
Research article**Nonparametric estimation of coefficient function derivatives in varying coefficient models****Junfeng Huo^{1,*}Mingquan Wang² and Xiuqing Zhou²**¹ School of Mathematics and Statistics, Wuhan University, Wuhan 430072, China² School of Mathematical Sciences, Nanjing Normal University, Nanjing 210023, China*** Correspondence:** Email: huojf1998@163.com.

Abstract: In varying-coefficient models, accurately estimating the derivatives of coefficient functions is crucial, particularly for optimal bandwidth selection and confidence interval construction. Despite its importance, methods have largely ignored derivative estimation. In this paper, we address this gap by proposing a novel weighted difference quotient approach to estimate both first and second order derivatives of coefficient functions under the differences in the smoothness of the coefficient functions. We derived the asymptotic properties of our estimator and introduced a data-driven tuning parameter selection method. Simulations and real-data analyses confirmed the superior performance of our approach compared to existing techniques. Our method provides the most accurate estimates, effectively estimating both the first and second derivatives.

Keywords: varying coefficient model; derivative estimation; asymptotic properties; symmetric quotients method; kernel method

Mathematics Subject Classification: 62F12, 62G05, 62H12

1. Introduction

As an important extension of nonparametric models, varying-coefficient models have evolved through three distinct methodological epochs since their conceptual inception by Shumway [1]. This framework fundamentally reimagines traditional parametric specifications by enabling covariate effects to fit the effects of predictors that change across levels of other variables, making them particularly useful in scenarios with heterogeneous data or when relationships between variables are not constant. Many statisticians proposed some effective methods to estimate it; see, e.g., Hastie and Tibshirani [2] for the smooth spline method, Fan and Zhang [3] for the two-step method, Huang and Shen [4] for the polynomial spline method, Cai et al. [5] for the local kernel smoothing method.

The derivative estimation of the coefficient functions is equally crucial. The first derivative of

coefficient functions precisely characterizes the instantaneous rate of change at any given point and facilitates the detection of inflection points of the coefficient function, while enabling the analysis of marginal effect in economic [6]. Crucially, the second derivative constitutes the foundation for statistical inference on coefficient functions [5], and its proper estimation is essential for constructing valid confidence intervals. However, a large number of researchers have concentrated predominantly on the estimation of the coefficient functions itself, inadvertently causing the derivative functions of the coefficient functions to languish in relative obscurity. Within the framework of the varying-coefficient models, the estimation of derivative functions has been restricted to the local polynomial method [7] and the smoothing spline method [8]. These methods require the selection of smoothing parameters. However, existing parameter selection approaches yield only satisfactory estimation for the coefficient functions, while producing suboptimal results for derivative function estimation under the same parameters. An intuitive explanation for the failure of smoothing parameters is that coefficient functions and their derivatives often have different smoothness levels. The selected smoothing parameters meet only requirements of the coefficient functions, resulting in poor estimation performance for the derivative functions. Together, these considerations require the development of novel estimation methods specifically designed for derivative functions of coefficient functions.

We now shift our focus to a special case of varying-coefficient models: nonparametric models. Within the framework of nonparametric models, researchers have developed a diverse array of derivative estimation methods that have significantly advanced the field. Consequently, we can draw upon these estimation techniques to devise a novel approach to estimate derivative functions of coefficient functions. In the realm of nonparametric models, local polynomial regression [9, 10] and smoothing spline [11] remain the predominant methods for derivative estimation. Although both methodologies necessitate the selection of a smoothing parameter, the challenge of optimally choosing this parameter remains inadequately resolved. Furthermore, Wang and Zhou [12] employed the Bayesian inference framework to estimate derivative functions under the assumption that the unknown function follows a Gaussian process. In addition, Liu and Li [13] proposed a plug-in kernel ridge regression estimator to estimate the derivative function.

Unlike these approaches, the difference-quotient method offers an alternative avenue for derivative estimation in nonparametric models. Müller et al. [14] introduced derivative estimators based on difference quotients. However, these estimators are characterized by high variance, which complicates the smoothing process. To mitigate this issue, Charnigo et al. [15] proposed a linear combination to reduce the variance of symmetric quotients, based on the assumption of an equispaced design between independent variables. However, the random variation inherent in independent variables within random design settings often invalidates theoretical conclusions derived from equispaced design assumptions, limiting their applicability. In response to this, several researchers have developed novel procedures in the equispaced design framework, which have produced promising results [16–18]. In pursuit of a symmetric difference quotient applicable under random design conditions, Liu and Brabanter [19] elucidated the theoretical properties of derivative estimation utilizing weighted symmetric quotients, assuming that independent variables follow a standard uniform distribution. Empirical results demonstrate the superiority of their method over conventional local polynomial and spline approaches. Furthermore, Liu and Kong [20] constructed empirical derivatives using weighted symmetric difference quotients to estimate derivatives under the condition of error correlation. In view of the great estimation performance of the weighted symmetric quotients procedure in nonparametric models, we can draw on

the weighted symmetric quotient method to estimate the derivatives of coefficient functions.

In this paper, we consider the estimation of the derivatives in a varying-coefficient model where there are differences in the smoothness of the coefficient functions. First, based on the weighted symmetric difference quotients idea, an effective method for estimating the first and second derivatives of the coefficient functions is proposed under the case where the index variables are a uniform random design. Compared to traditional local kernel estimation and spline methods that cannot select appropriate smoothing parameters, we propose an effective approach to choosing the tuning parameters in our estimator, which leads to superior estimation performance over conventional methods. Then, we derive the corresponding theoretical properties and generalize the estimators to the case that index variables are arbitrary distributions.

The paper is organized as follows. In Section 2, we show the first-order derivative estimation of coefficient functions based on variance reducing weighted difference quotients and establish the theoretical properties of conditional bias, variance. In addition, a method for selecting the optimal tuning parameter is given. In Section 3, we extend the procedure to second-order derivatives of the coefficient. In Sections 4 and 5, we perform Monte Carlo experiments and real data analysis to compare the proposed methodology with the local polynomial regression and the two-step method. Some conclusions are presented in Section 6, and the proof details of the theorems are provided in Appendix.

2. First order derivative estimation

2.1. Estimation approach of first order derivative

From now on, let $\{(Y_i; U_i, X_{i1}, \dots, X_{ip}), i = 1, \dots, n\}$ be a set of independent random samples from a population (\mathbf{X}, U, Y) ; then, we consider the varying coefficient model

$$Y = \sum_{j=1}^p a_j(U)X_j + \varepsilon, \quad (2.1)$$

where Y is response variable, X_1, \dots, X_p are covariate variables, U is index variable, $a_j(\cdot)$ ($j = 1, \dots, p$) are unknown coefficient functions, and ε is the regression error which satisfies $E(\varepsilon|U, X_1, \dots, X_p) = 0$ and $Var(\varepsilon|U, X_1, \dots, X_p) = \sigma^2$.

Now, we consider the special case where the index variable U follows a uniform distribution. Let $U \sim \mathcal{U}(0, 1)$, where $\mathcal{U}(0, 1)$ is the standard uniform distribution between 0 and 1. In addition, let $U_{(i)}, i = 1, 2, \dots, n$ be the order statistics of U and $\hat{a}_q(\cdot), q = 1, 2, \dots, p$ be a two-step estimator [3] of the coefficient functions which are obtained by using local linear fitting in the second step. Inspired by the success of symmetric difference quotient methods [18–20] in nonparametric regression for derivative estimation, we can construct weighted symmetric difference quotient estimators to estimate the derivative of $a_j(\cdot)$. Without loss of generality, we consider the derivative estimate of $a_p(\cdot)$. If we want to estimate the derivative of $a_q(\cdot)$, we can swap the position of $a_q(\cdot)$ with $a_p(\cdot)$. Therefore, the proposed weighted symmetric quotient estimators of first order derivative estimator for uniform order statistics is

$$\hat{a}_p^{(1)}(U_{(i)}) = \sum_{j=1}^k \omega_{i,j} \left(\frac{\hat{a}_p(U_{(i+j)}) - \hat{a}_p(U_{(i-j)})}{U_{(i+j)} - U_{(i-j)}} \right), \quad (2.2)$$

where $\hat{a}_p^{(1)}(U_{(i)})$ denotes the derivative estimator in $U_{(i)}$, and $\omega_{i,j}$ denote weights and satisfy $\omega_{i,j} \geq 0$, $\sum_{j=1}^k \omega_{i,j} = 1$, then k is a given turning parameter.

The idea behind the symmetric quotient estimator in Equation (2.2) mainly originates from the definition of a derivative. Given a point u_0 , the first-order derivative $a_1^{(1)}(u_0) = \lim_{\delta \rightarrow 0} \frac{a_1(u_0+\delta) - a_1(u_0-\delta)}{2\delta}$. Based on the definition, given a point $U_{(i)}$, we can naturally derive the following estimator for the first order derivative

$$\hat{a}_p^{(1)}(U_{(i)}) = \frac{\hat{a}_p(U_{(i+1)}) - \hat{a}_p(U_{(i-1)})}{U_{(i+1)} - U_{(i-1)}}, \quad (2.3)$$

(2.3) represents the special case of (2.2) when $k = 1$. However, in this scenario, the estimator utilizes only information from points $U_{(i-1)}$ and $U_{(i+1)}$, making it highly susceptible to noise interference. To mitigate this, we require additional estimators analogous to the derivative definition, denoted $\frac{\hat{a}_p(U_{(i+j)}) - \hat{a}_p(U_{(i-j)})}{U_{(i+j)} - U_{(i-j)}}$, and obtain a robust derivative estimator through weighted averaging, resulting in (2.2). In particular, k in this context plays a role similar to the smoothing parameter in the estimation of the local polynomial: when k is too small, excessive noise is introduced; When k is too large, the estimator becomes overly smoothed. We conducted an experiment to illustrate the impact of varying k on the estimation of the first-order derivative, the results of which are visualized in Figure 1. As shown in Figure 1, when $k = 1$, the estimator shows significant fluctuations due to noise interference. In contrast, when $k = 120$, excessive smoothing leads to poor estimation near the function's peak. Only when $k = 50$ does the estimator achieve satisfactory performance. Therefore, the choice of k is critical in determining the quality of the estimator.

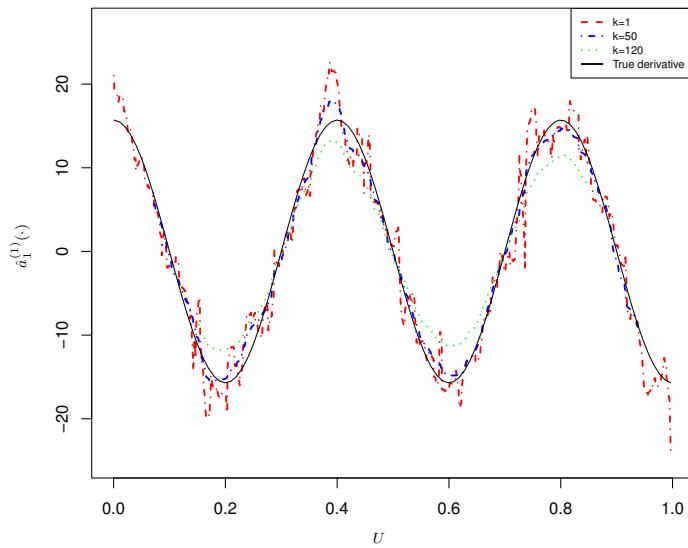


Figure 1. The estimated first order functions of $\hat{a}_1^{(1)}(\cdot)$. The samples $Y = \sin(5\pi U)X_1 + \sin(2\pi U)X_2 + \varepsilon$, U follows a uniform distribution on $[0, 1]$, $(X_1, X_2) \sim N(\mu, \Sigma)$, where $\mu = (0, 0)^T$, $\Sigma = \{(0)^{|i-j|}\}_{1 \leq i, j \leq 2}$, $\varepsilon \sim N(0, 0.1)$. The sample sizes are $n = 800$. The parameter $h_1 = 0.05$, $h_2 = 0.03$, $k = 1, 50, 120$, $\omega_{i,j} = 1/k$. The red dashed curves: The results of $k = 1$; the blue dot-dashed curves: The results of $k = 50$; the green dotted curves: The results of $k = 120$; the black solid curves: True derivative function.

Furthermore, we employ the two-step estimator instead of the local polynomial estimator, motivated by potential differences in the smoothness of the coefficient functions in (2.2). When there is a difference in the smoothness of the coefficient functions, local polynomial estimation may struggle to provide accurate estimates, resulting in a substantial difference between the estimated and true functions. This considerable estimation error can propagate to the estimation of (2.2), thereby hindering the accuracy of the derivative function estimates. However, according to Fan and Zhang [3], the two-step method effectively enhances the accuracy of the coefficient function estimation when their smoothness varies. Consequently, we adopt the two-step method to obtain the estimator $\hat{a}_p(\cdot)$.

In (2.2), the weights $\omega_{i,j}$ are unknown. Before delving into the analysis of these weights $\omega_{i,j}$, we first consider the conditional bias and conditional variance order of $\hat{a}_p^{(1)}(U_{(i)})$ when $j = 1$. Now, we present the following assumptions.

- (A1) $E(X_j^{2s}) < \infty$ holds for $s > 2$ and $j = 1, \dots, p$.
- (A2) $nh_2^4 > n^{\rho_1}$, for some $\rho_1 > 0$. $n^{-\rho_2}h_2^{-1} = o_p(h_2^2)$, for some $\rho_2 > 0$.
- (A3) $nh_1^\gamma/\log(h_1) \rightarrow \infty$, for any $\gamma > s/(s-2)$.
- (A4) The function $K(t)$ is a finite symmetric density function with compact support.
- (A5) The function $r_{kj}(u) = E(X_k X_j | U = u)$ has a bounded second derivative in a neighborhood of u_0 and $r_{kj}^{(2)}(u_0) \neq 0$.
- (A6) The coefficient $a_j(\cdot)$ has a continuous second derivative.

Assumptions (A1) and (A3)-(A6) are often used in the varying-coefficient model literature [3, 5, 21]. Assumption (A2) imposes specific requirements on the convergence rate of h_2 , and generally the condition holds as long as $h_2 = o(n^{-0.1})$ is satisfied for some $\rho_1 > 0$ and $\rho_2 > 0$.

Proposition 1. *Under the assumptions (A1)-(A6) and $h_1 \rightarrow 0, h_2 \rightarrow 0, k \rightarrow \infty$ as $n \rightarrow \infty$, such that $nh_1 \rightarrow \infty, h_1/h_2 \rightarrow 0$, second derivative $a_j^{(2)}(\cdot)$ is Lipschitz continuous, then let*

$$\mathcal{D} = (U_1, \dots, U_n, X_{11}, \dots, X_{1n}, \dots, X_{p1}, \dots, X_{pn})^T,$$

thus the asymptotic order of condition bias and condition variance of $\hat{a}_p^{(1)}(U_{(i)})$ for $j = 1$ are given by

$$Bias[\hat{a}_p^{(1)}(U_{(i)})|\mathcal{D}] = o_p(1), \quad (2.4)$$

$$Var[\hat{a}_p^{(1)}(U_{(i)})|\mathcal{D}] = O_p\left(\frac{n}{h_1}\right). \quad (2.5)$$

According to Proposition 1, under $j = 1$, as n grows large, the bias term asymptotically converges to zero, while the variance of the estimator grows without bound, resulting in an unreliable estimate. To mitigate this risk of uncontrolled variance, the proposed weighting scheme effectively controls the variance of the parameter $\hat{a}_p^{(1)}(U_{(i)})$. The following proposition provides a method to determine the optimal weights $\omega_{i,j}$.

Proposition 2. *For $k+1 \leq i \leq n-k$, the weights $\omega_{i,j}$ that minimize an upper bound of conditional variance of (2.2), satisfying $\sum_{j=1}^k \omega_{i,j} = 1$, are given by*

$$\omega_{i,j} = \frac{(U_{(i+j)} - U_{(i-j)})^2}{\sum_{c=1}^k (U_{(i+c)} - U_{(i-c)})^2}. \quad (2.6)$$

Note that optimal weights are obtained by minimizing an upper bound of the conditional variance of (2.2) rather than minimizing the conditional variance. However, in the next section, it is shown that under the optimal weights given by Proposition 2, the conditional variance can be controlled with appropriate parameter settings, and the performance of our proposed estimator is better than the method in simulation.

Remark 1. Our estimator is formulated under the assumption that the index variable follows a standard uniform distribution. In particular, this simplifying assumption aligns with established practices in the literature on varying-coefficient modeling [21–23], where researchers often impose uniformity on the index variable for empirical implementation. Although this convention facilitates tractable analysis, we explicitly address its limitations in the following sections by deriving derivative estimates when the variable follows nonuniform distributions.

2.2. Asymptotic properties of the first-order derivative estimator

We establish the upper bounds of asymptotic conditional bias and variance of our proposed estimator (2.2) for the interior points $k + 1 \leq i \leq n - k$. Now, we will use the following notation. Define

$$\mu_m = \int K(t)t^m dt, \quad M = \sup_t K(t),$$

and

$$\begin{aligned} r_{pp}^- &= \inf_{U \in [0,1]} E(X_p^2|U), \\ r_{dp}^+ &= \sup_{U \in [0,1]} |E(X_d X_p|U)|, \\ r_{dp}^* &= \sup_{U \in [0,1]} |e_{d,p}^T (E(\mathbf{X} \mathbf{X}^T|U))^{-1} e_{p,p}|, \end{aligned}$$

for $d = 1, \dots, p - 1$.

Theorem 1. Under assumptions (A1)-(A6) and $h_1 \rightarrow 0, h_2 \rightarrow 0, k \rightarrow \infty$ as $n \rightarrow \infty$, such that $nh_1 \rightarrow \infty, h_1/h_2 \rightarrow 0, k^2/n^2h_2 \rightarrow 0$, the upper bounds of conditional bias and the conditional variance of $\hat{a}_p(U_{(i)})$ are given by

$$\begin{aligned} \left| \text{Bias} \left[\hat{a}_p^{(1)}(U_{(i)}) | \mathcal{D} \right] \right| &\leq \sup_{u \in [0,1]} |a_p^{(2)}(u)| \\ &\times \left\{ \frac{3k(k+1)}{4(n+1)(2k+1)} + h_2^2 \mu_2 \frac{3(n+1)}{2(2k+1)} \right\} (1 + o_p(1)) \end{aligned} \quad (2.7)$$

and

$$\text{Var} \left[\hat{a}_p^{(1)}(U_{(i)}) | \mathcal{D} \right] \leq \frac{9(n+1)^2 M \sigma^2}{(2k+1)^2 n r_{pp}^-} \left\{ \frac{1}{h_2} + \frac{1}{h_1} \sum_{d=1}^{p-1} r_{dp}^+ r_{dp}^* \right\} (1 + o_p(1)). \quad (2.8)$$

As demonstrated in Theorem 1, we establish asymptotic upper bounds for conditional bias and conditional variance. This corollary provides a precise asymptotic order for conditional bias and conditional variance, thereby characterizing their exact rates of convergence.

Corollary 1. Under assumptions (A1)-(A6) and $h_1 \rightarrow 0, h_2 \rightarrow 0, k \rightarrow \infty$ as $n \rightarrow \infty$, such that $nh_1 \rightarrow \infty, h_1/h_2 \rightarrow 0, k^2/n^2h_2 \rightarrow 0$, then the exact asymptotic order for conditional bias and conditional variance of $\hat{a}_p^{(1)}(U_{(i)})$ are given by

$$Bias\left[\hat{a}_p^{(1)}(U_{(i)})|\mathcal{D}\right] = O_p\left(\max\left\{h_2^2, \frac{k}{n}\right\}\right), \quad (2.9)$$

$$Var\left[\hat{a}_p^{(1)}(U_{(i)})|\mathcal{D}\right] = O_p\left(\frac{n}{k^2h_1}\right). \quad (2.10)$$

According to Corollary 1, we can obtain

$$\begin{aligned} MSE\left[\hat{a}_p^{(1)}(U_{(i)})|\mathcal{D}\right] &= Bias^2\left[\hat{a}_p^{(1)}(U_{(i)})|\mathcal{D}\right] + Var\left[\hat{a}_p^{(1)}(U_{(i)})|\mathcal{D}\right] \\ &= O_p\left(\max\left\{h_2^4, \frac{k^2}{n^2}, \frac{n}{k^2h_1}\right\}\right). \end{aligned} \quad (2.11)$$

Now, we compare the MSE of this method with those of the two-step method and the local polynomial method. According to the proof of Fan and Zhang's theorem [3], when $a_j(\cdot)$ has a continuous second order derivative, we can obtain the convergence rates of the bias and variance of the estimated first-order derivative, where the local linear fitting is employed in the second step, thus $Bias\left[\hat{a}_{tse,p}^{(1)}(U_{(i)})|\mathcal{D}\right] = O_p\left(\max\left\{h_{tse,2}^2, \frac{h_{tse,1}^2}{h_{tse,2}}\right\}\right)$ and $Var\left[\hat{a}_{tse,p}^{(1)}(U_{(i)})|\mathcal{D}\right] = O_p\left(\frac{1}{nh_{tse,2}^3}\right)$, then $MSE\left[\hat{a}_{tse,p}^{(1)}(U_{(i)})|\mathcal{D}\right] = O_p\left(\max\left\{h_{tse,2}^4, \frac{h_{tse,1}^4}{h_{tse,2}^2}, \frac{1}{nh_{tse,2}^3}\right\}\right)$. If we select $h_2 = O(n^{-0.13}), h_1 = O(n^{-0.14}), h_{tse,2} = O(n^{-0.2})$ and $h_{tse,1} = O(n^{0.21})$, we have $MSE\left[\hat{a}_{tse,p}^{(1)}(U_{(i)})|\mathcal{D}\right] = O_p(n^{-0.4})$. Thus, we need only select $k = O(n^{-l})$ ($0.77 \leq l \leq 0.8$), then $MSE\left[\hat{a}_p^{(1)}(U_{(i)})|\mathcal{D}\right] = o_p(n^{-0.4})$. Notice that $h_{tse,2} = O(n^{-0.2})$ is optimal to minimize the mean squared error (MSE) of the coefficient function in the two-step estimation method. Generally, existing bandwidth selection methods aim to maximize the accuracy of the coefficient function estimation. Therefore, our choice of $h_{tse,2} = O(n^{-0.2})$ is theoretically justified. Therefore, when the parameters k, h_1 , and h_2 are appropriately chosen, the convergence rate of $MSE\left[\hat{a}_p^{(1)}(U_{(i)})|\mathcal{D}\right]$ is superior to that of $MSE\left[\hat{a}_{tse,p}^{(1)}(U_{(i)})|\mathcal{D}\right]$.

For local polynomial estimation based on linear expansion, $Bias\left[\hat{a}_{lpe,p}^{(1)}(U_{(i)})|\mathcal{D}\right] = O_p(h^2)$ and $Var\left[\hat{a}_{lpe,p}^{(1)}(U_{(i)})|\mathcal{D}\right] = O_p\left(\frac{1}{nh^3}\right)$, $MSE\left[\hat{a}_{lpe,p}^{(1)}(U_{(i)})|\mathcal{D}\right] = O_p\left(\max\left\{h^4, \frac{1}{nh^3}\right\}\right)$. If we select $h = O(n^{-0.2}), h_2 = O(n^{-0.13})$ and $h_1 = O(n^{-0.14})$, we have $MSE\left[\hat{a}_{lpe,p}^{(1)}(U_{(i)})|\mathcal{D}\right] = O_p(n^{-0.4})$. Then, by appropriately choosing $k = O(n^{-l})$ ($0.77 \leq l \leq 0.8$), we can ensure $MSE\left[\hat{a}_p^{(1)}(U_{(i)})|\mathcal{D}\right] = o_p(n^{-0.4})$. The bandwidth $h = O(n^{-0.2})$ also follows the optimal bandwidth that minimizes the mean squared error (MSE) of the coefficient function in the local polynomial framework.

We complete the theoretical analysis of $\hat{a}_p^{(1)}(U_{(i)})$, but the method for obtaining the parameter k remains unknown in (2.2). According to the analysis in Section 2.1, parameter k controls the estimation performance. Notice that the MSE measures the estimation performance of an estimator and controls the bias-variance trade-off of the estimator. Therefore, we can minimize the upper bound of the conditional MSE to choose the optimal turning parameter k .

Corollary 2. Under the assumptions of Theorem 1, the optimal k is obtained by minimizing the

following formula

$$k_{opt} = \arg \min_{k \in N^+} \left\{ \left(\sup_{u \in [0,1]} |a_p^{(2)}(u)| \left\{ \frac{3k(k+1)}{2(n+1)(2k+1)} + h_2^2 \mu_2 \frac{3(n+1)}{4(2k+1)} \right\} \right) \right. \\ \left. + \frac{9(n+1)^2 M \sigma^2}{(2k+1)^2 n r_{pp}^-} \left\{ \frac{1}{h_2} + \frac{1}{h_1} \sum_{d=1}^{p-1} r_{dp}^+ r_{dp}^* \right\} \right\}. \quad (2.12)$$

Although a fast and easy parameter selection method is provided in Corollary 2, some unknown quantities need to be estimated. The error variance can be estimated by

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \sum_{q=1}^p \hat{a}_q(U_i) X_{iq})^2.$$

The $\sup_{u \in [0,1]} |a_p^{(2)}(u)|$ can be estimated using the two-step method with a local polynomial expansion of 3 order in the second step. The $E(X_d X_p | U)$ can be obtained based on observation $(U_i, X_{id} X_{ip})$ using the local linear fitting with bandwidth that is selected by cross-validation. Then, $r_{pp}^-, r_{dp}^+, r_{dp}^*$ can be calculated by taking the maximum or minimum of the fitting values of the corresponding $E(X_d X_q | U)$. By adding the estimators above, the optimal value k_{opt} can be obtained by searching the grid over the set of integers $[1, \lfloor \frac{n-1}{2} \rfloor]$, where $\lfloor \cdot \rfloor$ denotes round down.

Due to limitations in these smoothing parameter selection schemes, the first order derivatives of the coefficient functions cannot be reliably estimated for local polynomial estimation, two-step methods, and spline estimation approaches. In Corollary 2, we propose an effective approach for selecting the parameter k . Subsequent simulation studies demonstrate the validity of our parameter selection method, showing that it yields derivative estimates with superior accuracy compared to alternative approaches. Furthermore, it should be noted that the estimation of the derivatives also requires the estimation of $\hat{a}_p(U_i)$, which involves the selection of smoothing parameters h_1 and h_2 . For this purpose, cross-validation methods are sufficient, as they adequately meet the requirement of producing a good estimate of $\hat{a}_p(U_i)$. Our approach needs only ensure the proper estimation of $\hat{a}_p(U_i)$, a requirement that standard cross-validation techniques can satisfy fully.

2.3. Smoothing the estimators and generalizing results and arbitrary distributions

In the previous discussion, we were only able to obtain the first-order derivative estimate at point U_i , $i = k+1, \dots, p-k$, and estimates at other points could not be obtained using the current method. Within the framework of univariate nonparametric estimation, by treating U_i as the covariate variable and $\hat{a}_p(U_i)$ as the response variable, we can apply nonparametric estimation techniques to fit the first-order derivative, thus estimating the first-order derivative at other points. According to the local polynomial estimation for nonparametric regression [7], given an arbitrary point $u_0 \in [0, 1]$ and kernel function $K(\cdot)$, we can obtain the derivative estimator through the following equation,

$$\tilde{a}_p^{(1)}(u_0) = e_{1,g}^T \left(\mathbf{U}^T(u_0) \mathbf{W}(u_0) \mathbf{U}(u_0) \right)^{-1} \mathbf{U}^T(u_0) \mathbf{W}(u_0) \hat{\mathbf{a}}_p^{(1)}, \quad (2.13)$$

where g represents the order of local polynomial fitting, $e_{1,g}$ represents a column vector of length g , and the first element is 1, and the remaining elements are 0,

$$\mathbf{W}(u_0) = \text{Diag}(K_h(U_{(k+1)} - u_0), \dots, K_h(U_{(n-k)} - u_0)), \quad K_h(\cdot) = K(\cdot)/h,$$

$$\mathbf{U}(u_0) = \begin{pmatrix} 1 & U_{(k+1)} - u_0 & \cdots & (U_{(k+1)} - u_0)^p \\ 1 & U_{(k+2)} - u_0 & \cdots & (U_{(k+2)} - u_0)^p \\ \vdots & \vdots & & \vdots \\ 1 & U_{(n-k)} - u_0 & \cdots & (U_{(n-k)} - u_0)^p \end{pmatrix}, \hat{\mathbf{a}}_p^{(1)} = \begin{pmatrix} \hat{a}_p^{(1)}(U_{(k+1)}) \\ \hat{a}_p^{(1)}(U_{(k+2)}) \\ \vdots \\ \hat{a}_p^{(1)}(U_{(n-k)}) \end{pmatrix}.$$

Note that u_0 is a predetermined point. Immediately after specifying u_0 and the order statistics $U_{(i)}$, we can directly derive $\mathbf{U}(u_0)$ and $\mathbf{W}(u_0)$.

However, in estimating local polynomials, a smoothing parameter h must be determined. Here, we use the plug-in bandwidth selector [10], which is a simple and effective bandwidth selection method. In subsequent simulations, we also compute the smoothed derivative estimates. The improvement in the accuracy of the estimation after smoothing further demonstrates the effectiveness of the smoothing method we propose.

In addition, the index variables U do not follow a standard uniform distribution, which means that the above estimation method fails under other distributions. In the frame of nonparametric derivative estimation, to generalize covariate Z with unknown distribution F , we can use probability integral transformation (PIT)

$$F(Z) \sim \mathcal{U}(0, 1), \quad (2.14)$$

where $F(\cdot)$ is continuous, such that the density function $f(x) = F^{(1)}(x)$ exists and let f be bounded away from zero. This means that the index variables U with the distribution function F satisfying the above conditions can be transformed to an index variable following a uniform distribution. Now, the index variable with unknown distribution denotes Z . Then, in the varying coefficient models, we use the chain rule

$$\sum_{q=1}^p \frac{d}{dZ} a_q(Z) X_q = \sum_{q=1}^p \frac{dr_q(U)}{dU} \frac{dU}{dZ} X_q = f(Z) \sum_{q=1}^p r_q^{(1)}(U) X_q,$$

where $a_q(Z) = r_q(F(Z))$. In practice, since the distribution $F(\cdot)$ and density f both are unknown, the kernel density estimator [24] can be used to estimate the distribution and density to obtain $\hat{f}(\cdot)$ and $\hat{F}(\cdot)$ based on the plug-in bandwidth h [10], namely,

$$\hat{f}(z) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{Zi - z}{h}\right), \quad \hat{F}(z) = \frac{1}{n} \sum_{i=1}^n \int_{-\infty}^{\frac{Zi - z}{h}} K(z) dz. \quad (2.15)$$

The detailed derivative estimation algorithm in Table 1 is given below.

Table 1. Algorithm 1.

Step 1. Input Z , using kernel density estimation to obtain estimator $\hat{F}(Z)$, $\hat{f}(Z)$, and construct new index variable $U = \hat{F}(Z)$.

Step 2. Input X, U, Y , using corollary 2 and formula (2.2) to obtain optimal k and $\hat{r}_q^{(1)}(U)$.

Step 3. Let $\hat{a}_q^{(1)}(Z) = \hat{f}(Z)\hat{r}_q^{(1)}(U)$. Establishing the estimation of $a_q^{(1)}(\cdot)$ under index variable is unknown.

3. Second order derivative estimation

3.1. Estimation approach and asymptotic results of a second derivative

In varying-coefficient models, many researchers are also interested in the second derivative; for example, in analyzing the curvature of certain coefficient functions. However, higher-order derivatives of coefficient functions become more difficult to estimate, and the performance of the traditional estimation method becomes worse and worse with increasing derivative order. In order to obtain a better estimator of the second derivative estimator of coefficient functions, we proposed second order derivative estimator for uniform order statistics, as

$$\hat{a}_p^{(2)}(U_{(i)}) = 2 \sum_{j=1}^{k_2} \omega_{i,j,2} \left\{ \frac{\frac{\hat{a}_p(U_{(i+j+k_1)}) - \hat{a}_p(U_{(i+j)})}{U_{(i+j+k_1)} - U_{(i+j)}} - \frac{\hat{a}_p(U_{(i-j-k_1)}) - \hat{a}_p(U_{(i-j)})}{U_{(i-j-k_1)} - U_{(i-j)}}}{U_{(i+j+k_1)} + U_{(i+j)} - U_{(i-j-k_1)} - U_{(i-j)}} \right\}. \quad (3.1)$$

where $\hat{a}_p^{(2)}(U_{(i)})$ denotes the derivative estimator in $U_{(i)}$, and $\omega_{i,j,2}$ denote weights and satisfy $\omega_{i,j,2} \geq 0$, $\sum_{j=1}^{k_2} \omega_{i,j,2} = 1$, then k_1, k_2 are given turning parameter.

The weight $\omega_{i,2} = \{\omega_{i,1,2}, \dots, \omega_{i,k_2,2}\}^T$ determination method is analogous to the first order derivative estimation, so the optimal weights can be derived by minimizing the conditional variance of $\hat{a}_p^{(2)}(U_{(i)})$. Thus, we obtain the optimal weights $\omega_{i,2}$ by solving the following optimization problem

$$\hat{\omega}_{i,2} = \arg \min_{\omega_{i,j,2}, j=1, \dots, k_2} \text{Var} \left[\hat{a}_p^{(2)}(U_{(i)}) | \mathcal{D} \right]. \quad (3.2)$$

Proposition 3. $\text{Var} \left[\hat{a}_p^{(2)}(U_{(i)}) | \mathcal{D} \right]$ is a quadratic function of $\omega_{i,2} = \{\omega_{i,1,2}, \dots, \omega_{i,k_2,2}\}$, namely,

$$\text{Var} \left[\hat{a}_p^{(2)}(U_{(i)}) | \mathcal{D} \right] = \sigma^2 \omega_{i,2}^T \Sigma_0 \omega_{i,2}, \quad (3.3)$$

where Σ_0 is a $k_2 \times k_2$ positive semidefinite matrix, and the specific expression for Σ_0 can be found in 6.

According to Proposition 3, we can know that the conditional variance of $\hat{a}_p^{(2)}(U_{(i)})$ can be expressed as a quadratic function of $\omega_{i,j,2}$, which inspires us to use a quadratic programming algorithm to solve (3.2). Consequently, optimal weights $\hat{\omega}_{i,2}$ can be derived by solving the following quadratic programming problem with the given samples:

$$\hat{\omega}_{i,2} = \arg \min_{\sum_{j=1}^{k_2} \omega_{i,j,2} = 1, \omega_{i,j,2} \geq 0} \omega_{i,2}^T \Sigma_0 \omega_{i,2}. \quad (3.4)$$

There are several well-established algorithms to solve quadratic programming problems; in this paper, we adopt the dual interior point method proposed by [25] to optimize and solve the equation (3.4). This method has been implemented in the R package 'LowRankQP' specifically for solving quadratic programming problems, thus we choose to use the R package 'LowRankQP' to solve the optimization problem. In practice, the computational speed depends on the tuning parameter k_2 . We conduct a simulation experiment to evaluate the time required to obtain $\hat{a}_p^{(2)}(U_{(i)})$, comparing it with the two-step method involving triple expansion in its second step. The result is shown in Table 2. All computations are performed on an Intel(R) Xeon(R) CPU E5-2683 v4 processor (2.10 GHz base frequency). Table 2 reveals that the computational speed of our method is slightly slower than that of

the two-step approach, indicating that its performance remains acceptable. Furthermore, employing faster optimization techniques could further reduce computation time.

Table 2. The CPU computation time (in seconds) of $\hat{a}_1^{(2)}(U_{(i)})$, $k_1 + k_2 + 1 \leq i \leq n - k_1 - k_2$.

The samples $Y = \sin(6\pi U)X_1 + \sin(3\pi U)X_2 + \varepsilon$, U follows a uniform distribution on $[0, 1]$, $(X_1, X_2) \sim N(\mu, \Sigma)$, where $\mu = (0, 0)^T$, $\Sigma = \{(0.1)^{|i-j|}\}_{1 \leq i, j \leq 2}$, $\varepsilon \sim N(0, 0.4)$. The sample sizes are $n = 800$. The parameter $h_1 = 0.04$, $h_2 = 0.05$, $k_1 = 20$.

	$k_2 = 2$	$k_2 = 5$	$k_2 = 10$	$k_2 = 15$	$k_2 = 20$
Our method	12.3	12.8	13.0	13.6	13.8
Two-step estimation	12.2	12.0	12.2	12.0	11.9

However, there is a disadvantage that optimal weights are obtained using the weight of the quadratic programming algorithm, that is, we cannot get a theoretical expression for the weights $\omega_{i,j,2}$, which means that difficulties arise for the theoretical analysis of (3.1). Thus, we assume that the weights $\omega_{i,j,2}$ are known for the following theorem.

Theorem 2. *Under assumptions (A1)–(A5), and $h_1 \rightarrow 0, h_2 \rightarrow 0, k_1 \rightarrow \infty$ as $n \rightarrow \infty$, such that $nh_1 \rightarrow \infty, h_1/h_2 \rightarrow 0, k_1^2/n^2h_2 \rightarrow 0$. Besides, we assume $a_p(\cdot)$ has a third-order derivative $a_p^{(3)}(\cdot)$, and the weights $\omega_{i,j,2}$ are known. The absolute conditional bias and conditional variance of (3.1) are*

$$\begin{aligned} \left| \text{Bias} \left[\hat{a}_p^{(2)}(U_i) | \mathcal{D} \right] \right| &\leq \sup_{u \in [0,1]} |a_p^{(3)}(u)| \\ &\times \sum_{j=1}^{k_2} \omega_{i,j,2} \left\{ \frac{j^2 + jk_1 + \frac{1}{3}k_1^2}{(n+1)(2j+k_1)} + h_2^2 \mu_2 \frac{n+1}{4j+2k_1} \right\} (1 + o_p(1)), \end{aligned} \quad (3.5)$$

and

$$\begin{aligned} \text{Var} \left[\hat{a}_p^{(2)}(U_{(i)}) | \mathcal{D} \right] &\leq \frac{(n+1)^4 M \sigma^2}{k_1^2 n r_{pp}^-} \sum_{j=1}^{k_2} \sum_{m=1}^{k_2} \frac{\omega_{i,j,2} \omega_{i,m,2}}{(2j+k_1)(2m+k_1)} \\ &\times \left\{ \frac{2}{h_2} + \frac{2}{h_1} \left(\sum_{d=1}^{p-1} r_{dp}^+ r_{dp}^* + \sum_{d=1}^{p-1} \mathcal{R}_{dp} \mathcal{R}_{dp}^* \right) \right\} (1 + o_p(1)), \end{aligned} \quad (3.6)$$

where $\mathcal{R}_{dp} = \max\{0, \sup_{U \in [0,1]} E(X_d X_p | U)\}$, $\mathcal{R}_{dp}^* = \max\{0, \sup_{U \in [0,1]} e_{d,p}^T (E(X X^T | U))^{-1} e_{p,p}\}$, and $\mu_m, M, r_{pp}^-, r_{dp}^+$, r_{dp}^* maintain identical definitions to those specified in Theorem 1.

Corollary 3. *Under the assumptions of Theorem 2, and k_2 is bounded, then the exact asymptotic order for conditional bias and conditional variance of $\hat{a}_p^{(2)}(U_{(i)})$ are given by*

$$\text{Bias} \left[\hat{a}_p^{(2)}(U_{(i)}) | \mathcal{D} \right] = O_p \left(\max \left\{ \frac{n}{k_1} h_2^2, \frac{k_1}{n} \right\} \right), \quad (3.7)$$

$$\text{Var} \left[\hat{a}_p^{(2)}(U_{(i)}) | \mathcal{D} \right] = O_p \left(\frac{n^3}{k_1^4 h_1} \right). \quad (3.8)$$

3.2. Optimal parameter selection for k_1 and k_2

Similarly to parameter k for the first-order derivative estimation, parameter k_1, k_2 should also be selected, which controls the bias-variance trade-off. Based on the upper bound of conditional MSE, the following corollary provides the selection method of k_1, k_2 .

Corollary 4. *Under the assumptions of Theorem 2, we can obtain optimal k_1, k_2 that minimizes the upper bound of conditional MSE is*

$$(k_1, k_2)_{opt} = \arg \min_{k_1, k_2 \in N^+} \frac{1}{n} \sum_{i=1}^n \left[\left(\sup_{u \in [0, 1]} |a_p^{(3)}(u)| \sum_{j=1}^{k_2} \omega_{i,j} \left\{ \frac{j^2 + jk_1 + \frac{1}{3}k_1^2}{(n+1)(2j+k_1)} + h_2^2 \mu_2 \frac{n+1}{4j+2k_1} \right\} \right)^2 \right. \\ \left. + \frac{(n+1)^4 M \sigma^2}{k_1^2 n r_{pp}^-} \sum_{j=1}^{k_2} \sum_{m=1}^{k_2} \frac{\omega_{i,j} \omega_{i,m}}{(2j+k_1)(2m+k_1)} \left\{ \frac{2}{h_2} + \frac{2}{h_1} \left(\sum_{d=1}^{p-1} r_{dp}^+ r_{dp}^* + \sum_{d=1}^{p-1} \mathcal{R}_{dp} \mathcal{R}_{dp}^* \right) \right\} \right] \right\}. \quad (3.9)$$

The unknown quantity $\sup_{u \in [0, 1]} |a_p^{(3)}(u)|$ can be obtained by the two-step method with local polynomial expansion of order $p = 4$ in the second step. Moreover, $\mathcal{R}_{dp}, \mathcal{R}_{dp}^*$ are estimated in a way similar to r_{dp}^* . The estimation methods for the remaining unknown quantity $\sigma^2, r_{pp}^-, r_{dp}^+, r_{dp}^*$ are the same as those in Corollary 2. The optimal value k_1, k_2 can then be obtained by a grid search.

3.3. Smoothing the second order estimators and generalizing it to arbitrary distributions

Similar to (2.13), we can also obtain the smooth estimator provided at arbitrary point $u_0 \in [0, 1]$ by local polynomial smoothing, that is,

$$\tilde{a}_p^{(2)}(u_0) = e_{1,g}^T \left(\mathbf{U}_2^T(u_0) \mathbf{W}_2(u_0) \mathbf{U}_2(u_0) \right)^{-1} \mathbf{U}_2^T(u_0) \mathbf{W}_2(u_0) \hat{\mathbf{a}}_p^{(2)}, \quad (3.10)$$

where g represents the order of local polynomial fitting, $e_{1,g}$ represents a column vector of length g , and the first element is 1 and the remaining elements are 0,

$$\mathbf{W}_2(u_0) = \text{Diag}(K_h(U_{(k_1+k_2+1)} - u_0), \dots, K_h(U_{(n-k_1-k_2)} - u_0)),$$

$$\mathbf{U}_2(u_0) = \begin{pmatrix} 1 & U_{(k_1+k_2+1)} - u_0 & \dots & (U_{(k_1+k_2+1)} - u_0)^p \\ 1 & U_{(k_1+k_2+2)} - u_0 & \dots & (U_{(k_1+k_2+2)} - u_0)^p \\ \vdots & \vdots & & \vdots \\ 1 & U_{(n-k_1-k_2)} - u_0 & \dots & (U_{(n-k_1-k_2)} - u_0)^p \end{pmatrix}, \quad \hat{\mathbf{a}}_p^{(2)} = \begin{pmatrix} \hat{a}_p^{(2)}(U_{(k_1+k_2+1)}) \\ \hat{a}_p^{(2)}(U_{(k_1+k_2+2)}) \\ \vdots \\ \hat{a}_p^{(2)}(U_{(n-k_1-k_2)}) \end{pmatrix}.$$

The smoothing parameter h is chosen in the same way as in Section 3.3.

To generalize to arbitrary distributions for the index variables U , we still use the Probability Integral Transform (PIT) as in (2.14) to transform the index variables Z to U . Then, let $a_q(Z) = r_q(F(Z))$, the second derivative of coefficient functions with respect to Z is

$$\frac{d^2}{dZ^2} a_q(Z) = \frac{d}{dZ} (f(Z) r_q^{(1)}(U)) = f^{(1)}(Z) r_q^{(1)}(U) + f(Z) r_q^{(2)}(U),$$

where the estimation of unknown $f^{(1)}(x)$ can be estimated using the kernel density derivative estimator. To take a similar procedure to Algorithm 1, we can obtain

$$\hat{a}_q^{(2)}(Z) = \hat{f}^{(1)}(Z) \hat{r}_q^{(1)}(U) + \hat{f}(Z) \hat{r}_q^{(2)}(U). \quad (3.11)$$

4. Simulation studies

4.1. Simulation 1

In this section, we perform simulations to demonstrate the performance of the proposed methods. We consider the following varying coefficient model

$$Y = a_1(U)X_1 + a_2(U)X_2 + a_3(U)X_3 + a_4(U)X_4 + \varepsilon.$$

For the above model, two examples are considered. The true coefficient functions in two examples are given in Table 3 based on the smoothness difference. In Example 1, the smoothness of $a_2(U)$ and $a_3(U)$ is the same, whereas in Example 2, the smoothness of all the coefficient functions differs.

Table 3. True coefficient functions in two examples.

True Functions	Example 1	Example 2
$a_1(U)$	$\sin(6\pi U)$	$\sin(5\pi U) + 2U$
$a_2(U)$	$\sin(3\pi U)$	$\sin(3\pi U)$
$a_3(U)$	$\cos(3\pi U)$	$9U^3 - 5U^2 + 2U$
$a_4(U)$	$\sin(\pi U)$	$\sin(\pi U)$

In two examples, U follows a uniform distribution on $[0, 1]$. For X_1 and X_2 , we set $(X_1, X_2, X_3, X_4)^T \sim N(\mu, \Sigma)$, where $\mu = (0, 0, 0, 0)^T$, $\Sigma = (\sigma_{ij})_{4 \times 4}$ and $\sigma_{ij} = \rho^{|i-j|}$, $\rho = 0.1$ or 0.5 .

Here, we set (X_1, X_2, X_3, X_4) , U , and ε to be independent of each other. Let the random error term ε follow a normal distribution with mean 0 and variance σ^2 , and σ^2 is chosen so that the signal-to-noise ratio is 5:1, namely,

$$\sigma^2 = 0.2\text{var}\{E(Y|U, X_1, X_2, X_3, X_4)\}.$$

For each example, we perform 200 simulations with sample sizes $n = 400, 800$ in each case. To evaluate the performance of the symmetric difference quotients method (SQE), the performance of the symmetric difference quotients smoothing method (S-SQE) of order $g = 3$, the local polynomial estimator (LPE) of order $g = 3$, and the two-step estimator (TSE), where we use local polynomial fitting of order $g = 3$ in the second step, which are also computed. Additionally, we incorporate a smoothing spline estimator (SPE) based on B-spline, where the derivative estimate can be obtained directly by differentiating the B-spline basis functions. For SQE, optimal k and k_1, k_2 are selected based on Corollary 2 and 4 in a set $\{1, 2, \dots, \lfloor n/2 \rfloor\}$ and a search space $\{1, 2, \dots, \lfloor n/4 \rfloor\} \otimes \{1, 2, \dots, \lfloor n/4 \rfloor\}$. For LPE, the optimal bandwidth h_{LLE} is selected by the cross-validation method. In the two-step estimation, the initial bandwidth $h_0 = 0.5h_{LLE}$ is taken in the first step, and then we use cross-validation to select the bandwidth in the second step. Furthermore, for SPE, we employ equidistant knots, with the number of knots selected via cross-validation. For each method, the Epanechnikov kernel function is used, i.e., $K(t) = 0.75(1 - t^2)_+$. All the aforementioned numerical simulations are implemented in the R software.

To systematically compare the estimation performance between methodologies, we calculate the mean absolute error (MAE) of the derivative estimates for each coefficient function. A critical challenge arises from the inherent scale incompatibility between coefficient functions. As their derivatives exhibit distinct dynamic ranges (varying upper and lower bounds), direct comparison of

raw MAE values becomes dimensionally inconsistent. To resolve this, we implement a range-based standardization procedure that normalizes MAE values relative to the characteristic scale of each derivative function. Furthermore, we address the boundary effect paradox observed in LPE, TSE, and SPE, where derivative estimation accuracy deteriorates severely near domain boundaries. In addition, SQE provides reliable estimates only within the interior region of the domain. Consequently, to ensure a fair comparison, we exclude boundary points and focus on evaluating performance exclusively within the interior domain. Thus, the performance measure is defined using the adjusted mean absolute error.

$$MAE_1 = \frac{1}{n-2k} \sum_{i=k+1}^{n-k} |\hat{a}_p^{(1)}(U_i) - a_p^{(1)}(U_i)| / \{\max(a_p^{(1)}(U)) - \min(a_p^{(1)}(U))\}, \quad (4.1)$$

and

$$MAE_2 = \frac{1}{n-2(k_1+k_2)} \sum_{i=k_1+k_2+1}^{n-k_1-k_2} |\hat{a}_p^{(2)}(U_i) - a_p^{(2)}(U_i)| / \{\max(a_p^{(2)}(U)) - \min(a_p^{(2)}(U))\}. \quad (4.2)$$

We also calculate the average adjusted mean absolute error of the coefficient functions, denoted as mean. For each method, the mean adjusted mean absolute error based on 200 simulations is presented in Tables 4 and 5. Moreover, in Figures 2 and 3, we present the graphs of the first- and second-order derivatives of the estimated coefficient functions for Example 2, when $n = 800$ and $\rho = 0.1$. Additionally, Figures 4 and 5 present boxplots of the root mean squared error (RMSE) for the derivative function estimates under the same examples.

The results in Tables 4 and 5 show that, in terms of the mean adjusted mean absolute error, the average error of the SQE method is significantly smaller than that of the TSE, SPE and LPE methods, thus highlighting the effectiveness of the proposed approach. In the estimation of the derivative with the highest smoothness, LPE consistently performs the worst. This is because the local polynomial method can only employ a fixed bandwidth when estimating the derivative, resulting in poor estimates for the derivatives of coefficient functions with higher smoothness. The limited bandwidth leads to an inadequate approximation of the derivative, especially for functions with higher smoothness. Furthermore, SPE consistently yields poor derivative estimates for coefficient functions with the lowest smoothness, demonstrating that SPE should generally be avoided for derivative estimation. In addition, the estimation error of S-SQE relative to SQE is smaller, demonstrating the effectiveness of our smoothing method in further improving estimation accuracy and reducing the estimation error. As seen in Example 1, the estimation of the SQE and S-SQE method in $\rho = 0.1$ is better than the estimation at $\rho = 0.5$. However, the mean of the adjusted mean absolute error of SQE and S-SQE in the case $\rho = 0.5$ does not increase much compared to the case $\rho = 0.1$. However, the LPE, TSE, and SPE methods exhibit a substantial decrease in the accuracy of the estimation as the correlation increases, suggesting that our method is less sensitive to the correlation between variables. Furthermore, it can be seen that the error in the second-order derivative is larger than that in the first-order derivative, but SQE and S-SQE are able to give an estimate with lower error. In addition, as the sample size increases, the adjusted mean absolute error of the four methods becomes smaller.

Table 4. Means of adjusted mean absolute error of first order derivative based on 200 samples.

Method	$n = 400$					$n = 800$				
	$\hat{a}_1^{(1)}(\cdot)$	$\hat{a}_2^{(1)}(\cdot)$	$\hat{a}_3^{(1)}(\cdot)$	$\hat{a}_4^{(1)}(\cdot)$	mean	$\hat{a}_1^{(1)}(\cdot)$	$\hat{a}_2^{(1)}(\cdot)$	$\hat{a}_3^{(1)}(\cdot)$	$\hat{a}_4^{(1)}(\cdot)$	mean
Example 1 $\rho = 0.1$										
SQE	0.077	0.065	0.061	0.058	0.065	0.059	0.061	0.053	0.036	0.052
S-SQE	0.068	0.057	0.059	0.058	0.060	0.051	0.049	0.046	0.036	0.046
LPE	0.056	0.062	0.062	0.189	0.092	0.038	0.061	0.064	0.183	0.086
TSE	0.114	0.063	0.060	0.080	0.079	0.081	0.057	0.059	0.036	0.057
SPE	0.309	0.101	0.075	0.067	0.138	0.309	0.038	0.061	0.032	0.111
Example 1 $\rho = 0.5$										
SQE	0.081	0.076	0.065	0.060	0.071	0.063	0.072	0.063	0.040	0.059
S-SQE	0.068	0.068	0.062	0.059	0.064	0.058	0.060	0.055	0.040	0.053
LPE	0.059	0.080	0.079	0.216	0.108	0.047	0.060	0.062	0.159	0.082
TSE	0.122	0.111	0.111	0.056	0.100	0.075	0.079	0.073	0.037	0.066
SPE	0.308	0.102	0.091	0.127	0.157	0.309	0.047	0.063	0.034	0.113
Example 2 $\rho = 0.1$										
SQE	0.077	0.067	0.098	0.067	0.077	0.048	0.052	0.011	0.037	0.037
S-SQE	0.075	0.060	0.098	0.066	0.075	0.042	0.050	0.011	0.037	0.035
LPE	0.054	0.063	0.390	0.190	0.176	0.034	0.049	0.047	0.140	0.067
TSE	0.162	0.074	0.270	0.090	0.150	0.068	0.073	0.014	0.036	0.047
SPE	0.303	0.086	0.074	0.043	0.126	0.301	0.058	0.052	0.035	0.112
Example 2 $\rho = 0.5$										
SQE	0.078	0.078	0.243	0.067	0.117	0.061	0.067	0.082	0.041	0.063
S-SQE	0.071	0.071	0.240	0.066	0.112	0.054	0.062	0.082	0.041	0.060
LPE	0.059	0.081	0.501	0.224	0.217	0.046	0.056	0.338	0.145	0.146
TSE	0.172	0.127	0.322	0.095	0.179	0.089	0.119	0.114	0.040	0.091
SPE	0.304	0.105	0.097	0.132	0.160	0.301	0.063	0.066	0.037	0.117

Table 5. Means of adjusted mean absolute error of second order derivative based on 200 samples.

Method	n = 400					n = 800				
	$\hat{a}_1^{(2)}(\cdot)$	$\hat{a}_2^{(2)}(\cdot)$	$\hat{a}_3^{(2)}(\cdot)$	$\hat{a}_4^{(2)}(\cdot)$	mean	$\hat{a}_1^{(2)}(\cdot)$	$\hat{a}_2^{(2)}(\cdot)$	$\hat{a}_3^{(2)}(\cdot)$	$\hat{a}_4^{(2)}(\cdot)$	mean
Example 1 $\rho = 0.1$										
SQE	0.102	0.110	0.083	0.220	0.129	0.105	0.089	0.12	0.119	0.100
S-SQE	0.085	0.097	0.081	0.217	0.120	0.065	0.091	0.081	0.118	0.089
LPE	0.153	0.078	0.082	0.583	0.224	0.110	0.083	0.083	0.118	0.237
TSE	0.139	0.079	0.080	0.213	0.128	0.099	0.076	0.081	0.109	0.089
SPE	0.327	0.144	0.167	0.198	0.309	0.322	0.068	0.138	0.086	0.153
Example 1 $\rho = 0.5$										
SQE	0.113	0.101	0.092	0.185	0.123	0.085	0.105	0.104	0.112	0.101
S-SQE	0.087	0.094	0.090	0.184	0.114	0.071	0.095	0.095	0.111	0.093
LPE	0.154	0.090	0.092	0.658	0.248	0.140	0.076	0.077	0.480	0.193
TSE	0.148	0.136	0.136	0.168	0.147	0.088	0.100	0.090	0.103	0.095
SPE	0.325	0.141	0.175	0.548	0.297	0.323	0.082	0.137	0.082	0.156
Example 2 $\rho = 0.1$										
SQE	0.077	0.097	0.071	0.157	0.117	0.092	0.089	0.062	0.120	0.090
S-SQE	0.075	0.109	0.082	0.218	0.125	0.073	0.084	0.062	0.119	0.085
LPE	0.140	0.084	0.369	0.533	0.280	0.122	0.0749	0.297	0.416	0.227
TSE	0.230	0.089	0.190	0.199	0.177	0.099	0.098	0.081	0.104	0.095
SPE	0.326	0.190	0.049	0.094	0.165	0.301	0.058	0.052	0.035	0.112
Example 2 $\rho = 0.5$										
SQE	0.100	0.123	0.19	0.233	0.162	0.096	0.134	0.074	0.125	0.107
S-SQE	0.093	0.115	0.189	0.232	0.157	0.078	0.119	0.074	0.125	0.099
LPE	0.134	0.094	0.470	0.609	0.327	0.133	0.082	0.081	0.402	0.235
TSE	0.242	0.257	0.205	0.201	0.201	0.107	0.121	0.097	0.111	0.109
SPE	0.311	0.228	0.062	0.535	0.284	0.327	0.125	0.048	0.094	0.148

From Figures 2 and 3, the estimation of the first-order derivatives of four methods is similar. However, the estimation of the fourth function in the second-order derivatives estimated by the other methods is poor, whereas S-SQE estimates the general shape of the functions. In addition, the LPE and SPE methods show significantly poorer performance in estimating the second-order derivative compared to the TSE and S-SQE methods, which is attributed to the chosen bandwidth not being optimal for second-order derivative estimation. According to 4 and 5, the boxplot width of S-SQE is nearly the smallest among all methods, while maintaining a stable width across all derivative functions without observable outliers. This demonstrates the superior stability of our proposed estimation method.

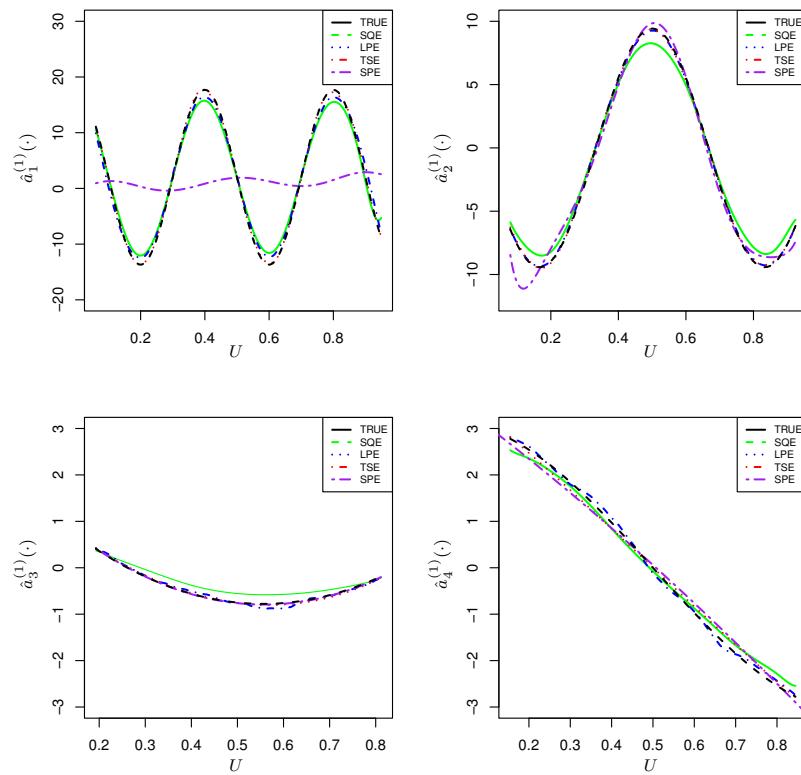


Figure 2. The average of the first order function estimations based on Example 2 and sample size $n = 800$ and $\rho = 0.1$. The green solid curves: The results of S-SQE; The red dot-dashed curves: The results of TSE; The blue dotted curves: The results of LPE; The purple dot-long dashed curves: The results of SPE; The black solid curves: True function.

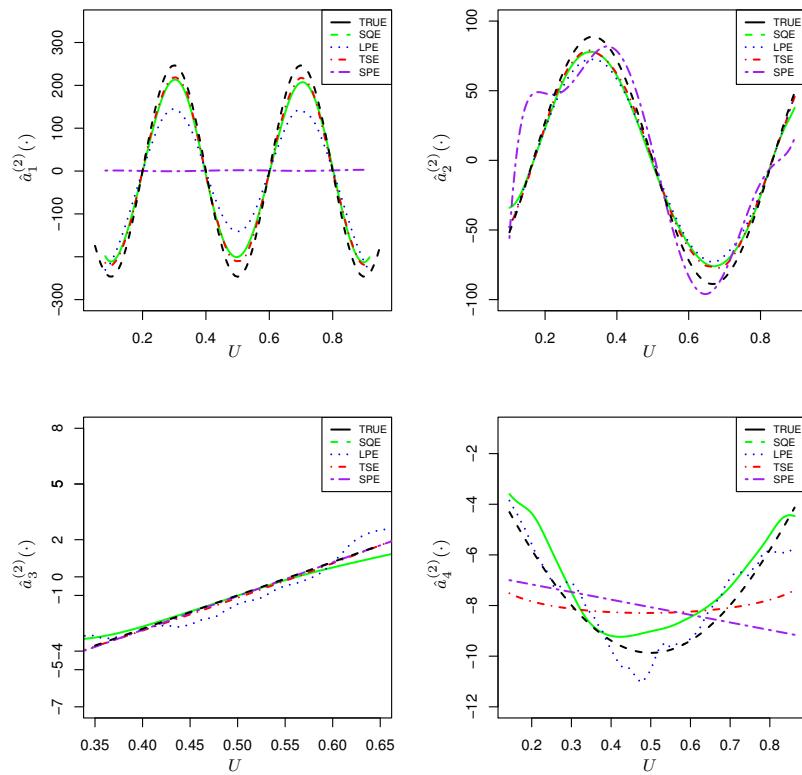


Figure 3. The average of the second order function estimations based on Example 2 and sample size $n = 800$ and $\rho = 0.1$. The green solid curves: The results of S-SQE; The red dot-dashed curves: The results of TSE; The blue dotted curves: The results of LPE; The purple dot-long dashed curves: The results of SPE; The black solid curves: True function.

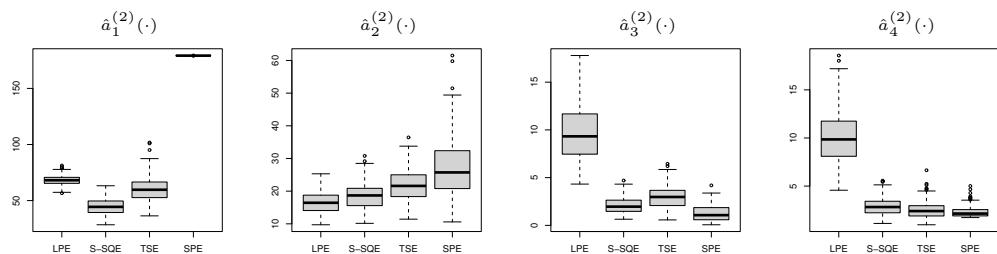


Figure 4. The boxplot of RMSE for the first order function estimations based on Example 2 and sample size $n = 800$ and $\rho = 0.1$.

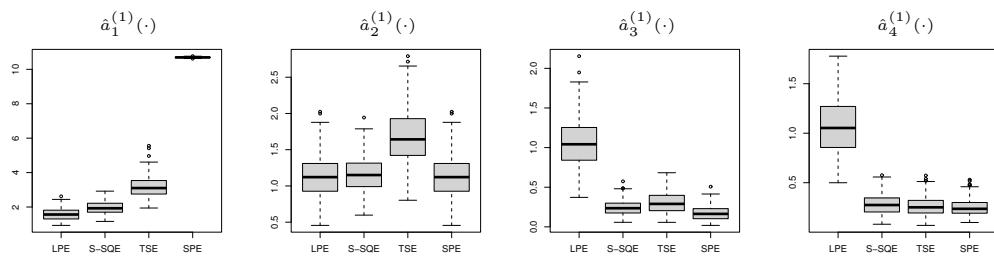


Figure 5. The boxplot of RMSE for the second order function estimations based on Example 2 and sample size $n = 800$ and $\rho = 0.1$.

4.2. Simulation 2

In this section, we consider a more complicated example: A six-dimensional varying-coefficient model. We consider the following varying coefficient model

$$Y = \sin(6\pi U)X_1 + (\sin(4\pi U) + 2U)X_2 + \cos(4\pi U)X_3 + (\sin(3\pi U) - U)X_4 + \cos(2\pi U)X_5 + \sin(\pi U)X_6 + \varepsilon.$$

In this example, U follows a uniform distribution on $[0, 1]$. For X_1 and X_2 , we set $(X_1, X_2, X_3, X_4, X_5, X_6)^T \sim N(\mu, \Sigma)$, where $\mu = (0, 0, 0, 0, 0, 0)^T$, $\Sigma = (\sigma_{ij})_{6 \times 6}$ and $\sigma_{ij} = 0.5^{|i-j|}$.

Here, we set $(X_1, X_2, X_3, X_4, X_5, X_6)$, U , and ε to be independent of each other. Let the random error term ε follows a normal distribution with mean 0 and variance σ^2 , and σ^2 is chosen so that the signal-to-noise ratio is 5:1, namely,

$$\sigma^2 = 0.2\text{var}\{E(Y|U, X_1, X_2, X_3, X_4, X_5, X_6)\}.$$

The mean of the adjusted mean absolute error based on 200 simulations is presented in Tables 6 and 7.

Table 6. Means of adjusted mean absolute error of first order derivative based on 200 samples.

Sample size	Methods	$\hat{a}_1^{(1)}(\cdot)$	$\hat{a}_2^{(1)}(\cdot)$	$\hat{a}_3^{(1)}(\cdot)$	$\hat{a}_4^{(1)}(\cdot)$	$\hat{a}_5^{(1)}(\cdot)$	$\hat{a}_6^{(1)}(\cdot)$	mean
$n = 400$	SQE	0.091	0.075	0.078	0.074	0.075	0.062	0.076
	S-SQE	0.075	0.071	0.076	0.074	0.074	0.061	0.072
	LPE	0.112	0.054	0.055	0.061	0.086	0.168	0.089
	TSE	0.148	0.258	0.155	0.061	0.119	0.059	0.133
	SPE	0.309	0.054	0.073	0.107	0.068	0.096	0.118
$n = 800$	SQE	0.063	0.056	0.063	0.064	0.059	0.053	0.060
	S-SQE	0.057	0.052	0.056	0.060	0.059	0.053	0.056
	LPE	0.047	0.049	0.046	0.065	0.091	0.179	0.080
	TSE	0.160	0.114	0.171	0.100	0.061	0.045	0.109
	SPE	0.310	0.047	0.069	0.095	0.083	0.114	0.120

Table 7. Means of adjusted mean absolute error of second order derivative based on 200 samples.

Sample size	Methods	$\hat{a}_1^{(2)}(\cdot)$	$\hat{a}_2^{(2)}(\cdot)$	$\hat{a}_3^{(2)}(\cdot)$	$\hat{a}_4^{(2)}(\cdot)$	$\hat{a}_5^{(2)}(\cdot)$	$\hat{a}_6^{(2)}(\cdot)$	mean
$n = 400$	SQE	0.112	0.107	0.104	0.091	0.084	0.187	0.136
	S-SQE	0.110	0.099	0.096	0.090	0.084	0.187	0.111
	LPE	0.224	0.128	0.140	0.091	0.112	0.414	0.185
	TSE	0.183	0.546	0.226	0.091	0.180	0.168	0.232
	SPE	0.312	0.070	0.103	0.228	0.107	0.287	0.185
$n = 800$	SQE	0.102	0.084	0.109	0.094	0.070	0.194	0.120
	S-SQE	0.097	0.081	0.091	0.087	0.070	0.194	0.103
	LPE	0.158	0.092	0.094	0.074	0.120	0.480	0.170
	TSE	0.259	0.166	0.334	0.144	0.082	0.167	0.192
	SPE	0.319	0.063	0.101	0.202	0.192	0.514	0.232

Like in Simulation 1, Tables 6 and 7 show that LPE, TSE, and SPE struggle to accurately estimate all functions among the second-order derivatives, but SQE and S-SQE maintain a relatively low estimation error. However, in this example, the TSE and SPE methods perform the worst, suggesting that as the dimensionality increases, the estimation accuracy of coefficient function derivatives using TSE and SPE declines significantly. In addition, compared to LPE, TSE, and SPE, the SQE and S-SQE methods demonstrate relatively stable estimation accuracy for the derivatives of each coefficient function. With the exception of the second-order derivative of the sixth coefficient function, the estimation errors for all other coefficient functions remain at a consistent level.

Furthermore, to demonstrate the statistically significant improvement in mean estimation accuracy of S-SQE over competing methods, we conduct paired t-tests [26] comparing S-SQE's mean term based on 200 simulations against those of other approaches. We obtain the corresponding p-values, all of which are less than 1×10^{-10} , demonstrating that S-SQE achieves significantly improved performance compared to other methods. This further validates that our method yields more stable estimates in the presence of varying smoothness across the coefficient functions.

5. Real data analysis

To further illustrate the proposed methodologies, we choose to implement these approaches to estimate derivative functions on the Boston Housing Dataset in this section. For this dataset, calculating the first order derivatives of the coefficient functions enables the analysis of the rate of change for each coefficient function at different points, while the second order of derivatives can be used to construct confidence intervals for local polynomial estimation of the coefficient functions. For the given dataset, we construct a varying-coefficient model. Consequently, the estimation of derivative functions carries substantial significance.

The dataset encompasses 506 observations spanning 14 distinct variables and is retrievable from the R package 'mlbench'. According to Fan and Huang's [27] modeling analysis of the data using a varying coefficient model, we designate MEDV, which denotes the median value of owner-occupied homes measured in thousands of dollars, as our response variable, and LSTAT, indicating the proportion of the population with a lower socioeconomic status, as our index variable. In terms of covariate

variables, we adopt the selection criteria established by Wang and Xia through the application of the KLASSO method [22]. This has led to the inclusion of four predictor variables: INT (the model intercept), CRIM (the crime rate per capita within each town), RM (the average number of rooms per residential unit), and PTRATIO (the ratio of pupils to teachers in each town). Prior to the deployment of these methodologies, the data undergoes a normalization process for the covariate variables, with the exception of INT and the response variable. In addition, the index variable is subjected to a transformation to ensure that its distribution conforms to a uniform distribution between 0 and 1.

We estimate the first and second order derivatives of the coefficient functions using SQE, LPE, TSE, and SPE. The plots of the estimated first and second derivatives are presented in Figures 6 and 7. To assess the accuracy of these estimates, we identify the points corresponding to the index variables at the maxima and minima of the first-order derivative for each coefficient function at the interior points. We then calculate the mean distance from the second-order derivative estimates at these points to zero, and the results are presented in Table 8. In theory, the second-order derivative at the extrema of the first-order derivative should be zero. Therefore, a smaller distance from the second-order derivative to the zero point indicates a more accurate estimate, while a larger distance suggests a poorer estimation.

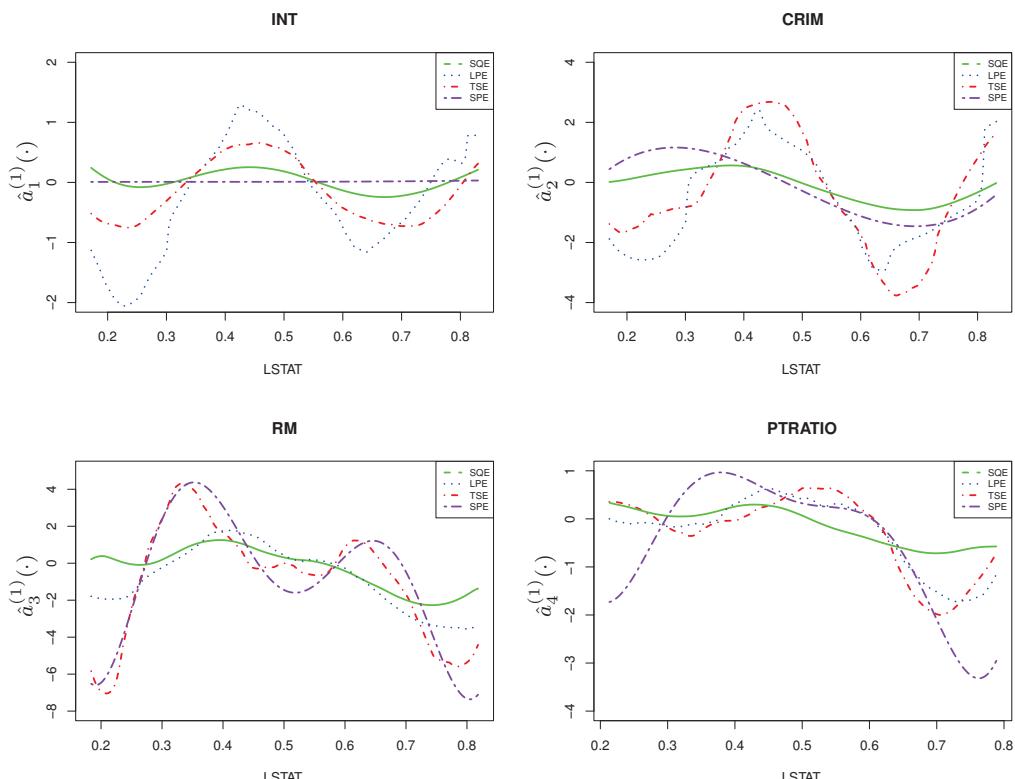


Figure 6. The estimated first order functions for four variables. The green solid curves: The results of SQE; the red dot-dashed curves: The results of TSE; the blue dotted curves: The results of LPE; the purple dot-long dashed curves: The results of SPE.

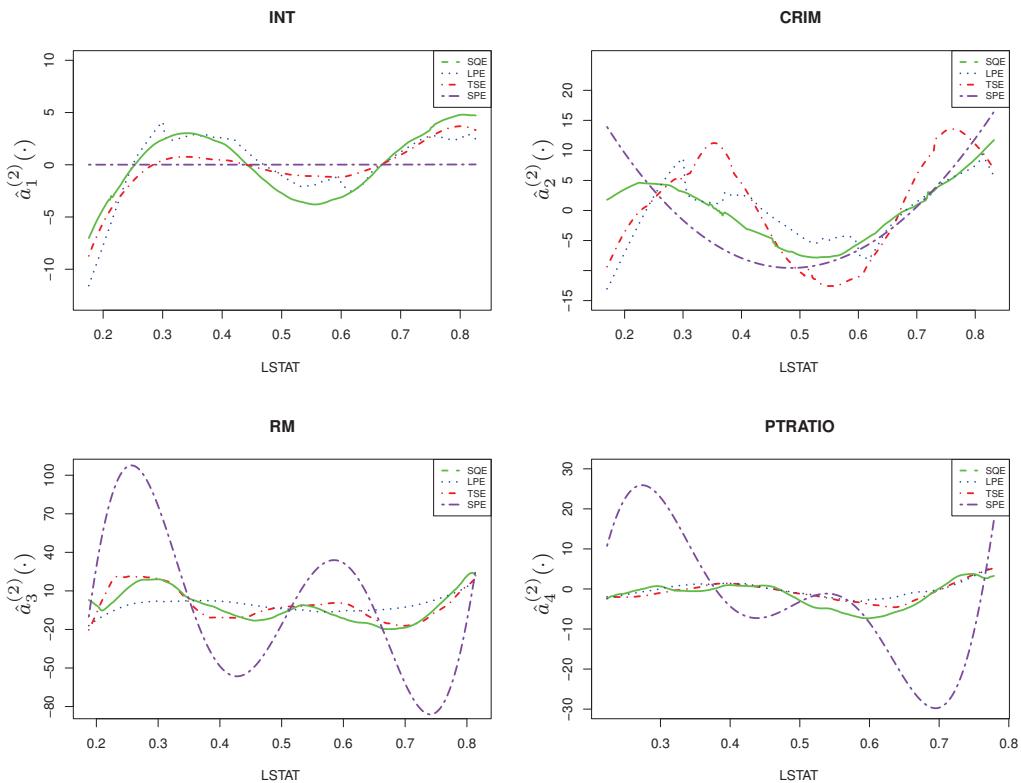


Figure 7. The estimated second order functions for four variables. The green solid curves: The results of SQE; the red dot-dashed curves: The results of TSE; the blue dotted curves: The results of LPE; the purple dot-long dashed curves: The results of SPE.

As shown in Figure 6, when LSTAT exceeds 0.5, the first-order derivatives of all coefficient functions are negative, indicating a declining trend. This aligns with empirical observations: Areas with higher proportions of lower socioeconomic status populations generally exhibit lower median home values. Notably, a local minimum occurs near LSTAT=0.7 (corresponding to 15%), where the rate of decline peaks. Beyond this threshold, the descent rate moderates, suggesting that housing price depreciation slows when LSTAT surpasses 15%. Furthermore, the first-order derivatives of coefficient functions estimated by SPE on INT are asymptotically zero, indicating that INT's coefficient function is essentially constant. This conclusion sharply contradicts Wang and Xia's finding [22], which confirmed the existence of a varying-coefficient intercept term. Such inconsistency significantly undermines the credibility of SPE.

According to Table 8, we observe substantial distance among LPE, TSE, and SPE estimates, indicating their uniformly poor performance in derivative estimation. For the RM variable, all methods yield substantially large distances, likely attributable to the high amplitude of its first-order derivative function. In addition, with the exception of the RM variable, the distances estimated by SQE are very close to zero, which suggests that our method provides the most accurate estimates. In summary, our method effectively estimates both the first and second derivatives, and accurately identifies the maxima and minima of the derivatives in this dataset.

Table 8. The mean distance from given points to the zero.

Methods	INT	CRIM	RM	PTRATIO
SQE	0.07	0.61	4.83	0.38
LPE	3.49	5.75	15.37	1.75
TSE	1.46	5.54	12.60	2.44
SPE	0.03	10.64	9.26	2.43

6. Conclusions

We consider the first and second order derivative function estimation problem of the varying coefficient model when the coefficient functions may have different degrees of smoothness, and a weighted symmetric difference quotient estimation method is proposed. Along with the proposed method for estimating derivative functions, we also present an effective approach for selecting tuning parameters. Both the simulation results and the real data analysis demonstrate the effectiveness of our method in estimating the derivative functions. Furthermore, under certain conditions, we provide the asymptotic upper bounds for the conditional bias and conditional variance of the estimator and exact order of conditional bias and variance. In addition, based on these bounds, we propose a parameter selection method derived from the asymptotic upper bound of the conditional MSE. Furthermore, relevant researchers may explore alternative methods to obtain a more precise estimate $\hat{a}_p(\cdot)$. Incorporating $\hat{a}_p(\cdot)$ higher-precision estimate of a into the weighted symmetric difference quotient estimator could potentially further enhance the accuracy of derivative estimation.

In future research, we plan to extend our method to time-varying coefficient models in economics. Additionally, we aim to investigate higher-order derivative estimation of coefficient functions, derivative estimation for multivariate covariates, and the extension of our approach to additive models for derivative function estimation. Furthermore, due to the difficulty of the proof, we do not present an exact expression of asymptotic conditional bias and conditional variance. Thus, obtaining precise asymptotic expressions for the conditional bias and variance will be essential, as this could facilitate the selection of more appropriate tuning parameters, ultimately leading to improved estimation results. Besides, when deriving the upper bounds for the asymptotic bias and variance of the second-order derivative estimator, we assume that the weights are known—a relatively stringent condition, and we aim to explore methods for obtaining the exact expression of the second-order derivative weights to circumvent this condition in the future.

Appendix: Proof of Theorem

We use the following notation.

$$\mathbf{X} = (X_1, \dots, X_p), \quad \mathbf{Y} = (Y_1, \dots, Y_n)^T, \quad \boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)^T.$$

$$\Omega_p(U) = E(\mathbf{X}^T \mathbf{X} | U), \quad K_h(u - u_0) = K\left(\frac{u - u_0}{h}\right).$$

$$A = \text{diag}(1, h_1), \quad Q = \text{diag}(1, h_2), \quad G = I_p \otimes \text{diag}(1, h_1),$$

where I_p is a $p \times p$ matrix.

$$\mathbf{X}_{2,(b)} = \begin{pmatrix} X_{1p} & X_{1p}(U_1 - U_{(b)}) \\ \vdots & \vdots \\ X_{np} & X_{np}(U_n - U_{(b)}) \end{pmatrix}.$$

$$\mathbf{X}_{0,b} = \begin{pmatrix} X_{11} & X_{11}(U_1 - U_b) & \cdots & X_{1p} & X_{1p}(U_1 - U_b) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ X_{n1} & X_{n1}(U_n - U_b) & \cdots & X_{np} & X_{np}(U_n - U_b) \end{pmatrix}.$$

$$\mathbf{X}_{(b)} = \begin{pmatrix} X_{11} & X_{11}(U_1 - U_{(b)}) & \cdots & X_{1p} & X_{1p}(U_1 - U_{(b)}) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ X_{n1} & X_{n1}(U_n - U_{(b)}) & \cdots & X_{np} & X_{np}(U_n - U_{(b)}) \end{pmatrix}.$$

$$W_{2,(b)} = \text{diag}(K_{h_2}(U_1 - U_{(b)}), \dots, K_{h_2}(U_n - U_{(b)})),$$

$$W_{0,b} = \text{diag}(K_{h_1}(U_1 - U_b), \dots, K_{h_1}(U_n - U_b)),$$

$$W_{(b)} = \text{diag}(K_{h_1}(U_1 - U_{(b)}), \dots, K_{h_1}(U_n - U_{(b)})).$$

$$B_n = \sum_{q=1}^{p-1} \begin{pmatrix} X_{1q} e_{2q-1,2p}^T (\mathbf{X}_{0,1}^T W_{0,1} \mathbf{X}_{0,1})^{-1} \mathbf{X}_{0,1} W_{0,1} \\ \vdots \\ X_{nq} e_{2q-1,2p}^T (\mathbf{X}_{0,n}^T W_{0,n} \mathbf{X}_{0,n})^{-1} \mathbf{X}_{0,n} W_{0,n} \end{pmatrix}.$$

A1. Proof of Proposition 2.

The proof of this proposition is essentially a simplified version of Corollary 1. For detailed proof, please refer to Corollary 1.

A2. Proof of Proposition 2.

$$\begin{aligned} \text{Var} \left[\hat{a}_p^{(1)}(U_i) | \mathcal{D} \right] &= \text{Var} \left[\sum_{j=1}^k \omega_{i,j} \left(\frac{\hat{a}_p(U_{(i+j)}) - \hat{a}_p(U_{(i-j)})}{U_{(i+j)} - U_{(i-j)}} \right) | \mathcal{D} \right] \\ &= \text{Var} \left[\sum_{j=1}^k \frac{\omega_{i,j}}{U_{(i+j)} - U_{(i-j)}} \{H_{(i+j)} - H_{(i-j)}\} \varepsilon | \mathcal{D} \right] \\ &= \sigma^2 \sum_{l=1}^n \left(\sum_{j=1}^k \frac{\omega_{i,j}}{U_{(i+j)} - U_{(i-j)}} (H_{l,(i+j)} - H_{l,(i-j)}) \right)^2 \\ &\leq \sigma^2 \sum_{j=1}^k \left(\frac{\omega_{i,j}}{U_{(i+j)} - U_{(i-j)}} \right)^2 \left\{ \sum_{l=1}^n \sum_{j=1}^k (H_{l,(i+j)} - H_{l,(i-j)})^2 \right\}. \end{aligned}$$

where $H_{(b)} = (1, 0)(\mathbf{X}_{2,(b)}^T W_{2,(b)} \mathbf{X}_{2,(b)})^{-1} \mathbf{X}_{2,(b)}^T W_{2,(b)} (I_n - B_n)$, $H_{l,(b)}$ is the l -th element of $H_{(b)}$. Therefore, we obtain an upper bound of $Var(\hat{a}_p^{(1)}(U_i)|\mathcal{D})$. Then, the Lagrange multiplier method is used to solve the minimum point of $\omega_{i,j}$ for this upper bound. The optimal solution is

$$\omega_{i,j} = \frac{(U_{(i+j)} - U_{(i-j)})^2}{\sum_{c=1}^k (U_{(i+c)} - U_{(i-c)})^2}.$$

When the asymptotic conditional bias and variance are calculated, the following lemma on the uniform convergence is used.

Lemma 1. *Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be i.i.d random vectors, where the Y_i 's are scalar random variables. Assume further that $E|y|^3 < \infty$ and $\sup_x \int |y|^s f(x, y) dy < \infty$, where f denotes the joint density of (X, Y) . Let K be a bounded positive function with a bounded support, satisfying a Lipschitz condition. Then,*

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

provided that $n^{2\epsilon-1}h \rightarrow \infty$ for some $\epsilon < 1 - s^{-1}$.

Proof. This follows immediately from the result obtained by Mack and Silverman [28].

Lemma 2. *Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be i.i.d random vectors, where the Y_i 's are scalar random variables, X is 2-dimention random vector with the continuous marginal density function $f(x)$, and $f(x)$ has a compact support. Besides, $nh^4 > n^{\rho_1}$, for some $\rho_1 > 0$, $n^{-\rho_2}h^{-1} = o_p(h^2)$, for some $\rho_2 > 0$. Then*

$$\sup_{t \in R^2} \left| \frac{1}{nh^2} \sum_{i=1}^n Y_i K\left(\frac{t_1 - X_{i1}}{h}\right) K\left(\frac{t_2 - X_{i2}}{h}\right) - \frac{1}{h^2} EY_1 K\left[\left(\frac{s_1 - X_{i1}}{h}\right) K\left(\frac{s_2 - X_{i2}}{h}\right)\right] \right| \rightarrow 0,$$

completely as $n \rightarrow 0$, where $t = (t_1, t_2)$.

Proof. This follows immediately from the result obtained by Cheng and Taylor [29].

Lemma 3. *Let $\mathbf{U} = \{U_1, \dots, U_n\}$ be a sample generated from the random variable U , where U follows the standard uniform distribution. Also, let $U_{(1)}, \dots, U_{(n)}$ be the corresponding order statistics. Then, we have:*

$$\begin{aligned} U_{(i+i)} - U_{(i-j)} &= E(U_{(i+j)} - U_{(i-j)}) + O_p \left\{ \sqrt{Var(U_{(i+j)} - U_{(i-j)})} \right\} \\ &= \frac{2j}{n+1} + O_p \left(\sqrt{\frac{j}{n^2}} \right). \\ U_{(i+i)} - U_{(i)} &= E(U_{(i+j)} - U_{(i)}) + O_p \left\{ \sqrt{Var(U_{(i+j)} - U_{(i)})} \right\} \\ &= \frac{j}{n+1} + O_p \left(\sqrt{\frac{j}{n^2}} \right). \end{aligned}$$

Proof. This follows immediately from the result obtained by David and Nagaraja [30].

A3. Proof of Theorem 1.

For the initial estimator $\hat{a}_{q,0}(U_{(b)})$ of the two-step estimator, we have

$$\hat{a}_{q,0}(U_i) = (\mathbf{X}_{(b)}^T W_{(b)} \mathbf{X}_{(b)})^{-1} \mathbf{X}_{(b)}^T W_{(b)} \mathbf{Y}.$$

Note that by Taylor's expansion, we have

$$a_p(U_i) = a_p(U_{(b)}) + a_p^{(1)}(U_{(b)})(U_i - U_{(b)}) + \frac{1}{2} a_p^{(2)}(\eta_{i,(b)})(U_i - U_{(b)})^2,$$

where $\eta_{i,(b)}$ is between U_i and $U_{(b)}$ for $i = 1, \dots, n$.

Therefore,

$$\begin{aligned} & \hat{a}_p(U_{(b)}) \\ &= (1, 0)(\mathbf{X}_{2,(b)}^T W_{2,(b)} \mathbf{X}_{2,(b)})^{-1} \mathbf{X}_{2,(b)}^T W_{2,(b)} \begin{pmatrix} Y_1 - \sum_{q=1}^{p-1} \hat{a}_{q,0}(U_1) X_{q1} \\ \vdots \\ Y_n - \sum_{q=1}^{p-1} \hat{a}_{q,0}(U_n) X_{qn} \end{pmatrix} \\ &= (1, 0)(\mathbf{X}_{2,(b)}^T W_{2,(b)} \mathbf{X}_{2,(b)})^{-1} \mathbf{X}_{2,(b)}^T W_{2,(b)} \\ & \quad \times \begin{pmatrix} \sum_{q=1}^{p-1} (a_q(U_1) - \hat{a}_{q,0}(U_1)) X_{1q} + a_p(U_1) X_{1p} + \varepsilon_1 \\ \vdots \\ \sum_{q=1}^{p-1} (a_q(U_n) - \hat{a}_{q,0}(U_n)) X_{nq} + a_p(U_1) X_{np} + \varepsilon_n \end{pmatrix} \\ &= a_p(U_b) + \frac{1}{2} J_1 + J_2 + (1, 0)(\mathbf{X}_{2,(b)}^T W_{2,(b)} \mathbf{X}_{2,(b)})^{-1} \mathbf{X}_{2,(b)}^T W_{2,(b)} \boldsymbol{\varepsilon}, \end{aligned}$$

where

$$\begin{aligned} J_1 &= (1, 0)(\mathbf{X}_{2,(b)}^T W_{2,(b)} \mathbf{X}_{2,(b)})^{-1} \mathbf{X}_{2,(b)}^T W_{2,(b)} \\ & \quad \times \begin{pmatrix} a_p^{(2)}(\eta_{1,(b)})(U_1 - U_{(b)})^2 X_{1p} \\ \vdots \\ a_p^{(2)}(\eta_{n,(b)})(U_n - U_{(b)})^2 X_{np} \end{pmatrix}, \end{aligned}$$

$$\begin{aligned} J_2 &= (1, 0)(\mathbf{X}_{2,(b)}^T W_{2,(b)} \mathbf{X}_{2,(b)})^{-1} \mathbf{X}_{2,(b)}^T W_{2,(b)} \\ & \quad \times \begin{pmatrix} \sum_{q=1}^{p-1} (a_q(U_1) - \hat{a}_{q,0}(U_1)) X_{1q} \\ \vdots \\ \sum_{q=1}^{p-1} (a_q(U_n) - \hat{a}_{q,0}(U_n)) X_{nq} \end{pmatrix}. \end{aligned}$$

According to the Lemma 1 and continuity assumption for $a_j^{(2)}(\cdot)$, we can obtain

$$\mathbf{X}_{2,(b)}^T W_{2,(b)} \mathbf{X}_{2,(b)} = n r_{pp}(U_{(b)}) Q \begin{pmatrix} \mu_0 & 0 \\ 0 & \mu_2 \end{pmatrix} Q (1 + o_p(1)), \quad (\text{A.1})$$

$$\begin{aligned}
& \mathbf{X}_{2,(i+j)}^T W_{2,(b)} \begin{pmatrix} a_p^{(2)}(\eta_{1,(b)})(U_1 - U_{(b)})^2 X_{1p} \\ \vdots \\ a_p^{(2)}(\eta_{n,(b)})(U_n - U_{(b)})^2 X_{np} \end{pmatrix} \\
&= nh_2^2 r_{pp}(U_{(b)}) a_p(U_{(b)}) Q \begin{pmatrix} \mu_2 \\ 0 \end{pmatrix} (1 + o_p(1)). \tag{A.2}
\end{aligned}$$

By substituting (A.1) and (A.2) into J_1 , we have

$$E(J_1|\mathcal{D}) = h_2^2 \mu_2 a_p^{(2)}(U_{(b)}) (1 + o_p(1)). \tag{A.3}$$

Similarly, we have

$$E(J_2|\mathcal{D}) = -\frac{h_1^2 \mu_2}{2 r_{pp}(U_{(b)})} \sum_{q=1}^{p-1} a_q^{(2)}(U_{(b)}) r_{qp}(U_{(b)}) (1 + o_p(1)). \tag{A.4}$$

By combining (A.3) and (A.4), we have

$$E(\hat{a}_p(U_{(b)})|\mathcal{D}) = a_p(U_{(b)}) + \frac{1}{2} h_2^2 \mu_2 a_p^{(2)}(U_{(b)}) + o_p(h_2^2). \tag{A.5}$$

According to the standard uniform distribution property in Lemma 3, we have

$$\begin{aligned}
& \left| E \left(\sum_{j=1}^k \omega_{i,j} \left(\frac{a_p(U_{(i+j)}) - a_p(U_{(i-j)})}{U_{(i+j)} - U_{(i-j)}} \right) \middle| \mathcal{D} \right) - a_p^{(1)}(U_{(i)}) \right| \\
&= \frac{1}{2} \left| \sum_{j=1}^k \omega_{i,j} \frac{(U_{(i+j)} - U_{(i)})^2 a_p^{(2)}(\eta_{i,(i+j)}) - (U_{(i-j)} - U_{(i)})^2 a_p^{(2)}(\eta_{i,(i-j)})}{U_{(i+j)} - U_{(i-j)}} \right| \\
&\leq \frac{1}{2} \sup_{u \in [0,1]} |a_p^{(2)}(u)| \frac{(U_{(i+j)} - U_{(i-j)}) \left[(U_{(i+j)} - U_{(i)})^2 - (U_{(i-j)} - U_{(i)})^2 \right]}{\sum_{l=1}^k U_{(i+l)} - U_{(i-l)}} \\
&= \sup_{u \in [0,1]} |a_p^{(2)}(u)| \frac{3k(k+1)}{4(n+1)(2k+1)} \left\{ 1 + O_p \left(\frac{1}{\sqrt{k}} \right) \right\} \tag{A.6}
\end{aligned}$$

According to the (A.5) and (A.6), we can obtain an upper bound of the absolute conditional bias

$$\begin{aligned}
\left| Bias \left[\hat{a}_p^{(1)}(U_{(i)}) \middle| \mathcal{D} \right] \right| &= \left| E \left(\sum_{j=1}^k \omega_{i,j} \left(\frac{\hat{a}_p(U_{(i+j)}) - \hat{a}_p(U_{(i-j)})}{U_{(i+j)} - U_{(i-j)}} \right) \middle| \mathcal{D} \right) - a_p^{(1)}(U_{(i)}) \right| \\
&= \left| \sum_{j=1}^k \omega_{i,j} \left\{ \frac{a_p(U_{(i+j)}) - a_p(U_{(i-j)})}{U_{(i+j)} - U_{(i-j)}} \right. \right. \\
&\quad \left. \left. + \frac{h_2^2 \mu_2 (a_p^{(2)}(U_{(i+j)}) - a_p^{(2)}(U_{(i-j)}))}{2(U_{(i+j)} - U_{(i-j)})} + o_p(h_2^2) \right\} - a_p^{(1)}(U_{(i)}) \right| \\
&\leq \sup_{u \in [0,1]} |a_p^{(2)}(u)| \left\{ \frac{3k(k+1)}{4(n+1)(2k+1)} + h_2^2 \mu_2 \frac{3(n+1)}{2(2k+1)} \right\} (1 + o_p(1)).
\end{aligned}$$

Then, we deal with conditional variance. First, for $H_{(b)}H_{(c)}^T$, we have

$$\begin{aligned} & H_{(b)}H_{(c)}^T \\ = & e_{1,2}^T(\mathbf{X}_{2,(b)}^T W_{2,(b)} \mathbf{X}_{2,(b)})^{-1} \mathbf{X}_{2,(b)}^T \\ & \times \left\{ W_{2,(b)} W_{2,(c)} - 2W_{2,(b)} B_n W_{2,(c)} + W_{2,(b)} B_n B_n^T W_{2,(c)} \right\} \\ & \times \mathbf{X}_{2,(c)} (\mathbf{X}_{2,(c)}^T W_{2,(c)} \mathbf{X}_{2,(c)})^{-1} e_{1,2}. \end{aligned}$$

According to Lemma 2, we can obtain

$$\begin{aligned} & \mathbf{X}_{2,(b)}^T W_{2,(b)} W_{2,(c)} \mathbf{X}_{2,(c)} \\ = & nr_{pp}(U_{(b)}) K_{h_2}(U_{(b)} - U_{(c)}) \\ \times & Q \left(\frac{\mu_0}{U_{(b)} - U_{(c)} + o_p(1)} \frac{o_p(1)}{\mu_2 + \frac{(U_{(b)} - U_{(c)})^2}{h_2} o_p(1) + o_p(1)} \right) Q(1 + o_p(1)), \end{aligned} \quad (\text{A.7})$$

and

$$\begin{aligned} & B_n W_{2,(c)} \mathbf{X}_{2,(c)} \\ = & \sum_{q=1}^{p-1} \begin{pmatrix} X_{1q} K_{h_2}(U_1 - U_c) e_{2q-1,2p}^T (\Omega_p^{-1}(U_1) \otimes A^{-1}) C_1 Q \\ \vdots \\ X_{nq} K_{h_2}(U_n - U_c) e_{2q-1,2p}^T (\Omega_p^{-1}(U_n) \otimes A^{-1}) C_n Q \end{pmatrix} (1 + o_p(1)). \end{aligned} \quad (\text{A.8})$$

where

$$C_l = \begin{pmatrix} r_{1p}(U_l) & r_{1p}(U_l) \frac{U_l - U_{(c)}}{h_2} + o_p(1) \\ o_p(1) & o_p(1) \frac{U_l - U_{(c)}}{h_2} + o_p(1) \\ \vdots & \vdots \\ r_{pp}(U_l) & r_{pp}(U_l) \frac{U_l - U_{(c)}}{h_2} + o_p(1) \\ o_p(1) & o_p(1) \frac{U_l - U_{(c)}}{h_2} + o_p(1) \end{pmatrix}.$$

Notice, $e_{2q-1,2p}^T (\Omega_p^{-1}(U_l) \otimes A^{-1}) C_l = (0, o_p(1))$. Therefore, we have

$$e_{1,2}^T (\mathbf{X}_{2,(b)}^T W_{2,(b)} \mathbf{X}_{2,(b)})^{-1} \mathbf{X}_{2,(b)}^T W_{2,(b)} B_n W_{2,(c)} \mathbf{X}_{2,(c)} (\mathbf{X}_{2,(c)}^T W_{2,(c)} \mathbf{X}_{2,(c)})^{-1} e_{1,2} = o_p(1).$$

Let

$$\mathbf{X}_{2,(b)}^T W_{2,(b)} B_n B_n^T W_{2,(c)} \mathbf{X}_{2,(c)} = (\tau_{rs})_{2 \times 2}.$$

We expand the τ_{rs} , we have

$$\begin{aligned} & \tau_{rs} \\ = & \sum_{t=1}^n \sum_{l=1}^n \left\{ X_{tp}(U_t - U_{(b)})^r K_{h_2}(U_t - U_{(b)}) \left[\sum_{q=1}^{p-1} X_{tq} e_{2q-1,2p}^T (\mathbf{X}_{0,t}^T W_{0,t} \mathbf{X}_{0,t})^{-1} \mathbf{X}_{0,t}^T W_{0,t} \right] \right\} \end{aligned}$$

$$\begin{aligned}
& \times \left\{ X_{lp}(U_l - U_{(c)})^s K_{h_2}(U_l - U_{(c)}) \left[\sum_{q=1}^{p-1} X_{lq} e_{2q-1,2p}^T (\mathbf{X}_{0,l}^T W_{0,l} \mathbf{X}_{0,l})^{-1} \mathbf{X}_{0,l}^T W_{0,l} \right]^T \right\} \\
= & \sum_{q=1}^{p-1} \sum_{d=1}^{p-1} \sum_{a=1}^n \left\{ \left[\sum_{t=1}^n X_{tq} X_{tp}(U_t - U_{(b)})^r K_{h_2}(U_t - U_{(b)}) \right. \right. \\
& \times e_{2q-1,2p}^T (\mathbf{X}_{0,t}^T W_{0,t} \mathbf{X}_{0,t})^{-1} X_{a(t)} K_{h_1}(U_a - U_t) \\
& \left. \left. \times \left[\sum_{l=1}^n X_{ld} X_{lp}(U_l - U_{(c)})^s K_{h_2}(U_l - U_{(c)}) e_{2d-1,2p}^T (\mathbf{X}_{0,l}^T W_{0,l} \mathbf{X}_{0,l})^{-1} K_{h_1}(U_a - U_l) \right] \right\} \right].
\end{aligned}$$

where

$$X_{a(t)} = (X_{a1}, X_{a1}(U_a - U_t), \dots, X_{ap}, X_{ap}(U_a - U_t))^T.$$

Then, using Lemma 1, we have

$$\begin{aligned}
& \sum_{t=1}^n X_{tq} X_{tp}(U_t - U_{(b)})^r K_{h_2}(U_t - U_{(b)}) e_{2q-1,2p}^T (\mathbf{X}_{0,t}^T W_{0,t} \mathbf{X}_{0,t})^{-1} X_{a(t)} K_{h_1}(U_a - U_t) \\
= & r_{qp}(U_{(b)}) (K_{h_1}(U_a - U_{(b)}) e_{2q-1,2p}^T G^{-1} (\Omega_p^{-1}(U_{(b)}) \otimes A^{-1}) G^{-1} \\
& \times \begin{pmatrix} X_{a1} \\ X_{a2} \\ \vdots \\ X_{ap} \end{pmatrix} \otimes \begin{pmatrix} h_2^r \mu_r \\ h_2^r \mu_r(U_a - U_{(b)}) + h_2^{r+1} \mu_{r+1} \end{pmatrix} (1 + o_p(1))). \tag{A.9}
\end{aligned}$$

Substituting (A.9) into τ_{rs} , we have

$$\begin{aligned}
& \tau_{rs} \\
= & \sum_{q=1}^{p-1} \sum_{d=1}^{p-1} \left\{ \sum_{a=1}^n r_{dp}(U_{(c)}) r_{qp}(U_{(b)}) K_{h_1}(U_a - U_{(b)}) K_{h_1}(U_a - U_{(c)}) \right. \\
& \times e_{2q-1,2p}^T (\Omega_p^{-1}(U_{(b)}) \otimes A^{-1}) G^{-1} \\
& \times \left\{ \begin{pmatrix} X_{a1} & X_{a2} & \dots & X_{ap} \end{pmatrix}^T \begin{pmatrix} X_{a1} & X_{a2} & \dots & X_{ap} \end{pmatrix} \otimes C_{rs} \right\} \\
& \left. \times G^{-1} (\Omega_p^{-1}(U_{(c)}) \otimes A^{-1}) e_{2d-1,2p} \right\} (1 + o_p(1)).
\end{aligned}$$

where

$$C_{rs} = (h_2^r \mu_r \quad h_2^r \mu_r(U_a - U_{(b)}) + h_2^{r+1} \mu_{r+1})^T (h_2^r \mu_r \quad h_2^r \mu_r(U_a - U_{(b)}) + h_2^{r+1} \mu_{r+1}).$$

According to Lemma 2 and the properties of the Kronecker product, we have

$$\tau_{rs}$$

$$\begin{aligned}
&= n \sum_{q=1}^{p-1} \sum_{d=1}^{p-1} h_2^{r+s} \mu_r \mu_s r_{qp}(U_{(b)}) r_{dp}(U_{(c)}) K_{h_1}(U_{(b)} - U_{(c)}) e_{q,p}^T \Omega_p^{-1}(U_{(b)}) e_{d,p} (1 + o_p(1)) \\
&= -n h_2^{r+s} \mu_r \mu_s r_{pp}(U_{(b)}) K_{h_1}(U_{(b)} - U_{(c)}) \\
&\quad \times \sum_{d=1}^{p-1} r_{dp}(U_{(c)}) e_{d,p}^T \Omega_p^{-1}(U_{(b)}) e_{p,p} (1 + o_p(1)). \tag{A.10}
\end{aligned}$$

Substituting (A.7), (A.8), and (A.10) into the expansion of $H_{(b)} H_{(c)}^T$, we have

$$\begin{aligned}
&H_{(b)} H_{(c)}^T \\
&= \frac{1}{nr_{pp}(U_{(c)})} \left\{ K_{h_2}(U_{(b)} - U_{(c)}) \right. \\
&\quad \left. - K_{h_1}(U_{(b)} - U_{(c)}) \sum_{d=1}^{p-1} r_{dp}(U_{(c)}) e_{d,p}^T \Omega_p^{-1}(U_{(b)}) e_{p,p} \right\} (1 + o_p(1)) \\
&\leq \frac{1}{nr_{pp}^-} \left\{ \frac{M}{h_2} + \frac{M}{h_1} \sum_{d=1}^{p-1} r_{dp}^+ r_{dp}^* \right\} (1 + o_p(1)) \tag{A.11}
\end{aligned}$$

Now, we provide an upper bound for $Var[\hat{a}_p^{(1)}(U_{(i)})|\mathcal{D}]$. First, the $Var[\hat{a}_p^{(1)}(U_{(i)})|\mathcal{D}]$ is expanded by

$$\begin{aligned}
&Var[\hat{a}_p^{(1)}(U_{(i)})|\mathcal{D}] \\
&= Var \left[\sum_{j=1}^k \omega_{i,j} \left(\frac{\hat{a}_p(U_{(i+j)}) - \hat{a}_p(U_{(i-j)})}{U_{(i+j)} - U_{(i-j)}} \right) |\mathcal{D} \right] \\
&= Var \left[\sum_{j=1}^k \frac{\omega_{i,j}}{U_{(i+j)} - U_{(i-j)}} \{H_{(i+j)} - H_{(i-j)}\} \varepsilon |\mathcal{D} \right] \\
&= \sigma^2 \sum_{j=1}^k \sum_{m=1}^k \frac{\omega_{i,j} \omega_{i,m}}{(U_{(i+j)} - U_{(i-j)})(U_{(i+m)} - U_{(i-m)})} \\
&\quad \times \{H_{(i+j)} - H_{(i-j)}\} \{H_{(i+m)} - H_{(i-m)}\}^T \\
&= \sigma^2 \sum_{j=1}^k \sum_{m=1}^k \frac{\omega_{i,j} \omega_{i,m}}{(U_{(i+j)} - U_{(i-j)})(U_{(i+m)} - U_{(i-m)})} \\
&\quad \times \{H_{(i+j)} H_{(i+m)}^T - H_{(i-j)} H_{(i+m)}^T - H_{(i+j)} H_{(i-m)}^T + H_{(i-j)} H_{(i-m)}^T\}.
\end{aligned}$$

By substituting (A.11) into the above expansion, we can obtain the upper bound for $Var[\hat{a}_p^{(1)}(U_{(i)})|\mathcal{D}]$,

$$\begin{aligned}
Var[\hat{a}_p^{(1)}(U_{(i)})|\mathcal{D}] &\leq \frac{M\sigma^2}{nr_{pp}^-} \left\{ \sum_{j=1}^k \sum_{m=1}^k \frac{\omega_{i,j} \omega_{i,m}}{(U_{(i+j)} - U_{(i-j)})(U_{(i+m)} - U_{(i-m)})} \right\} \\
&\quad \times \left\{ \frac{1}{h_2} + \frac{1}{h_1} \sum_{d=1}^{p-1} r_{dp}^+ r_{dp}^* \right\} (1 + o_p(1))
\end{aligned}$$

$$= \frac{9(n+1)^2 M \sigma^2}{(2k+1)^2 n r_{pp}^-} \left\{ \frac{1}{h_2} + \frac{1}{h_1} \sum_{d=1}^{p-1} r_{dp}^+ r_{dp}^* \right\} (1 + o_p(1)).$$

A4. Proof of Corollary 1.

$$\begin{aligned} & a_p(U_{(i+j)}) - a_p(U_{(i-j)}) \\ = & a_p^{(1)}(U_{(i)}) \left(U_{(i+j)} - U_{(i-j)} \right) \\ & + \frac{1}{2} a_p^{(2)}(U_{(i)}) \left\{ \left(U_{(i+j)} - U_{(i)} \right)^2 - \left(U_{(i-j)} - U_{(i)} \right)^2 \right\} + O_p \left(\frac{j^2}{n^2} \right). \end{aligned} \quad (\text{A.12})$$

According to the upper bound of the absolute conditional bias and (A.12), we have

$$\begin{aligned} \left| \text{Bias} \left[\hat{a}_p^{(1)}(U_{(i)}) \mid \mathcal{D} \right] \right| &= \left| \sum_{j=1}^k \omega_{i,j} \left\{ \frac{a_p(U_{(i+j)}) - a_p(U_{(i-j)})}{U_{(i+j)} - U_{(i-j)}} \right. \right. \\ &\quad \left. \left. + \frac{h_2^2 \mu_2 (a_p^{(2)}(U_{(i+j)}) - a_p^{(2)}(U_{(i-j)}))}{2(U_{(i+j)} - U_{(i-j)})} + o_p(h_2^2) \right\} - a_p^{(1)}(U_{(i)}) \right| \\ &= \left| \left\{ \frac{a_p^{(1)}(U_{(i)}) \sum_{j=1}^k (U_{(i+j)} - U_{(i-j)})^2 + O_p(k^4/n^3)}{\sum_{j=1}^k (U_{(i+j)} - U_{(i-j)})^2} \right. \right. \\ &\quad \left. \left. + \sum_{j=1}^k \omega_{i,j} \frac{h_2^2 \mu_2 (a_p^{(2)}(U_{(i+j)}) - a_p^{(2)}(U_{(i-j)}))}{2(U_{(i+j)} - U_{(i-j)})} + o_p(h_2^2) \right\} - a_p^{(1)}(U_{(i)}) \right| \\ &= O_p \left(\max \left(h_2^2, \frac{k}{n} \right) \right). \end{aligned}$$

According to the upper bound of the absolute conditional variance, we have

$$\begin{aligned} & \text{Var} \left[\hat{a}_p^{(1)}(U_{(i)}) \mid \mathcal{D} \right] \\ = & \sigma^2 \sum_{j=1}^k \sum_{m=1}^k \frac{\omega_{i,j} \omega_{i,m}}{(U_{(i+j)} - U_{(i-j)})(U_{(i+m)} - U_{(i-m)})} \\ & \times \left\{ H_{(i+j)} H_{(i+m)}^T - H_{(i-j)} H_{(i+m)}^T - H_{(i+j)} H_{(i-m)}^T + H_{(i-j)} H_{(i-m)}^T \right\} \\ = & O_p \left(\frac{n}{k^2 h_1} \right). \end{aligned}$$

Thus, Corollary 1 is proven.

A4.1. Proof of Proposition 3.

$$\text{Var} \left[\hat{a}_p^{(2)}(U_{(i)}) \mid \mathcal{D} \right]$$

$$\begin{aligned}
&= \text{Var} \left[\sum_{j=1}^{k_2} \omega_{i,j,2} \left(\frac{\hat{a}_p(U_{(i+j+k_1)}) - \hat{a}_p(U_{(i+j)})}{U_{(i+j+k_1)} - U_{(i+j)}} - \frac{\hat{a}_p(U_{(i-j-k_1)}) - \hat{a}_p(U_{(i-j)})}{U_{(i-j-k_1)} - U_{(i-j)}} \right) \middle| \mathcal{D} \right] \\
&= \sigma^2 \left\{ \sum_{j=1}^{k_2} \frac{\omega_{i,j,2}}{(U_{(i+j+k_1)} + U_{(i+j)} - U_{(i-j-k_1)} - U_{(i-j)})} \right. \\
&\quad \times \left\{ \frac{H_{(i+j+k_1)} - H_{(i+j)}}{U_{(i+j+k_1)} - U_{(i+j)}} - \frac{H_{(i-j-k_1)} - H_{(i-j)}}{U_{(i-j-k_1)} - U_{(i-j)}} \right\} \\
&\quad \times \left\{ \sum_{j=1}^{k_2} \frac{\omega_{i,j,2}}{(U_{(i+j+k_1)} + U_{(i+j)} - U_{(i-j-k_1)} - U_{(i-j)})} \right. \\
&\quad \times \left. \left\{ \frac{H_{(i+j+k_1)} - H_{(i+j)}}{U_{(i+j+k_1)} - U_{(i+j)}} - \frac{H_{(i-j-k_1)} - H_{(i-j)}}{U_{(i-j-k_1)} - U_{(i-j)}} \right\} \right\}^T. \tag{A.13}
\end{aligned}$$

Let

$$\xi_j = \frac{\left\{ \frac{H_{(i+j+k_1)} - H_{(i+j)}}{U_{(i+j+k_1)} - U_{(i+j)}} - \frac{H_{(i-j-k_1)} - H_{(i-j)}}{U_{(i-j-k_1)} - U_{(i-j)}} \right\}}{(U_{(i+j+k_1)} + U_{(i+j)} - U_{(i-j-k_1)} - U_{(i-j)})},$$

and $\xi = (\xi_1, \dots, \xi_{k_2})^T$. We can obtain

$$\text{Var} \left[\hat{a}_p^{(2)}(U_{(i)}) \middle| \mathcal{D} \right] = \sigma^2 \boldsymbol{\omega}_{i,2}^T \xi \xi^T \boldsymbol{\omega}_{i,2}.$$

Then, let $\Sigma_0 = \xi \xi^T$, thus Proposition 3 is proven.

A5. Proof of Theorem 2.

Similar to the proof of Theorem 1, we can obtain the upper bound of the conditional bias of $\hat{a}_p^{(2)}(U_i)$

$$\begin{aligned}
&\left| \text{Bias} \left[\hat{a}_p^{(2)}(U_i) \middle| \mathcal{D} \right] \right| \\
&= \left| E \left(\sum_{j=1}^{k_2} \omega_{i,j,2} \left(\frac{\hat{a}_p(U_{(i+j+k_1)}) - \hat{a}_p(U_{(i+j)})}{U_{(i+j+k_1)} - U_{(i+j)}} - \frac{\hat{a}_p(U_{(i-j-k_1)}) - \hat{a}_p(U_{(i-j)})}{U_{(i-j-k_1)} - U_{(i-j)}} \right) \middle| \mathcal{D} \right) - a_p^{(2)}(U_{(i)}) \right| \\
&= \left| \sum_{j=1}^{k_2} \omega_{i,j,2} \left(\frac{a_p(U_{(i+j+k_1)}) - a_p(U_{(i+j)})}{U_{(i+j+k_1)} - U_{(i+j)}} - \frac{a_p(U_{(i-j-k_1)}) - a_p(U_{(i-j)})}{U_{(i-j-k_1)} - U_{(i-j)}} \right) \right. \\
&\quad \left. + \frac{1}{2} h_2^2 \mu_2 \frac{a_p^{(2)}(U_{(i+j+k_1)}) - a_p^{(2)}(U_{(i+j)})}{U_{(i+j+k_1)} - U_{(i+j)}} - \frac{a_p^{(2)}(U_{(i-j-k_1)}) - a_p^{(2)}(U_{(i-j)})}{U_{(i-j-k_1)} - U_{(i-j)}} \right) + o_p \left(\frac{nh_2^2}{k_1^2} \right) - a_p^{(2)}(U_{(i)}) \right| \\
&\leq \sup_{u \in [0,1]} |a_p^{(3)}(u)| \sum_{j=1}^{k_2} \omega_{i,j,2} \left\{ \frac{j^2 + jk_1 + \frac{1}{3}k_1^2}{(n+1)(2j+k_1)} + h_2^2 \mu_2 \frac{n+1}{4j+2k_1} \right\} (1 + o_p(1)).
\end{aligned}$$

Similar to the proof of the upper bound of the conditional variance in Theorem 1, we can substitute (A.11) into (A.13), and we can have

$$\text{Var} \left[\hat{a}_p^{(2)}(U_{(i)}) \middle| \mathcal{D} \right]$$

$$\begin{aligned}
&\leq \sum_{j=1}^{k_2} \sum_{m=1}^{k_2} \left\{ \frac{\omega_{i,j,2}}{(U_{(i+j+k_1)} + U_{(i+j)} - U_{(i-j-k_1)} - U_{(i-j)})} \right. \\
&\quad \times \frac{\omega_{i,m,2}}{(U_{(i+m+k_1)} + U_{(i+m)} - U_{(i-m-k_1)} - U_{(i-m)})} \Big\} \\
&\quad \times \frac{4M\sigma^2}{nr_{pp}^-} \frac{(n+1)^2}{k_1^2} \left\{ \frac{2}{h_2} + \frac{2}{h_1} \left(2 \sum_{d=1}^{p-1} r_{dp}^+ r_{dp}^* + 2 \sum_{d=1}^{p-1} \mathcal{R}_{dp} \mathcal{R}_{dp}^* \right) \right\} (1 + o_p(1)) \\
&= \frac{4(n+1)^4 M \sigma^2}{k_1^2 nr_{pp}^-} \left\{ \frac{2}{h_2} + \frac{2}{h_1} \left(\sum_{d=1}^{p-1} r_{dp}^+ r_{dp}^* + \sum_{d=1}^{p-1} \mathcal{R}_{dp} \mathcal{R}_{dp}^* \right) \right\} \\
&\quad \times \left\{ \sum_{j=1}^{k_2} \sum_{m=1}^{k_2} \frac{\omega_{i,j,2} \omega_{i,m,2}}{(2j+k_1)(2m+k_1)} \right\} (1 + o_p(1)).
\end{aligned}$$

Thus, Theorem 2 is proven.

A6. Proof of Corollary 3.

According to the proof of Theorem 2, we can immediately obtain Corollary 3.

Author contributions

Junfeng Huo proposed the methodology, proved the theoretical properties, implemented the code, performed part of the numerical simulation, and was primarily responsible for writing and revising the manuscript. Mingquan Wang contributed to the numerical simulation and participated in the writing and revision parts of the manuscript. Xiuqing Zhou contributed to the formulation of the research topic and provided critical revisions of the manuscript.

Use of Generative-AI tools declaration

The authors declare that generative AI tools were used to refine the language of sentences in the Introduction section of this article. Specifically, ChatGPT was used to improve language clarity and readability. All other content, including research design, data analysis, and conclusions, was independently developed by the authors without AI assistance. The authors bear full responsibility for the integrity and accuracy of the manuscript.

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. B. R. H. Shumway, *Applied Statistical Time Series Analysis*, New Jersey: Prentice Hall, 1988. <http://dx.doi.org/10.1007/978-1-4757-3261-0>

2. T. J. Hastie, R. J. Tibshirani, Varying-coefficient models, *J. Roy. Statist. Soc. Ser. B*, **4** (1993), 757–796. <https://doi.org/10.1111/j.2517-6161.1993.tb01939.x>
3. J. Fan, W. Zhang, Statistical estimation in varying coefficient models, *Ann. Statist.*, **27** (1999), 1491–1518. <https://doi.org/10.1214/aos/1017939139>
4. J. Z. Huang, H. Shen, Functional coefficient regression models for non-linear time series: A polynomial spline approach, *Scand. J. Statist.*, **31** (2004), 515–534. <https://doi.org/10.1111/j.1467-9469.2004.00404.x>
5. Z. Cai, J. Fan, R. Li, Efficient estimation and inferences for varying-coefficient models, *J. Amer. Statist. Assoc.*, **95** (2000), 888–902. <https://doi.org/10.1080/01621459.2000.10474280>
6. T. V. Somanathan, V. A. Nageswaran, *The economics of derivatives*, Cambridge University Press, London, 2015. <https://doi.org/10.1017/CBO9781316134566>
7. J. Fan, W. Zhang, Statistical methods with varying coefficient models, *Stat. Interface*, **2008** (2008), 179–195. <https://doi.org/10.4310/SII.2008.v1.n1.a15>
8. C. T. Chiang, J. A. Rice, C. O. Wu, Smoothing spline estimation for varying coefficient models with repeatedly measured dependent variables, *J. Amer. Statist. Assoc.*, **96** (2001), 605–619. <https://doi.org/10.1198/016214501753168280>
9. W. S. Cleveland, S. J. Devlin, Locally weighted regression: an approach to regression analysis by local fitting, *J. Amer. Statist. Assoc.*, **83** (1988), 596–610. <https://doi.org/10.1080/01621459.1988.10478639>
10. J. Fan, I. Gijbels, *A Local polynomial modelling and its applications*, Chapman & Hall, London, 1996. <https://doi.org/10.1201/9780203748725>
11. P. J. Green, B. W. Silverman, *Nonparametric regression and generalized linear models*, Chapman & Hall, London, 1994. <http://dx.doi.org/10.1007/978-1-4899-4473-3>
12. H. Wang, X. Zhou, Explicit estimation of derivatives from data and differential equations by Gaussian process regression, *Int. J. Uncertain. Quantif.*, **11** (2021), 41–57. <https://doi.org/10.1615/Int.J.UncertaintyQuantification.2021034382>
13. Z. Liu, M. Li, On the estimation of derivatives using plug-in kernel ridge regression estimators, *J. Mach. Learn. Res.*, **24** (2023), 1–37. <https://doi.org/10.1007/s00180-007-0102-8>
14. H. G. Müller, U. Stadtmüller, T. Schmitt, Bandwidth choice and confidence intervals for derivatives of noisy data, *Biometrika*, **74** (1987), 743–749. <https://doi.org/10.1093/biomet/74.4.743>
15. R. Charnigo, B. Hall, C. Srinivasan, A generalized Cp criterion for derivative estimation, *Technometrics*, **53** (2011), 238–253. <https://doi.org/10.1198/TECH.2011.09147>
16. K. D. Brabanter, J. D. Brabanter, B. D. Moor, I. Gijbels, Derivative estimation with local polynomial fitting, *J. Mach. Learn. Res.*, **14** (2013), 281–301. <https://doi.org/10.1002/cem.2489>
17. W. Wang, L. Lin, Derivative estimation based on difference sequence via locally weighted least squares regression, *J. Mach. Learn. Res.*, **16** (2015), 2617–2641. <https://doi.org/10.5555/2789272.2912083>
18. W. Dai, T. Tong, M. G. Genton, Optimal estimation of derivatives in nonparametric regression, *J. Mach. Learn. Res.*, **17** (2016), 1–37. <https://doi.org/10.5555/2946645.3053446>

19. Y. Liu, K. D. Brabanter, Smoothed nonparametric derivative estimation using weighted difference quotients, *J. Mach. Learn. Res.*, **21** (2020), 1–45. <https://doi.org/10.5555/3455716.3455781>
20. S. Liu, X. Kong, A generalized correlated C_p criterion for derivative estimation with dependent errors, *Comput. Statist. Data Anal.*, **21** (2022), 1–23. <https://doi.org/10.1016/j.csda.2022.107473>
21. J. Huo, X. Zhou, Efficient estimation in varying coefficient regression models, *J. Stat. Comput. Simul.*, **93** (2023), 2297–2320. <https://doi.org/10.1080/00949655.2023.2179626>
22. H. Wang, Y. Xia, Shrinkage estimation of the varying coefficient model, *J. Amer. Statist. Assoc.*, **104** (2009), 747–757. <https://doi.org/10.1198/jasa.2009.0138>
23. Y. Zhao, J. Lin, X. Huang, H. Wang, Adaptive jump-preserving estimates in varying-coefficient models, *J. Multivariate Anal.*, **149** (2016), 65–80. <https://doi.org/10.1016/j.jmva.2016.03.005>
24. E. Parzen, On estimation of a probability density function and mode, *Ann. Math. Statist.*, **33** (1962), 1065–1076. <https://doi.org/10.1214/aoms/1177704472>
25. J. Ormerod, M. P. Wand, I. Koch, Penalised spline support vector classifiers: computational issues, *Comput. Statist.*, **23** (2008), 623–641. <https://doi.org/10.1007/s00180-007-0102-8>
26. S. Kotz, N. Balakrishnan, C. B. Read, B. Vidakovic, N. L. Johnson, *Encyclopedia of Statistical Sciences, Volume 1*. John Wiley & Sons, 2005. <https://doi.org/10.1002/0471667196>
27. J. Fan, T. Huang, Profile likelihood inferences on semiparametric varying-coefficient partially linear models, *Bernoulli*, **11** (2005), 1031–1057. <https://doi.org/10.3150/bj/1137421639>
28. Y. P. Mack, B. W. Silverman, Weak and strong uniform consistency of kernel regression estimates, *Z. Wahrsch. Verw. Gebiete*, **61** (1982), 405–415. <https://doi.org/10.1007/BF00539840>
29. K. F. Cheng, R. L. Taylor, On the uniform complete convergence of estimates for multivariate density functions and regression curves, *Ann. Inst. Statist. Math.*, **32** (1980), 187–199. <https://doi.org/10.1007/BF00539840>
30. H. A. David, H. W. Nagaraja, *Order Statistics*, 3 Eds., John Wiley and Sons Inc, 2005. <http://dx.doi.org/10.1002/9781118445112.stat00830>



AIMS Press

© 2025 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>)