asymptotic expressions of* WREMS *for $\beta \in \Delta_n$ as follows:*

(i) *For any $u \in D_1$,*

$$\mathrm{WREMS}_v(u) = \sigma_*^2(u) + R_{1,v}(u), \quad v = c, l, r,$$

where $R_{1,v}(u)$, $v = c, l, r$ are asymptotically zero with probability 1 *and uniformly in $u \in D_1$.*

(ii) *For any $u \in D_{2,l}$, that is, $u = \underline{u}_j - \tau h$ with $\tau \in (0, \delta)$, we have*

$$\begin{aligned}
\mathrm{WREMS}_l(u) &= \sigma_*^2(u) + R_{2,l}(u),\\
\mathrm{WREMS}_r(u) &= \sigma_*^2(u) + \mu_{0,r}^{-1}H(u)d_j^2C_{-\tau,r}^2 + \mu_{0,r}^{-1}G(u)d_jC_{-\tau,r} + R_{2,r}(u),\\
\mathrm{WREMS}_c(u) &= \sigma_*^2(u) + \mu_{0,c}^{-1}H(u)d_j^2C_{-\tau,c}^2 + \mu_{0,c}^{-1}G(u)d_jC_{-\tau,c} + R_{2,c}(u),
\end{aligned}$$

where $R_{2,v}(u)$, $v = c, l, r$ are asymptotically zero with probability 1 *and uniformly in $u \in [\underline{u}_j - \tau h, \underline{u}_j)$.*

(iii) *For any $u \in D_{2,r}$, that is, $u = \underline{u}_j + \tau h$ with $\tau \in [0, \delta)$, we have*

$$\begin{aligned}
\mathrm{WREMS}_r(u) &= \sigma_*^2(u) + R_{3,r}(u),\\
\mathrm{WREMS}_l(u) &= \sigma_*^2(u) + \mu_{0,l}^{-1}H(u)d_j^2C_{\tau,l}^2 + \mu_{0,l}^{-1}G(u)d_jC_{\tau,l} + R_{3,l}(u),\\
\mathrm{WREMS}_c(u) &= \sigma_*^2(u) + \mu_{0,c}^{-1}H(u)d_j^2C_{\tau,c}^2 + \mu_{0,c}^{-1}G(u)d_jC_{\tau,c} + R_{3,c}(u),
\end{aligned}$$

where $R_{3,v}(u)$, $v = c, l, r$ are asymptotically zero with probability 1 *and uniformly in $u \in [\underline{u}_j, \underline{u}_j + \tau h]$, in which*

$$\begin{aligned}
&C_{\tau,v}^2\\
&= \int_{-\tau}^{\delta}\Big\{d_j\int_{-\delta}^{-\tau}K_{h,v}(w)\frac{\mu_{2,v}-\mu_{1,v}w}{\mu_{0,v}\mu_{2,v}-\mu_{1,v}^2}\mathrm{d}w - sd_j\int_{-\tau}^{\delta}K_{h,v}(w)\frac{\mu_{0,v}w-\mu_{1,v}}{\mu_{0,v}\mu_{2,v}-\mu_{1,v}^2}\mathrm{d}w\Big\}^2K_{h,v}(s)\mathrm{d}s\\
&\quad + \int_{-\delta}^{-\tau}\Big\{d_j\int_{-\tau}^{\delta}K_{h,v}(w)\frac{\mu_{2,v}-\mu_{1,v}w}{\mu_{0,v}\mu_{2,v}-\mu_{1,v}^2}\mathrm{d}w + sd_j\int_{-\tau}^{\delta}K_{h,v}(w)\frac{\mu_{0,v}w-\mu_{1,v}}{\mu_{0,v}\mu_{2,v}-\mu_{1,v}^2}\mathrm{d}w\Big\}^2K_{h,v}(s)\mathrm{d}s.
\end{aligned}$$

Proof From Step 2 in Subsection 2.2, it follows that

$$\mathrm{WREMS}_v(u) = \frac{\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}[\varepsilon_{i,v} - \varepsilon_{i,v}^{-1}]^2 K_{h,v}(U_i - u)}{\dfrac{1}{n}\displaystyle\sum_{i=1}^{n} K_{h,v}(U_i - u)}, \tag{3.8}$$

where $\varepsilon_{i,v} = Y_i \exp(-\widehat{a}_v(u) - \widehat{b}_v(u)(U_i - u))$ and $\varepsilon_{i,v}^{-1} = Y_i^{-1}\exp(\widehat{a}_v(u) + \widehat{b}_v(u)(U_i - u))$. Using Lemma 3.1, the denominator of (3.8) can be written as

$$\frac{1}{n}\sum_{i=1}^{n} K_{h,v}(U_i - u) = f(u)\mu_{0,v} + o_p(1). \tag{3.9}$$

Then, the numerator of (3.8) can be expressed as

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} K_{h,v}(U_i - u)\{(\varepsilon_i - \varepsilon_i^{-1}) + (\varepsilon_i + \varepsilon_i^{-1})[g(U_i) - \widehat{a}_v(u) - \widehat{b}_v(u)(U_i - u)]\}^2 + o_p(1)\\
&= \frac{1}{n}\sum_{i=1}^{n} K_{h,v}(U_i - u)(\varepsilon_i - \varepsilon_i^{-1})^2\\
&\quad + \frac{1}{n}\sum_{i=1}^{n} K_{h,v}(U_i - u)(\varepsilon_i + \varepsilon_i^{-1})^2[g(U_i) - \widehat{a}_v(u) - \widehat{b}_v(u)(U_i - u)]^2\\
&\quad + \frac{2}{n}\sum_{i=1}^{n} K_{h,v}(U_i - u)(\varepsilon_i - \varepsilon_i^{-1})(\varepsilon_i + \varepsilon_i^{-1})[g(U_i) - \widehat{a}_v(u) - \widehat{b}_v(u)(U_i - u)] + o_p(1)\\
&\triangleq \mathrm{M}_1 + \mathrm{M}_2 + \mathrm{M}_3 + o_p(1).
\end{aligned} \tag{3.10}$$

Suppose $u \in D_1$. By Lemma 3.2, we have $g(U_i) - \widehat{a}_v(u) - \widehat{b}_v(u)(U_i - u) = O_p(h^2 + \|\widehat{\theta}_v - \theta\|) = O_p\big(h^2 + \frac{1}{\sqrt{nh}}\big) = o_p(1)$. From Theorem 3.1 and Conditions (C4)–(C5), it is clear to obtain

$$\mathrm{M}_1 = \sigma_*^2(u) f(u)\mu_{0,v} + o_p(1), \tag{3.11}$$

$$\mathrm{M}_2 = O_p\Big(h^4 + \frac{1}{nh}\Big) O_p\Big(\frac{1}{\sqrt{nh}}\Big) = o_p(1), \tag{3.12}$$

$$\mathrm{M}_3 = O_p\Big(h^2 + \frac{1}{\sqrt{nh}}\Big) O_p\Big(\frac{1}{\sqrt{nh}}\Big) = o_p(1). \tag{3.13}$$

Therefore, combining (3.8)–(3.13) gives

$$\mathrm{WREMS}_v(u) = \sigma_*^2(u) + o_p(1).$$

Suppose $u \in D_{2,l}$, that is, $u = \underline{u}_j - \tau h$ with $\tau \in (0, \delta)$. $\mathrm{WREMS}_l(u)$ can be obtained in the same way as (i). Next, we prove the expression of $\mathrm{WREMS}_r(u)$. Firstly, from the proof of the second part of Theorem 3.1, we obtain

$$\widehat{a}_r(u) = g(u-) + d_j \int_{\tau}^{\delta} K_r(w) \frac{\mu_{2,r} - \mu_{1,r} w}{\mu_{0,r}\mu_{2,r} - \mu_{1,r}^2} \mathrm{d}w + o_p(1) \tag{3.14}$$

and

$$\widehat{b}_r(u) = g'(u-) + \frac{d_j}{h}\int_\tau^\delta K_r(w)\frac{\mu_{0,r}w - \mu_{1,r}}{\mu_{0,r}\mu_{2,r} - \mu_{1,r}^2}\mathrm{d}w + o_p\Big(\frac{1}{h}\Big). \tag{3.15}$$

Plugging (3.14) and (3.15) into the expression of M_2, we obtain

$$\begin{aligned}
\mathrm{M}_2 &= \frac{1}{n}\sum_{i=1}^n K_{h,r}(U_i-u)(\varepsilon_i+\varepsilon_i^{-1})^2[g(U_i)-\widehat{a}_r(u)-\widehat{b}_r(u)(U_i-u)]^2K_{h,r}(U_i-u)\\
&= \frac{1}{n}\sum_{U_i<\underline{u}_j} K_{h,r}(U_i-u)(\varepsilon_i+\varepsilon_i^{-1})^2\Big[g(U_i)-g(\underline{u}_j-)-d_j\int_\tau^\delta K_r(w)\frac{\mu_{2,r}-\mu_{1,r}w}{\mu_{0,r}\mu_{2,r}-\mu_{1,r}^2}\mathrm{d}w\\
&\quad - g'(\underline{u}_j-)-d_j\Big(\frac{U_i-u}{h}\Big)\int_\tau^\delta K_r(w)\frac{\mu_{0,r}w-\mu_{1,r}}{\mu_{0,r}\mu_{2,r}-\mu_{1,r}^2}\mathrm{d}w+o_p(1)+o_p\Big(\frac{1}{h}\Big)\Big]^2\\
&\quad + \frac{1}{n}\sum_{U_i\geq\underline{u}_j} K_{h,r}(U_i-u)(\varepsilon_i+\varepsilon_i^{-1})^2\Big[g(U_i)-g(\underline{u}_j-)-d_j\int_\tau^\delta K_r(w)\frac{\mu_{2,r}-\mu_{1,r}w}{\mu_{0,r}\mu_{2,r}-\mu_{1,r}^2}\mathrm{d}w\\
&\quad - g'(\underline{u}_j-)-d_j\Big(\frac{U_i-u}{h}\Big)\int_\tau^\delta K_r(w)\frac{\mu_{0,r}w-\mu_{1,r}}{\mu_{0,r}\mu_{2,r}-\mu_{1,r}^2}\mathrm{d}w+o_p(1)+o_p\Big(\frac{1}{h}\Big)\Big]^2\\
&= f(u)H(u)\int_{-\delta}^\tau\Big\{d_j\int_\tau^\delta K_{h,r}(w)\frac{\mu_{2,r}-\mu_{1,r}w}{\mu_{0,r}\mu_{2,r}-\mu_{1,r}^2}\mathrm{d}w\\
&\quad + sd_j\int_\tau^\delta K_{h,r}(w)\frac{\mu_{0,r}w-\mu_{1,r}}{\mu_{0,r}\mu_{2,r}-\mu_{1,r}^2}\mathrm{d}w\Big\}^2K_{h,r}(s)\mathrm{d}s\\
&\quad + f(u)H(u)\int_\tau^\delta\Big\{d_j\int_{-\delta}^\tau K_{h,r}(w)\frac{\mu_{2,r}-\mu_{1,r}w}{\mu_{0,r}\mu_{2,r}-\mu_{1,r}^2}\mathrm{d}w\\
&\quad - sd_j\int_\tau^\delta K_{h,r}(w)\frac{\mu_{0,r}w-\mu_{1,r}}{\mu_{0,r}\mu_{2,r}-\mu_{1,r}^2}\mathrm{d}w\Big\}^2K_{h,r}(s)\mathrm{d}s\\
&= d_j^2C_{-\tau,r}^2H(u)f(u)+o_p(1),
\end{aligned} \tag{3.16}$$

where $H(u) = E\{(\varepsilon+\varepsilon^{-1})^2 \mid U=u\}$ is defined in Condition (C5). Similarly, using Condition (C5), we can obtain $\mathrm{M}_3 = d_jC_{-\tau,r}G(u)f(u)+o_p(1)$. Using Lemma 3.1 and Condition (C5), we obtain $\mathrm{M}_1 = \sigma_*^2(u)f(u)\mu_{0,v}+o_p(1)$. It follows from (3.8)–(3.10) and (3.15)–(3.16) that

$$\mathrm{WREMS}_r(u) = \sigma_*^2(u)+\mu_{0,r}^{-1}H(u)d_j^2C_{-\tau,r}^2+\mu_{0,r}^{-1}G(u)d_jC_{-\tau,r}+o_p(1).$$

(iii) Suppose $u\in D_{2,r}$. The third part can be obtained in the same way.

Remark 3.3 If the error term ε satisfies the condition $G(u)=E\{\varepsilon^2-\varepsilon^{-2}|U=u\}=0$, then the asymptotic expressions (ii) and (iii) degenerate to (ii)$'$ and (iii)$'$, respectively.

(ii)$'$ For any $u\in D_{2,l}$, that is, $u=\underline{u}_j-\tau h$ with $\tau\in(0,\delta)$, we have

$$\begin{aligned}
\mathrm{WREMS}_l(u) &= \sigma_*^2(u)+R_{2,l}(u),\\
\mathrm{WREMS}_r(u) &= \sigma_*^2(u)+\mu_{0,r}^{-1}H(u)d_j^2C_{-\tau,r}^2+R_{2,r}(u),\\
\mathrm{WREMS}_c(u) &= \sigma_*^2(u)+\mu_{0,c}^{-1}H(u)d_j^2C_{-\tau,c}^2+R_{2,c}(u).
\end{aligned}$$

(iii)$'$ For any $u \in D_{2,r}$, that is, $u = \underline{u}_j + \tau h$ with $\tau \in [0, \delta)$, we have

$$\begin{aligned}\text{WREMS}_r(u) &= \sigma_*^2(u) + R_{3,r}(u),\\ \text{WREMS}_l(u) &= \sigma_*^2(u) + \mu_{0,l}^{-1} H(u) d_j^2 C_{\tau,l}^2 + R_{3,l}(u),\\ \text{WREMS}_c(u) &= \sigma_*^2(u) + \mu_{0,c}^{-1} H(u) d_j^2 C_{\tau,c}^2 + R_{3,c}(u),\end{aligned}$$

where $R_{j,v}(u)$, $j = 2, 3$, $v = c, l, r$ are asymptotically zero with probability 1 and uniformly in u.

Remark 3.4 From Theorem 3.2, we can conclude that the three WREMSs are consistent estimator of $\sigma_*^2(u) = E\{(\varepsilon - \varepsilon^{-1})^2 \mid U = u\}$ in the continuity regions. However, when $u \in D_{2,l}$, only $\text{WREMS}_l(u)$ is consistent, and when $u \in D_{2,r}$, only $\text{WREMS}_r(u)$ is consistent.

Theorem 3.3 *Under Conditions* (C1)–(C5), *for any* $u = X^{\mathrm{T}}\beta \in D_1$ *as* $n \to \infty$, *we have*

$$\sqrt{nh}\Big\{\widehat{g}(u) - g(u) - \frac{1}{2}h^2 g''(u) B_v\Big\} \xrightarrow{d} N\Big(0, \frac{\sigma_*^2(u) V_v}{F^2(u) f(u)}\Big),$$

for any $u = X^{\mathrm{T}}\beta \in D_{2,l}$ *located in the left neighborhood of* $\underline{u}_j$, *we have*

$$\sqrt{nh}\Big\{\widehat{g}(u) - g(u) - \frac{1}{2}h^2 g''(\underline{u}_j-) B_v\Big\} \xrightarrow{d} N\Big(0, \frac{\sigma_*^2(u) V_v}{F^2(u) f(u)}\Big),$$

for any $u = X^{\mathrm{T}}\beta \in D_{2,r}$ *located in the right neighborhood of* $\underline{u}_j$, *we have*

$$\sqrt{nh}\Big\{\widehat{g}(u) - g(u) - \frac{1}{2}h^2 g''(\underline{u}_j+) B_v\Big\} \xrightarrow{d} N\Big(0, \frac{\sigma_*^2(u) V_v}{F^2(u) f(u)}\Big),$$

where B_v *and* V_v *are defined in Theorem* 3.1.

Proof For $u \in D_1$, similar to the proof of Theorem 3.1, we have

$$\sqrt{nh}\Big\{\widehat{a}_v(u) - g(u) - \frac{1}{2}h^2 g''(u) B_v\Big\} \xrightarrow{d} N\Big(0, \frac{\sigma_*^2(u) V_v}{F^2(u) f(u)}\Big). \tag{3.17}$$

In addition, from Theorem 3.1, we can get that for the left-side estimator, (3.17) holds for $u \in D_1 \cup D_{2,l}$ and for the right-side estimator, (3.17) holds for $u \in D_1 \cup D_{2,r}$.

For any $u \in D_1 \cup D_2$, $\widehat{g}(u) = \widehat{a}_c(u) I\{D_1(u)\} + \widehat{a}_l(u) I\{D_{2,l}(u)\} + \widehat{a}_r(u) I\{D_{2,r}(u)\}$. It is easy to know that $D_1(u)$, $D_{2,l}$ and $D_{2,r}$ are mutually exclusive, and

$$I\{D_1(u)\} + I\{D_{2,l}(u)\} + I\{D_{2,r}(u)\} = 1.$$

For any $u \in D_1$, by Theorem 3.2, $\text{diff}(u) \to 0$ and $\lambda \to 0$ as $n \to \infty$. It means that in the continuity regions, $\widehat{g}(u) = \widehat{a}_c(u)$ a.s.

For any $u \in D_{2,l}$, suppose $u = \underline{u}_j - \tau h, \tau \in (0, \delta)$. From the second part of Theorem 3.2, we obtain

$$\text{diff}(u) = \max\{\text{WREMS}_c(u) - \text{WREMS}_r(u), \text{WREMS}_c(u) - \text{WREMS}_l(u)\}$$

$$= \max\{(\mu_{0,c}^{-1} - \mu_{0,r}^{-1})d_j C_{-\tau,c}[H(u)d_j C_{-\tau,c} + G(u)] + R_{2,c}(u) - R_{2,r}(u),$$
$$\mu_{0,c}^{-1} d_j C_{-\tau,c}[H(u)d_j C_{-\tau,c} + G(u)] + R_{2,c}(u) - R_{2,l}(u)\}.$$

Since

$$\lim_{n\to\infty} \mathrm{diff}(u) = \mu_{0,c}^{-1} d_j C_{-\tau,c}[H(u)d_j C_{-\tau,c} + G(u)],$$

where $0 < \lambda < \mu_{0,c}^{-1} d_j C_{-\tau,c}[H(u)d_j C_{-\tau,c} + G(u)]$, we have $I\{D_1(u)\} = 0$. Note that

$$\begin{aligned} \mathrm{WREMS}_r(u) - \mathrm{WREMS}_l(u) &= \mu_{0,c}^{-1} d_j C_{-\tau,c}[H(u)d_j C_{-\tau,c} + G(u)] + R_{2,c}(u) - R_{2,l}(u) \\ &\to \mu_{0,c}^{-1} d_j C_{-\tau,c}[H(u)d_j C_{-\tau,c} + G(u)] > 0, \end{aligned}$$

so $I\{D_{2,r}(u)\} = 0$. Consequently, $I\{D_{2,l}(u)\} = 1$ a.s., i.e., $\widehat{g}(u) = \widehat{a}_l(u)$.

For $u \in D_{2,r}$, similarly, we have $\widehat{g}(u) = \widehat{a}_r(u)$.

Because $\widehat{a}_c(u)$, $\widehat{a}_l(u)$ and $\widehat{a}_r(u)$ are asymptotically normal in D_1, $D_{2,l}$ and $D_{2,r}$, respectively, Theorem 3.3 is proved.

Theorem 3.4 *Under Conditions* (C1)–(C5)*, as $n \to \infty$, we have*

$$\sqrt{n}\{\widehat{\beta}_{\mathrm{JPLLPRE}} - \beta_0\} \xrightarrow{d} N(0, Q^- \Sigma Q^-),$$

where Q^- is a generalized inverse of Q,

$$\Sigma = E\Big\{(\varepsilon - \varepsilon^{-1})^2 \Big[g'(X^{\mathrm{T}}\beta_0)X - \frac{E\{(\varepsilon + \varepsilon^{-1})g'(X^{\mathrm{T}}\beta_0)X \mid X^{\mathrm{T}}\beta_0\}}{F(X^{\mathrm{T}}\beta_0)}\Big]^{\otimes 2}\Big\}$$

and

$$Q = E\Big\{(\varepsilon + \varepsilon^{-1}) \Big[g'(X^{\mathrm{T}}\beta_0)X - \frac{E\{(\varepsilon + \varepsilon^{-1})g'(X^{\mathrm{T}}\beta_0)X \mid X^{\mathrm{T}}\beta_0\}}{F(X^{\mathrm{T}}\beta_0)}\Big]^{\otimes 2}\Big\}.$$

Proof Theorem 3.4 can be proved in the same way as [16], so it is omitted here.

Remark 3.5 From Theorem 3.4, we can see that the parametric parts can be estimated at the usual parametric rate of convergence. Notice that from Condition (C2), $g'(X^{\mathrm{T}}\beta_0)$ is replaced by $g'(\underline{u}_j-)$ with $\underline{u}_j = X^{\mathrm{T}}\beta_0 \in D_{2,l}$. Otherwise, in the right neighborhood $g'(X^{\mathrm{T}}\beta_0)$ is replaced by $g'(\underline{u}_j+)$. Theorem 3.4 can be proved in the same way as [16].

4 Numerical Simulations

In this section, we perform numerical experiments to assess the performance of our proposed JPLLPRE estimator. We compare our method with the jump-preserving local least squares estimate (JPLLSE for short) method, which is similar to the approach in [14], as well as the local linear product relative error (LLPRE for short) method, under various scenarios. In the simulation study, the following two examples are considered.

Example 4.1 $Y = \exp(g_1(\beta^{\mathrm{T}} X))\varepsilon^{0.3}$, where

$$g_1(u) = \begin{cases} \cos(8\pi(0.5-u)), & 0 \le u < 0.5, \\ -\cos(8\pi(0.5-u)), & 0.5 \le u \le 1. \end{cases}$$

Example 4.2 $Y = \exp(g_2(\beta^{\mathrm{T}} X))\varepsilon^{0.1}$, where

$$g_2(u) = \begin{cases} -3u+2, & 0 \le u < 0.3, \\ -3u+3-\sin\Big(\dfrac{(u-0.3)\pi}{0.2}\Big), & 0.3 \le u < 0.7, \\ \dfrac{u}{2}+1.55, & 0.7 \le u \le 1. \end{cases}$$

Let $\beta = \big(\frac{7}{9}, \frac{4}{9}, \frac{4}{9}\big)^{\mathrm{T}}$ and $X = (X_1, X_2, X_3)$, where X_1, X_2 and X_3 follow uniform distributions on $\big[0, \frac{3}{7}\big]$, $\big[0, \frac{3}{4}\big]$ and $\big[0, \frac{3}{4}\big]$, respectively. For each example, we carry out $N = 200$ simulations with sample size $n = 100, 200$. Here, we employ the Epanechnikov kernel function. In addition, the following four different error distributions are considered.

Case 1 ε follows the distribution with density function,

$$f_1(x) = c_1 \exp(-x - x^{-1} - \log x + 2) I(x > 0),$$

where c_1 is a normalization constant;

Case 2 ε follows the log-normal distribution, i.e., $\log(\varepsilon) \sim N(0,1)$;

Case 3 ε follows the log-uniform distribution on $U(-1,1)$, i.e.,

$$\log(\varepsilon) \sim U(-1,1);$$

Case 4 ε follows the uniform distribution on $U(0.5, q)$ with q being chosen such that $E(\varepsilon) = E\big(\frac{1}{\varepsilon}\big)$.

To assess the performance of the proposed method and other methods, the mean integrated square error (MISE for short) is introduced. MISE is given by

$$\mathrm{MISE} = \frac{1}{N}\sum_{t=1}^{N}\frac{1}{n}\sum_{i=1}^{n}(\widehat{g}_t(X_i^{\mathrm{T}}\widehat{\beta}) - g(X_i^{\mathrm{T}}\widehat{\beta}))^2,$$

where N denotes simulation times. $\{\widehat{g}_t(\widehat{\beta}^{\mathrm{T}}X_i), i = 1, \cdots, n\}$ are the estimate results of the tth simulated sample. Furthermore, in the neighborhood of a jump point, the approximated local MISE (MISE) is given by

$$\mathrm{MISE}_{\underline{u}_j} = \frac{1}{N}\sum_{t=1}^{N}\Big(\frac{1}{n}\sum_{i=1}^{n}(\widehat{g}_t(u_i) - g(u_i))^2 I[\underline{u}_j - 0.05 < X_i^{\mathrm{T}}\beta_0 < \underline{u}_j + 0.05]\Big),$$

where $\underline{u}_j$ denotes the jth jump point and $I(\cdot)$ is an indicator function. To assess the performance of $\widehat{\beta}$, we report the mean of empirical biases (Bias), standard deviations (SD for short) and mean absolute deviation (MAD for short), which is defined by

$$\mathrm{MAD} = \frac{1}{N}\sum_{t=1}^{N}\|\widehat{\beta}_t - \beta\|_2,$$

where $\widehat{\beta}_t$ is the estimation result of the tth simulated sample and $\|\cdot\|_2$ denotes the Euclidean norm.

Observing Tables 1–2 and Figures 1–2, we can make the following conclusions.

Table 1 The simulation results of Example 4.1.

n	ε	Method	$\widehat{\boldsymbol{\beta}}_1$		$\widehat{\boldsymbol{\beta}}_2$		$\widehat{\boldsymbol{\beta}}_3$		MAD	MISE	$\mathrm{MISE}_{0.5}$
			Bias	SD	Bias	SD	Bias	SD			
100	Case 1	JPLLPRE	0.0000	0.0081	0.0003	0.0084	-0.0003	0.0067	0.0002	0.0248	0.0128
		JPLLSE	0.0003	0.0083	0.0006	0.0081	-0.0006	0.0065	0.0002	0.0264	0.0140
		LLPRE	0.0002	0.0082	0.0005	0.0080	-0.0005	0.0066	0.0002	0.0366	0.0252
200		JPLLPRE	-0.0002	0.0056	-0.0004	0.0057	0.0003	0.0042	0.0001	0.0152	0.0078
		JPLLSE	0.0000	0.0053	-0.0003	0.0055	0.0001	0.0040	0.0001	0.0164	0.0083
		LLPRE	-0.0006	0.0061	-0.0005	0.0059	0.0005	0.0047	0.0001	0.0279	0.0211
100	Case 2	JPLLPRE	-0.0002	0.0099	0.0009	0.0096	-0.0006	0.0076	0.0002	0.0426	0.0202
		JPLLSE	-0.0003	0.0096	0.0012	0.0094	-0.0007	0.0072	0.0002	0.0400	0.0181
		LLPRE	0.0008	0.0144	0.0011	0.0102	-0.0014	0.0121	0.0005	0.0535	0.0315
200		JPLLPRE	0.0002	0.0064	0.0003	0.0067	-0.0003	0.0050	0.0001	0.0254	0.0122
		JPLLSE	0.0002	0.0068	0.0002	0.0069	-0.0003	0.0053	0.0001	0.0263	0.0128
		LLPRE	0.0004	0.0091	0.0012	0.0113	-0.0011	0.0118	0.0004	0.0366	0.0221
100	Case 3	JPLLPRE	0.0010	0.0085	0.0005	0.0083	-0.0010	0.0071	0.0002	0.0227	0.0108
		JPLLSE	0.0014	0.0101	0.0007	0.0076	-0.0013	0.0075	0.0002	0.0235	0.0115
		LLPRE	0.0010	0.0089	0.0010	0.0085	-0.0013	0.0073	0.0002	0.0384	0.0278
200		JPLLPRE	-0.0005	0.0053	-0.0002	0.0050	0.0004	0.0041	0.0001	0.0127	0.0063
		JPLLSE	-0.0003	0.0052	-0.0003	0.0050	0.0003	0.0040	0.0001	0.0127	0.0066
		LLPRE	-0.0003	0.0058	-0.0001	0.0052	0.0002	0.0042	0.0001	0.0269	0.0215
100	Case 4	JPLLPRE	-0.0002	0.0058	0.0001	0.0055	0.0000	0.0045	0.0001	0.0127	0.0068
		JPLLSE	-0.0003	0.00061	0.0010	0.0078	-0.0004	0.0051	0.0001	0.0121	0.0065
		LLPRE	0.0000	0.0068	0.0008	0.0066	-0.0005	0.0053	0.0001	0.00276	0.0226
200		JPLLPRE	-0.0002	0.0043	0.0001	0.0037	0.0000	0.0032	0.0000	0.0071	0.0035
		JPLLSE	0.0000	0.0045	0.0000	0.0042	-0.0001	0.0034	0.0000	0.0075	0.0043
		LLPRE	0.0000	0.0054	0.0000	0.0048	-0.0001	0.0041	0.0001	0.0228	0.0200

Table 2 The simulation results of Example 4.2.

n	ε	Method	$\widehat{\boldsymbol{\beta}}_1$		$\widehat{\boldsymbol{\beta}}_2$		$\widehat{\boldsymbol{\beta}}_3$		MAD	MISE	$\mathrm{MISE}_{0.3}$	$\mathrm{MISE}_{0.7}$
			Bias	SD	Bias	SD	Bias	SD				
100	Case 1	JPLLPRE	0.0017	0.0071	-0.0012	0.0091	-0.0004	0.0069	0.0002	0.0133	0.0052	0.0045
		JPLLSE	0.0043	0.0185	0.0027	0.0192	-0.0049	0.0258	0.0014	0.0153	0.0053	0.0047
		LLPRE	0.0040	0.0185	0.0019	0.0188	-0.0043	0.0258	0.0014	0.0164	0.0059	0.0061
200		JPLLPRE	-0.0006	0.0039	0.0003	0.0036	0.0001	0.0026	0.0000	0.0078	0.0034	0.0022
		JPLLSE	-0.0007	0.0040	0.0008	0.0042	-0.0001	0.0032	0.0000	0.0086	0.0033	0.0026
		LLPRE	-0.0005	0.0048	-0.0001	0.0041	0.0003	0.0038	0.0001	0.0116	0.0051	0.0052
100	Case 2	JPLLPRE	0.0026	0.0154	0.0020	0.0152	-0.0032	0.0189	0.0008	0.0171	0.0053	0.0057
		JPLLSE	0.0032	0.0154	0.0012	0.0152	-0.0030	0.0192	0.0008	0.0173	0.0054	0.0061
		LLPRE	0.0034	0.0158	0.0022	0.0144	-0.0037	0.0189	0.0008	0.0177	0.0067	0.0062
200		JPLLPRE	-0.0046	0.0228	0.0005	0.0127	0.0019	0.0057	0.0007	0.0101	0.0031	0.0034
		JPLLSE	-0.0037	0.0208	-0.0003	0.0094	0.0019	0.0062	0.0006	0.0109	0.0032	0.0035
		LLPRE	-0.0003	0.0053	0.0004	0.0047	-0.0001	0.0037	0.0001	0.0123	0.0049	0.0055
100	Case 3	JPLLPRE	0.0009	0.0199	0.0018	0.0179	-0.0024	0.0268	0.0014	0.0140	0.0053	0.0034
		JPLLSE	0.0072	0.0285	0.0064	0.0281	-0.0099	0.0403	0.0034	0.0205	0.0055	0.0046
		LLPRE	0.0071	0.0287	0.0067	0.0275	-0.0100	0.0402	0.0033	0.0205	0.0058	0.0065
200		JPLLPRE	0.0004	0.0049	0.0004	0.0043	-0.0005	0.0037	0.0001	0.0096	0.0037	0.0030
		JPLLSE	0.0001	0.0054	-0.0001	0.0048	0.0000	0.0035	0.0001	0.0098	0.0038	0.0037
		LLPRE	-0.0004	0.0050	0.0002	0.0055	0.0001	0.0045	0.0001	0.0122	0.0054	0.0054
100	Case 4	JPLLPRE	0.0075	0.0318	0.0073	0.0291	-0.0110	0.0438	0.0039	0.0178	0.0073	0.0048
		JPLLSE	0.0085	0.0327	0.0079	0.0297	-0.0121	0.0454	0.0042	0.0189	0.0075	0.0051
		LLPRE	0.0088	0.0330	0.0063	0.0263	-0.0110	0.0424	0.0037	0.0203	0.0078	0.0083
200		JPLLPRE	-0.0007	0.0032	-0.0008	0.0041	0.0008	0.0029	0.0000	0.0067	0.0031	0.0023
		JPLLSE	-0.0005	0.0029	-0.0008	0.0043	0.0007	0.0029	0.0000	0.0070	0.0032	0.0024
		LLPRE	0.0001	0.0047	-0.0005	0.0048	0.0002	0.0043	0.0001	0.0124	0.0057	0.0052

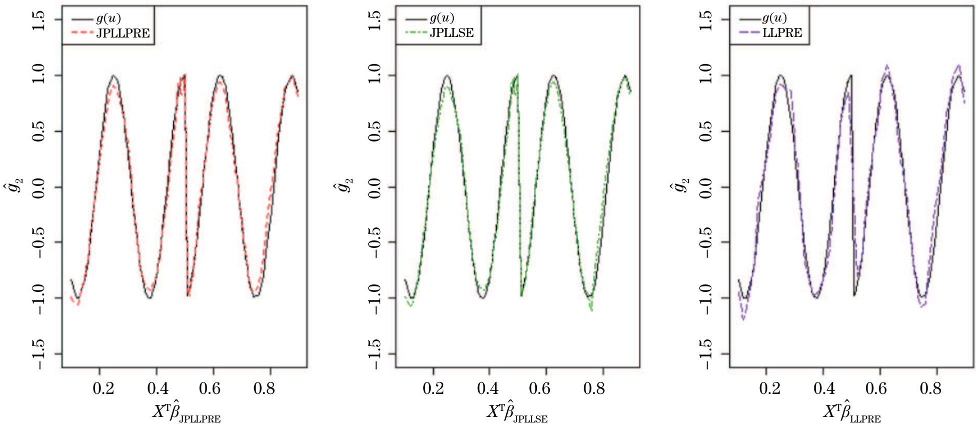

Figure 1 One typical simulation result of Example 4.1 when $n = 200$ and $\varepsilon \sim f_1$. The left panel depicts the true function (solid curve) and the JPLLPRE estimators (dashed curve). The middle panel depicts the true function (solid curve) and the JPLLSE estimators (dashed curve). The right panel depicts the true function (solid curve) and the LLPRE estimators (dashed curve).

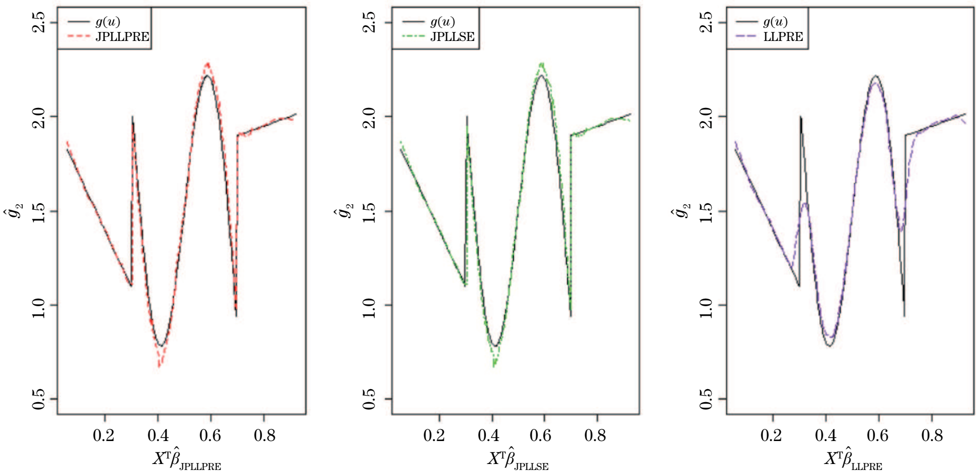

Figure 2 One typical simulation result of Example 4.2 when $n = 200$ and $\varepsilon \sim f_1$. The left panel depicts the true function (solid curve) and the JPLLPRE estimators (dashed curve). The middle panel depicts the true function (solid curve) and the JPLLSE estimators (dashed curve). The right panel depicts the true function (solid curve) and the LLPRE estimators (dashed curve).

(1) The findings shown in Table 1 demonstrate that the Bias, SD and MAD values for the parameter estimations for the three methodologies of estimation for the four different error distributions are small and not significantly distinct. Overall, it demonstrates that for estimating the parameter component, the estimate approaches JPLLPRE and JPLLSE are roughly as effective as the LLPRE approach.

(2) As can be seen from Table 1, the MISE and $\text{MISE}_{0.5}$ values for the estimation methods JPLLPRE and JPLLSE are significantly smaller than the values corresponding to the LLPRE method under the four error distributions. This result indicates that the two jump-preserving-based estimation methods are able to retain the information at the jump points well when the function is discontinuous. Also, Figure 1 provides a more visual representation of the conclusion. Furthermore, the JPLLPRE method has the smallest MISE with $\text{MISE}_{0.5}$ when the errors follow Case 1 and Case 3. This indicates that the JPLLPRE method is more effective than the JPLLSE method in such cases. When the error satisfies Case 2 and Case 4, the JPLLPRE and JPLLSE methods are estimated to be about the same.

(3) As the sample size increases, MAD, MISE and $\text{MISE}_{0.5}$ of all estimators decrease.

(4) Table 2 records the case where the function $g_2(\cdot)$ has two jump points. Similar results to those in Example 4.1 can be obtained. However, because the jump magnitude of $u = 0.3$ is larger than that of $u = 0.7$, $\text{MISE}_{0.3}$'s are larger than $\text{MISE}_{0.7}$'s. Furthermore, it is evident that the JPLLPRE estimation method outperforms others in terms of global function estimation, as it exhibits the smallest MISE. One typical simulation result is presented in Figure 2.

To sum up, the above results show that the proposed JPLLPRE works well under these different error distributions. Moreover, as the sample size increases, all approaches become more accurate. Thus, we recommend the JPLLPRE method to address the SIMM with discontinuous link function.

5 Real Data Analysis

In this section, we apply the JPLLPRE, JPLLSE and LLPRE methods to the sea-level pressure dataset, which can be found in [13]. This dataset denotes the December sea-level pressure during 1921–1992 observed by the Bombay weather station in India. Meteorologists have pointed out a jump around the year 1960. Atmospheric pressure can change as temperature varies, so the dataset also includes the annual meteorological temperatures during the same time period. [13] proposed their jump-preserving backfitting procedure in the additive models to confirm the existence of this jump. We use this dataset to show the performance of the proposed approach.

In this dataset, we consider the year (X_1) and temperature (X_2) as the independent variables, and pressure (Y) as the dependent variable. We apply the Epanechnikov kernel function for the analysis. In this study, we focus on the following single-index multiplicative model

$$y = \exp(g(\beta_1 X_1 + \beta_2 X_2))\varepsilon.$$

The estimates of the index coefficient for the three estimation methods are $\widehat{\beta}_{\text{JPLLPRE}} = (0.0275, 0.9996)$, $\widehat{\beta}_{\text{JPLLSE}} = (0.0254, 0.9997)$ and $\widehat{\beta}_{\text{LLPRE}} = (0.0163, 0.9999)$, respectively. All estimation approaches indicate a highly significant influence of temperature on sea-level pressure. Furthermore, the disparities among the three estimators are not considerable. From Figure 3(a), it can be observed that the fitting curve obtained by the JPLLPRE method is divided into four segments, indicating the presence of three potential jump points. The positions of these jump points are at $X^{\mathrm{T}}\widehat{\beta}_{\mathrm{JPLLPRE}} = 80.8277, 81.0706, 81.7657$, corresponding to the years 1943, 1954, and 1960, respectively. Moving to Figure 3(b), it is evident that the JPLLSE method identifies two jump points at $X^{\mathrm{T}}\widehat{\beta}_{\mathrm{JPLLSE}} = 76.8696, 77.5518$, corresponding to the years 1954 and 1960. Finally, Figure 3(c) reveals that the function fitted by the LLPRE method does not exhibit any discontinuity, indicating that the LLPRE method blurs potential jump points. In contrast, both the JPLLPRE method and the JPLLSE method effectively retain information about the jump points. The research findings from [13] indicate the existence of jump points in the years 1943 and 1960. Our proposed method's research results similarly affirm this conclusion and offer dependable evidence to suggest that the year 1954 is also a possible jump point.

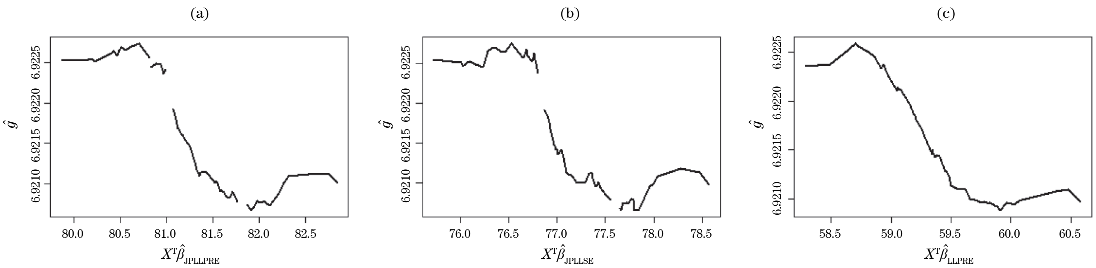

Figure 3 Sea-level pressure dataset. (a)–(c): The estimated link function of JPLLPRE, JPLLSE and LLPRE methods.

6 Discussion

In this paper, we have investigated a broader class of single-index multiplicative models with an unknown discontinuous link function, specifically designed for analyzing data with positive response. To address the challenges associated with this class of models, we have proposed a jump-preserving estimation approach based on the local least product relative error method. This approach effectively estimates the nonparametric link function in SIMM while preserving the discontinuities. We have derived the asymptotic properties of the proposed estimators under certain moderate conditions, which are weaker than the conditions stated in the work of [16]. This finding expands the applicability of SIMM by accommodating discontinuous link functions, making a significant contribution to the field. Through extensive experiments using both simulated and real data, we have demonstrated the reliability and effectiveness of the proposed estimators. However, it is worth noting that the local linear kernel smooth approach utilized in the jump-preserving LPRE method may have limitations in terms of computational efficiency. Therefore, further research is necessary to develop more efficient estimation approaches that can enhance the practical applicability of our proposed approach. By addressing the challenge of estimating link functions in SIMM, our study advances the understanding and application of these models in diverse contexts.

Acknowledgements The authors extend their appreciation to the anonymous reviewers for their insightful suggestions and comments, which significantly contributed to enhancing the quality of the initial draft.

Declarations

Conflicts of interest The authors declare no conflicts of interest.

References

[1] Cao, G. Q., Wang, Z. F., Wu, Y. H. and Zhao, L. C., Inference of change-point in single index models, *Sci. China Ser. A*, **51**, 2008, 1855–1870.

[2] Chen, K. N., Guo, S. J., Lin, Y. Y. and Ying, Z. L., Least absolute relative error estimation, *J. Amer. Statist. Assoc.*, **105**, 2010, 1104–1112.

[3] Chen, K. N., Lin, Y. Y., Wang, Z. F. and Ying, Z. L., Least product relative error estimation, *J. Multivariate Anal.*, **144**, 2016, 91–98.

[4] Chen, Y. J., Liu, H. L. and Ma, J. J., Local least product relative error estimation for single-index varying-coefficient multiplicative model with positive responses, *J. Comput. Appl. Math.*, **415**, 2022, 114478, 17 pp.

[5] Desmet, L. and Gijbels, I., Curve fitting under jump and peak irregularities using local linear regression, *Comm. Statist. Theory Methods*, **40**, 2011, 4001–4020.

[6] Desmet, L., Gijbels, I. and Lambert, A., Estimation of irregular probability densities, *Inst. Math. Stat. (IMS) Collect.*, **7**, 2010, 46–61.

[7] Gijbels, I., Lambert, A. and Qiu, P. H., Jump-preserving regression and smoothing using local linear fitting: A compromise, *Ann. Inst. Statist. Math.*, **59**, 2007, 235–272.

[8] Han, Z. C., Lin, J. G., Wang, H. X. and Huang, X. F., A robust and efficient estimation method for nonparametric models with jump points, *Comm. Statist. Simulation Comput.*, **46**, 2017, 6283–6297.

[9] Han, Z. C., Lin, J. G. and Zhao, Y. Y., Adaptive semiparametric estimation for single index models with jumps, *Comput. Statist. Data Anal.*, **151**, 2020, 107013, 12 pp.

[10] Jin, B. S., Dong, C. L., Tan, C. C. and Miao, B. Q., Estimator of a change point in single index models, *Sci. China Math.*, **57**, 2014, 1701–1712.

[11] Joo, J. H. and Qiu, P. H., Jump detection in a regression curve and its derivative, *Technometrics*, **51**, 2009, 289–305.

[12] Kang, K. H., Koo, J. Y. and Park, C. W., Kernel estimation of discontinuous regression functions, *Statist. Probab. Lett.*, **47**, 2000, 277–285.

[13] Kang, Y. C., Shi, Y. Y., Jiao, Y. L., et al., Fitting jump additive models, *Comput. Statist. Data Anal.*, **162**, 2021, 107266, 12 pp.

[14] Liu, G. X., Du, X. L., Wang, M. M., et al., Semiparametric jump-preserving estimation for single-index models, *J. Nonparametr. Stat.*, **30**, 2018, 556–580.

[15] Liu, G. X., Wang, M. M., Du, X. L., et al., Jump-detection and curve estimation methods for discontinuous regression functions based on the piecewise B-spline function, *Comm. Statist. Theory Methods*, **47**, 2018, 5729–5749.

[16] Liu, H. L. and Xia, X. C., Estimation and empirical likelihood for single-index multiplicative models, *J. Statist. Plann. Inference*, **193**, 2018, 70–88.

[17] McDonald, J. A. and Owen, A. B., Smoothing with split linear fits, *Technometrics*, **28**, 1986, 195–208.

[18] Qing, Y. and Zhang, Y., Change-point detection for the link function in a single-index model, *Statist. Probab. Lett.*, **186**, 2022, 109468, 10 pp.

[19] Qiu, P. H., A jump-preserving curve fitting procedure based on local piecewise-linear kernel estimation, *J. Nonparametr. Stat.*, **15**, 2003, 437–453.

[20] Qiu, P. H., Asano, C. and Li, X. P., Estimation of jump regression functions, *Bull. Inform. Cybernet.*, **24**, 1991, 197–212.

[21] Qiu, P. H. and Yandell, B., A local polynomial jump detection algorithm in nonparametric regression, *Technometrics*, **40**, 1998, 141–152.

[22] Wang, Z. F., Chen, Z. M. and Wu, Y. H., A relative error estimation approach for multiplicative single index model, *J. Syst. Sci. Complex.*, **30**, 2017, 1160–1172.

[23] Wang, Z. F., Liu, W. X. and Lin, Y. Y., A change-point problem in relative error-based regression, *TEST*, **24**, 2015, 835–856.

[24] Wu, J. S. and Chu, C. K., Kernel type estimators of jump points and values of a regression function, *Ann. Statist.*, **21**, 1993, 1545–1566.

[25] Zhang, J., Zhu, J. P. and Feng, Z. H., Estimation and hypothesis test for single-index multiplicative models, *TEST*, **28**, 2019, 242–268.

[26] Zhao, Y. Y. and Lin, J. G., Estimation and test of jump discontinuities in varying-coefficient models with empirical applications, *Comput. Statist. Data Anal.*, **139**, 2019, 145–163.

[27] Zhao, Y. Y., Lin, J. G., Huang, X. F. and Wang, H. X., Adaptive jump-preserving estimates in varying-coefficient models, *J. Multivariate Anal.*, **149**, 2016, 65–80.

[28] Zhao, Y. Y., Lin, J. G., Wang, H. X. and Huang, X. F., Jump-detection-based estimation in time-varying coefficient models and empirical applications, *TEST*, **26**, 2017, 574–599.