



Penalized Spline Estimation for Nonparametric Multiplicative Regression Models

Wei Chen^{1*}, Mingzhen Wan², Jiahui Xu², Jing Zhong², Yuying Xia², Minmin Zhang²

¹School of Zhangjiagang, Jiangsu University of Science and Technology, Zhangjiagang, China

²Suzhou Institute of Technology, Jiangsu University of Science and Technology, Zhangjiagang, China

Email: *chenweixiyang@sina.com.cn, 504951453@qq.com

How to cite this paper: Chen, W., Wan, M.Z., Xu, J.H., Zhong, J., Xia, Y.Y. and Zhang, M.M. (2024) Penalized Spline Estimation for Nonparametric Multiplicative Regression Models. *Open Access Library Journal*, **11**: e11352.
<http://doi.org/10.4236/oalib.1111352>

Received: February 26, 2024

Accepted: April 16, 2024

Published: April 19, 2024

Copyright © 2024 by author(s) and Open Access Library Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

In this paper, we consider the estimation problem of the unknown link function in the nonparametric multiplicative regression model. Combining the penalized splines technique, the least product relative error estimation method is proposed, where an effective model degree of freedom is defined, then the smoothing parameter is chosen by some information criteria. Simulation studies show that these strategies work well. Some asymptotic properties are established. A real data set is analyzed to illustrate the usefulness of the proposed approach. Finally, some possible extensions are discussed.

Subject Areas

Mathematical Statistics

Keywords

Multiplicative Regression, Nonparametric Estimation, Penalized Splines, Relative Error, Smoothing Parameter Selection

1. Introduction

In many applications, such as the studies of financial and biomedical data, the response variable usually is positive. For modelling the relationship between the positive response and a set of explanatory variables, a natural idea is that first take an appropriate transformation for the response, e.g., the logarithmic transformation, then some common regression models, such as the linear regression or quantile regression, can be employed based on the transformed data. As argued by [1], the least square or least absolute deviation criteria are both based on absolute errors, which is not desirable in many practical applications. Rather, the relative errors are more of a concern. In the early literature, many authors have

contributed fruitfully to this issue, see [2] [3] and [4]. Especially, since the work of [1], more attention has focused on the multiplicative regression (MR) model, and various extensions have been investigated, for example, see [5] [6] [7] and references therein.

It is worth noting that all the existing studies on MR are in the framework of linear (parametric) or semi-parametric MR models. Once the parametric forms or structures are misspecified, the resulting estimation and inference may be biased and the resulting conclusions become unreliable. Rather, nonparametric modelling is conceptually appealing and more robust. To the best of our knowledge, there are seldom studies on this problem. To fill this gap, we will address this problem in detail in this paper.

When estimating the nonparametric function $g(z)$ in some semi-parametric MR models, such as the partially linear MR model ([6] [8] [9]), single index MR model ([10] [11] [12]), varying coefficient MR model ([13]), and others ([14]), almost all the researchers use the local linear smoothing technique and approximate it in a neighborhood of z for obtaining its estimation, where a good choice of the bandwidth is quietly critical and its value is sensitive to the performance of the resulting estimation and inference. Besides, due to the fact that the value of the function at every observation of z is estimated separately, the optimal selection of bandwidth for all observations may be not the same and the numerical problem will become untractable then the sample size is large. As a result, the researchers had to make a compromise and assume that the bandwidths used for estimating $g(z)$ are the same.

In the nonparametric regression literature, spline-based methods, such as regression splines, smoothing splines, and penalized splines, are also popular and applied extensively in many fields. Recently, [15] proposed the multiplicative additive models based on the least product relative error criterion (LPRE), where the B-spline basis functions are used to estimate the nonparametric functions. Approximating a smooth function by a spline function has many desirable benefits, as presented in [16] and [17]. It is well known that the number and location of internal knots should be well-addressed when using the B-spline approximation approach. Too many knots may cause overfitting, few knots then bring underfitting. Although some information criteria can be adopted to select an appropriate number of knots, the overall computation burden is rather heavy. Fortunately, the penalized spline (P-splines) can avoid this difficulty and have gained remarkable popularity, especially due to computational expediency and easy adaptability to more complex models, see [18]. However, this technique has not been studied in the context of MR, and this is the first work to employ P-splines estimation for MR, although this issue is of importance and meaningful. All these motivate us to conduct this study in a formal manner.

This paper is organized as follows. In Section 2, we first introduce the nonparametric multiplicative regression model. Combining the penalized splines and the least product relative error criterion, a new estimation method is proposed, and some remarks about the selection of smoothing parameters and knots, and

the asymptotic properties of the proposed estimator are given. Some simulation studies are carried out to assess the performance of our method under finite sample situations in Section 3. To illustrate its usefulness, we also apply our method to one real data set in Section 4. Finally, some discussions in Section 5 conclude the paper.

2. Methodology

In this section, we mainly introduce the model and estimation approach. At the same time, some related issues will be answered in detail.

2.1. Model and Estimation

Consider the following nonparametric multiplicative regression model

$$Y = \exp(g(\mathbf{Z}))\varepsilon, \quad (1)$$

where Y is the response variable, \mathbf{Z} is a p -vector of covariates, ε is the positive random error and independent of \mathbf{Z} , $g(\cdot)$ is an unknown link function. Without loss of generality and for simplicity, we will assume that $p=1$ in the following, *i.e.*, the covariate \mathbf{Z} is univariate. But the method discussed later can be easily be extended to the general case. Assume that an i.i.d. sample $(Y_i, Z_i), i=1, \dots, n$ from model (1) are observed. For model identifiability, it is required that $E(\varepsilon) = E(\varepsilon^{-1})$ in [5]. On the other hand, after taking logarithmic transformation on both sides of Equation (1), it follows that

$$\ln(Y) = g(Z) + \ln(\varepsilon) = (g(Z) + c) + (\ln(\varepsilon) - c) \triangleq \tilde{g}(Z) + \ln(\tilde{\varepsilon}), \quad (2)$$

holds for any real number c , which means that the former requirement is not enough to ensure the nonparametric component unique. This phenomenon has been found by [8], where they imposed a strong condition on the model, namely, $E(\ln(\varepsilon)) = 0$. In our opinion, this is not necessary. The unknown constant c can be set to be $E(\ln(\varepsilon))$ such that $E(\ln(\tilde{\varepsilon})) = 0$ holds. Thus the function $g(\cdot)$ is identifiable. Meanwhile, suppose that one obtains a initial estimator of $g(\cdot)$, denoted by $\tilde{g}_n(\cdot)$. Combining the Equation (2), a more efficient estimator of $g(\cdot)$, denoted by $\hat{g}_n(\cdot)$, can be obtained by subtracting the mean of $E(\ln(\varepsilon))$ from the initial estimator $\tilde{g}_n(\cdot)$, where

$$\hat{g}_n(z) = \tilde{g}_n(\cdot) - \bar{c}, \quad \bar{c} = \frac{1}{n} \sum_{i=1}^n [\ln(Y_i) - \tilde{g}_n(Z_i)]. \quad (3)$$

In this process, it only needs that $E(\ln(\varepsilon))$, the expectation of $\ln(\varepsilon)$ is finite, which is allowed to be nonzero. Therefore, our condition is weaker than that stated in [8].

Following the LPRE technique used in [5], the estimator of $g(\cdot)$ in model (1) can be obtained by minimizing the objective function

$$S_n(g) = \frac{1}{n} \sum_{i=1}^n \left\{ \left| \frac{Y_i - \exp(g(Z_i))}{Y_i} \right| \times \left| \frac{Y_i - \exp(g(Z_i))}{\exp(g(Z_i))} \right| \right\}. \quad (4)$$

When all Z_i 's are mutually different, the computation involved in Equation (4)

can be accomplished individually for each $g(Z_i)$. However, the resulting estimator is not continuous, which is unsatisfactory in some applications. To overcome these drawbacks, we approximate $g(\cdot)$ by B-spline functions. Exactly speaking, let $-\infty < a = t_0 < t_1 < \dots < t_{k_n} < t_{k_n+1} = b < \infty$ be a sequence of knots, where $[a, b]$ denotes the support of Z . As discussed in [19], a set of B-spline basis function of degree d can be constructed and denoted by $\varphi_j(\cdot)$, $j = 1, \dots, K_n = k_n + 1 + d$, where $d = 2$ or 3 corresponds to the quadratic spline or the cubic spline. Then $g(z)$ can be approximated by a linear combination of basis functions, i.e., $g(z) \approx \sum_{j=1}^{K_n} \alpha_j \varphi_j(z) \triangleq \boldsymbol{\alpha}^T \mathbf{B}(z)$, where $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_{K_n})^T$ is the spline coefficient vector and $\mathbf{B}(z) = (\varphi_1(z), \dots, \varphi_{K_n}(z))^T$. Further, the model (1) can be rewritten as

$$Y = \exp(\boldsymbol{\alpha}^T \mathbf{B}(Z)) \varepsilon. \quad (5)$$

Meanwhile, $\boldsymbol{\alpha}$ can be estimated by $\tilde{\boldsymbol{\alpha}}$, the minimizer of the loss function

$$\begin{aligned} L_n(\boldsymbol{\alpha}) &= \frac{1}{n} \sum_{i=1}^n \left\{ \left| \frac{Y_i - \exp(\boldsymbol{\alpha}^T \mathbf{B}_i)}{Y_i} \right| \times \left| \frac{Y_i - \exp(\boldsymbol{\alpha}^T \mathbf{B}_i)}{\exp(\boldsymbol{\alpha}^T \mathbf{B}_i)} \right| \right\} \\ &= \frac{1}{n} \sum_{i=1}^n \left[Y_i \exp(-\boldsymbol{\alpha}^T \mathbf{B}_i) + Y_i^{-1} \exp(\boldsymbol{\alpha}^T \mathbf{B}_i) - 2 \right], \end{aligned}$$

where $\mathbf{B}_i = \mathbf{B}(Z_i)$. Then the B-spline estimator of $g(\cdot)$ can be obtained through Equation (3), where $\tilde{g}_n(z)$ is replaced by $\tilde{\boldsymbol{\alpha}}^T \mathbf{B}(z)$.

However, for reasons discussed in Section 1, here we estimate the nonparametric function in model (1) using penalized splines by adding a roughness penalty to the above minimization problem. In the literature, there are mainly two kinds of definition of roughness. One is based on the integrated squared q -th derivative of the spline function, the other is based on the q -th order difference of the spline coefficient vector, which is more popular for its simplicity and adopted in this paper. Exactly, let Δ denote the backward difference operator. Then we have $\Delta \alpha_j = \alpha_j - \alpha_{j-1}$ and $\Delta^2 \alpha_j = \alpha_j - 2\alpha_{j-1} + \alpha_{j-2}$. In General, let $\Delta^q = \Delta \Delta^{q-1}$. With these notations, the P-spline estimator of $g(\cdot)$ is defined as \hat{g} , the corrected spline function in (3), where $\tilde{g}_n(z)$ is replaced by $\hat{\boldsymbol{\alpha}}^T \mathbf{B}(z)$, and $\hat{\boldsymbol{\alpha}}$ is the minimizer of

$$Q_n(\boldsymbol{\alpha}) = L_n(\boldsymbol{\alpha}) + \lambda \sum_{j=q+1}^{K_n} \left| \Delta^q \alpha_j \right|^\gamma,$$

where $\gamma > 0$ and $\lambda > 0$ is the smoothing parameter controlling the smoothness of the fitting curve. If $\lambda = 0$, the resulting estimator becomes the B-spline estimator. In practice, common choices of q and γ are $q = 2$ and $\gamma = 2$ for computational convenience. At this time, the above minimization problem can be rewritten as

$$Q_n(\boldsymbol{\alpha}) = L_n(\boldsymbol{\alpha}) + \lambda \boldsymbol{\alpha}^T \mathbf{D}^T \mathbf{D} \boldsymbol{\alpha}, \quad (6)$$

where \mathbf{D} is the $(K_n - 2) \times K_n$ matrix representation of the second order difference operator Δ^2 .

2.2. Computation

Note that for given smoothing parameter and knots, the loss function (6) is differential and convex, minimizing it is not difficult, and some common algorithms can be applied. In the later simulation studies and real data analysis, we employ the function `nmk` in the R package `dfoptim`.

However, other alternatives are still helpful. As in the least square mean regression and penalized quantile regression (e.g. [20]), an iterative reweighted least square (IRLS) algorithm can be developed. Meanwhile, the measure for the effective model degree of freedom can be introduced based on the trace of the hat matrix, as used in [20]. In the literature, the degree of freedom usually is adopted as the measure of model complexity to facilitate further model comparison, which has not been discussed for multiplicative regression using local smoothing technique. Thus, the smoothing parameter selection can be achieved naturally using some existing information criteria or cross-validation.

When the smoothing parameter and knots are prepared, $\hat{\alpha}$ can also be seen as the root of the first derivative of objective function (6), namely,

$$-\frac{1}{n} \sum_{i=1}^n [Y_i \exp(-\hat{\alpha}^T \mathbf{B}_i) - Y_i^{-1} \exp(\hat{\alpha}^T \mathbf{B}_i)] \mathbf{B}_i + 2\lambda \mathbf{D}^T \mathbf{D} \hat{\alpha} = 0, \quad (7)$$

which can be rewritten as

$$-\frac{1}{n} \sum_{i=1}^n 2(\ln Y_i - \hat{\alpha}^T \mathbf{B}_i) W_i(\hat{\alpha}) \mathbf{B}_i + 2\lambda \mathbf{D}^T \mathbf{D} \hat{\alpha} = 0, \quad (8)$$

where

$$W_i(\hat{\alpha}) = \frac{Y_i \exp(-\hat{\alpha}^T \mathbf{B}_i) - Y_i^{-1} \exp(\hat{\alpha}^T \mathbf{B}_i)}{2(\ln Y_i - \hat{\alpha}^T \mathbf{B}_i)}. \quad (9)$$

Write $\tilde{\mathbf{Y}} = (\ln Y_1, \dots, \ln Y_n)^T$, $\mathbf{B} = (\mathbf{B}_1, \dots, \mathbf{B}_n)$ and $\mathbf{W}(\hat{\alpha}) = \text{diag}(W_1(\hat{\alpha}), \dots, W_n(\hat{\alpha}))$. The IRLS algorithm is implemented as follows.

Step 0. For given smoothing parameter λ , provide an initial value of α , denoted as $\hat{\alpha}^0$, which can be taken as the unpenalized B-spline estimator, or the ordinary least square estimator based on the logarithmic transformed data. Assume that we have the k -th iterative value $\hat{\alpha}^k$.

Step 1. At the $k+1$ -th step, update the parameter by

$$\hat{\alpha}^{k+1} = (\mathbf{B} \mathbf{W}(\hat{\alpha}^k) \mathbf{B}^T + 2\lambda \mathbf{D}^T \mathbf{D})^{-1} \mathbf{B} \mathbf{W}(\hat{\alpha}^k) \tilde{\mathbf{Y}}.$$

Repeat Step 1 until convergence. The value at the terminated step is defined as the final estimator of α , also denoted by $\hat{\alpha}$. At the same time, as in the linear regression, the hat matrix here can be defined through

$$\mathbf{H}_\lambda = \mathbf{B}^T (\mathbf{B} \mathbf{W}(\hat{\alpha}) \mathbf{B}^T + 2\lambda \mathbf{D}^T \mathbf{D})^{-1} \mathbf{B} \mathbf{W}(\hat{\alpha}).$$

2.3. Choice of Smoothing Parameter and Knots

To implement the estimator, the smoothing parameter λ and the number and

location of knots had to be determined in advance. Compared with the knots, the smoothing parameter plays a more significant role in the penalized spline estimation. As done in [20], we set the trace of the hat matrix mentioned above to be the model degrees of freedom, denoted by $df = \text{tr}(\mathbf{H}_\lambda)$, where $\text{tr}(A)$ denotes the trace of a square matrix A . In spline context, some criterions have been proposed to select the smoothing parameter λ , such as the Bayesian Information Criterion (BIC), Generalized Approximation Cross-Validation criterion (GACV) and Generalized Cross-Validation criterion (GCV). Especially, in this paper, they are implemented by minimizing the functions:

$$\begin{aligned} \text{BIC}(\lambda) &= \ln(L_{1n}(\hat{\boldsymbol{\alpha}})) + \frac{df \cdot \ln n}{2n}, \\ \text{GACV}(\lambda) &= \frac{L_{1n}(\hat{\boldsymbol{\alpha}})}{1 - df/n}, \\ \text{GCV}(\lambda) &= \frac{\sum_{i=1}^n W_i(\hat{\boldsymbol{\alpha}})(\ln(Y_i) - \hat{\boldsymbol{\alpha}}^T \mathbf{B}_i)/n}{(1 - df/n)^2}, \end{aligned}$$

respectively.

[18] suggested some rules for the number and location of knots. In the same spirits, we follow the strategy in [20], where quadratic B-splines with 20 equally-spaced quantiles knots are used unless otherwise specified. With the BIC, GACV, GCV criterions above, we use 51 equally-spaced log-scaled grid points on $\log_{10}(\lambda) \in [-5, 5]$ for the choice of λ . All these settings work well in the numerical studies and a real example analysis.

2.4. Asymptotic Results

For any probability measure P , define L_2 -norm $\|f\|_2 = \left(\int f^2 dP\right)^{1/2}$. Denote $p = d + 1$. To derive the asymptotic results, some regularity conditions are needed and listed as follows.

(A1) The true function in model (1) is g_0 . Suppose that $g_0 \in \mathcal{G} = \left\{g \in C^r[a, b]: \|g^{(j)}\|_\infty \leq M_0, j = 1, \dots, r, |g^{(r)}(z_1) - g^{(r)}(z_2)| \leq M_1 |z_1 - z_2|\right\}$, where M_0 and M_1 are some positive constants, $\|\cdot\|_\infty$ is the superior norm, $q \leq r \leq d + 1$.

(A2) Define $h_i = t_i - t_{i-1}$. Assume that $\max_i |h_{i+1} - h_i| = o(1/k_n)$. Moreover, the ratio of maximum and minimum spacings of knots is uniformly bounded.

(A3) $k_n = o(n)$. (A4) The covariate Z has abounded support $[a, b]$ with corresponding density $f_Z(\cdot)$, which has second derivative and is bounded away from zero and infinity.

(A5) ε is independent of Z , and $E(\varepsilon) = E(\varepsilon^{-1})$, $E(\ln(\varepsilon))$ is finite.

(A6) $E(\varepsilon + \varepsilon^{-1}) < +\infty$.

Conditions (A1)-(A3) are common requirements in the penalized spline theory. (A4) is the regularization condition used in the study of MR. (A5) is an identification condition for the LPRE estimation, which is similar to the zero mean condition in the classical linear mean regression. (A6) is required for proof,

which is also used in [15].

Theorem 2.1 Under conditions (A1)-(A6) mentioned above, we have that

1) If $r = p$, $k_n = O(n^{1/(2p+1)})$ and $\lambda = O(n^{-\beta})$ with $\beta > (1 + p - q)/(2p + 1)$, then it yields that $\|\hat{g} - g_0\|_2^2 = O_p(n^{-2p/(2p+1)})$.

2) If $r = q$, $\lambda = O(n^{-1/(2q+1)})$ and $k_n = O(n^\beta)$ with $1 > \beta > 1/(2q + 1)$, then it yields that $\|\hat{g} - g_0\|_2^2 = O_p(n^{-2q/(2q+1)})$.

Proof. Note that

$$\begin{aligned} Q_n(\alpha) &= \frac{1}{n} \sum_{i=1}^n [Y_i \exp(-\alpha^T B_i) + Y_i^{-1} \exp(\alpha^T B_i) - 2] + \lambda \alpha^T D^T D \alpha \\ &= \frac{1}{n} \sum_{i=1}^n [\varepsilon_i \exp(R_i + (g^*(Z_i) - \alpha^T B_i)) + \varepsilon_i^{-1} \exp(-R_i - (g^*(Z_i) - \alpha^T B_i)) - 2] \\ &\quad + \lambda \alpha^T D^T D \alpha, \end{aligned}$$

where $\varepsilon_i = Y_i / \exp(g_0(Z_i))$, $R_i = g(Z_i) - g^*(Z_i)$, and $g^*(Z) = \alpha^* B(Z)$ denotes the best spline function approximation. Define $X = n^{-1} \sum_{i=1}^n B_i B_i^T$,

$G_\lambda = X + \lambda D^T D$ and $\theta = G_\lambda^{1/2}(\alpha^* - \alpha)$. Then we have that

$g^*(Z_i) - \alpha^T B_i = B_i^T(\alpha^* - \alpha) = B_i^T G_\lambda^{-1/2} \theta$. Let a_n denote the convergence rate in the Theory. Since $\hat{\theta} = a_n(\hat{\alpha} - \alpha^*) = \arg \min_{\theta} [Q_n(\alpha^* + a_n^{-1} \theta) - Q_n(\alpha^*)]$, we only need to prove that for every $\nu > 0$, there exists a large constant $\delta > 0$ such that

$$P\left(\inf_{\|\theta\|_2 = \delta} Q_n(\alpha^* + a_n^{-1} \theta) - Q_n(\alpha^*) \geq 1 - \nu\right) \geq 1 - \nu.$$

In the line of the proof of Theorem 1 in the Appendix of [17], some similar techniques can be applied. Finally, by the Corollary 1 of [16], the desirable results are derived. The proof is completed.

3. Numerical Studies

In this section, numerical studies were conducted to evaluate the finite sample performance of our proposed method under various situations. To fairly compare the un-penalized B-spline estimator with the penalized spline estimators obtained by the BIC, GCV, GACV criteria and for simplicity, we set the degree of spline basis to be $q = 2$, the number of internal knots $k_n = 20$, which are located on equally-spaced quantiles for all methods. All results below are based on 500 replicates, where the sample size $n = 100, 300, 500$, respectively. All simulations are implemented using the software R.

3.1. Model A and Results

Model A. We generated (Y, Z) from the following model

$$Y = \exp\left(1 + 2 \sin\left(\frac{\pi Z}{30}\right)\right) \varepsilon, \tag{10}$$

where $Z \sim Unif(0, 50)$ and is independent with the random error ε . Further, three error distributions are considered, namely,

Case (i): $\varepsilon = \exp(U)$, $U \sim N(0, 1)$,

Case (ii): $\varepsilon = \exp(U)$, $U \sim Unif(-2, 2)$,

Case (iii): $\varepsilon \sim Unif(0.0001, 4.635506)$.

Note that under case (i)-(ii), it holds that $E(\varepsilon - \varepsilon^{-1}) = 0$ and $E(\log(\varepsilon)) = 0$. But under case (iii), $E(\varepsilon - \varepsilon^{-1}) = 0$, $E(\log(\varepsilon)) = 0.5339771$. We use the averaged Integrated Absolute Bias (IABIAS) and Mean Integrated Square Error (MISE), where for one estimator $\hat{g}_j (j = 1, \dots, 500)$ obtained from the j -th sample,

$$IABIAS = \frac{1}{500} \sum_{j=1}^{500} \left[\frac{1}{ngrid} \sum_{k=1}^{ngrid} |\hat{g}_j(u_k) - g_0(u_k)| \right],$$

$$MISE = \frac{1}{500} \sum_{j=1}^{500} \left[\frac{1}{ngrid} \sum_{k=1}^{ngrid} [\hat{g}_j(u_k) - g_0(u_k)]^2 \right],$$

at the fixed grid points $\{u_k\}$ equally-spaced in $[0, 50]$ and $ngrid = 501$. The values below them are the associated sample standard deviation. To study the different smoothing parameter selection methods at a finer scale, we also calculate the mean and standard deviation of effective model degree of freedom (DF).

Table 1 reported the results. It can be seen that all three penalized estimators

Table 1. Results of Model A under cases (i)-(iii) with different sample sizes and criterions ($\times 10^{-2}$).

	Case (i)				Case (ii)				Case (iii)			
	BIC	GACV	GCV	BS	BIC	GACV	GCV	BS	BIC	GACV	GCV	BS
<i>n</i> = 100												
IABIAS	29.835	33.621	30.643	43.104	29.550	34.495	30.006	46.696	49.812	54.306	51.422	64.237
	6.069	6.499	6.140	8.036	7.027	7.847	7.026	10.649	9.797	12.765	9.684	16.827
MISE	14.581	19.176	15.379	37.450	14.436	20.481	14.873	54.989	37.339	58.578	39.666	106.553
	5.875	14.661	6.106	54.348	6.459	17.833	6.547	135.870	20.192	329.668	22.028	587.813
DF	1227.584	1409.248	1267.132	-	1181.893	1376.901	1204.128	-	1260.861	1429.779	1318.995	-
	106.715	106.075	102.198	-	106.072	109.127	100.670	-	104.080	111.360	108.925	-
<i>n</i> = 300												
IABIAS	17.885	19.834	18.962	23.643	15.972	17.978	16.595	22.295	33.432	36.265	36.146	40.697
	3.370	3.529	3.476	3.645	3.363	3.624	3.413	3.838	6.675	6.458	6.577	6.291
MISE	5.188	6.407	5.852	9.179	4.225	5.348	4.552	8.318	18.751	21.386	21.270	26.320
	1.845	2.161	2.036	2.753	1.671	2.089	1.783	2.961	9.249	9.319	9.419	9.831
DF	1196.849	1407.369	1314.363	-	1142.739	1354.685	1222.428	-	1289.492	1485.475	1471.698	-
	68.528	76.614	78.794	-	68.308	77.785	69.231	-	77.216	75.052	88.981	-
<i>n</i> = 500												
IABIAS	14.356	15.654	15.113	18.230	12.5844	13.735	12.969	16.557	29.411	31.519	31.803	34.598
	2.626	2.768	2.750	2.910	2.4399	2.530	2.388	2.715	7.374	6.952	7.045	6.740
MISE	3.372	4.003	3.740	5.439	2.6288	3.130	2.788	4.567	16.395	18.093	18.354	20.913
	1.184	1.354	1.303	1.651	0.9924	1.113	0.993	1.470	11.118	10.905	11.013	10.937
DF	1210.032	1426.518	1339.452	-	1157.6377	1365.706	1251.546	-	1317.323	1522.904	1553.775	-
	61.335	68.784	73.771	-	52.0414	68.176	55.066	-	75.347	74.644	103.840	-

have smaller IABIAS and MISE than those of the unpenalized B-spline estimator. This indicates the benefits of the proposed method. As the sample size increases, the differences between them decrease and become comparable. The model degree using GACV are almost larger than those of using BIC and GCV in all settings, and those using BIC are the smallest, which is also observed in [20]. These imply that GACV tends to select more complex models, BIC tends to select simpler models, GCV makes a compromise and lies between them. **Figure 1** displays the average estimated curves and their 95% point-wise confidence band under cases (i) with sample size $n = 300$ for model A. We can see that all estimated curves are close to the true one, but the B-spline estimation has a drastic deviation near the boundaries. **Figure 2** shows the boxplots of the estimates of $g(z)$ at $z = 15, 30, 45$, respectively. We can see that most estimates are centered around the true values of $(3, 1, -1)$ with small deviations, although the BS estimator has a little larger standard deviation. The corresponding QQ-plots of them are presented in **Figure 3**, which show that the resulting estimators seem to converge asymptotically to the normal distribution. Estimated curves under other situations are similar, and are not displayed, but are available in the supplementary materials.

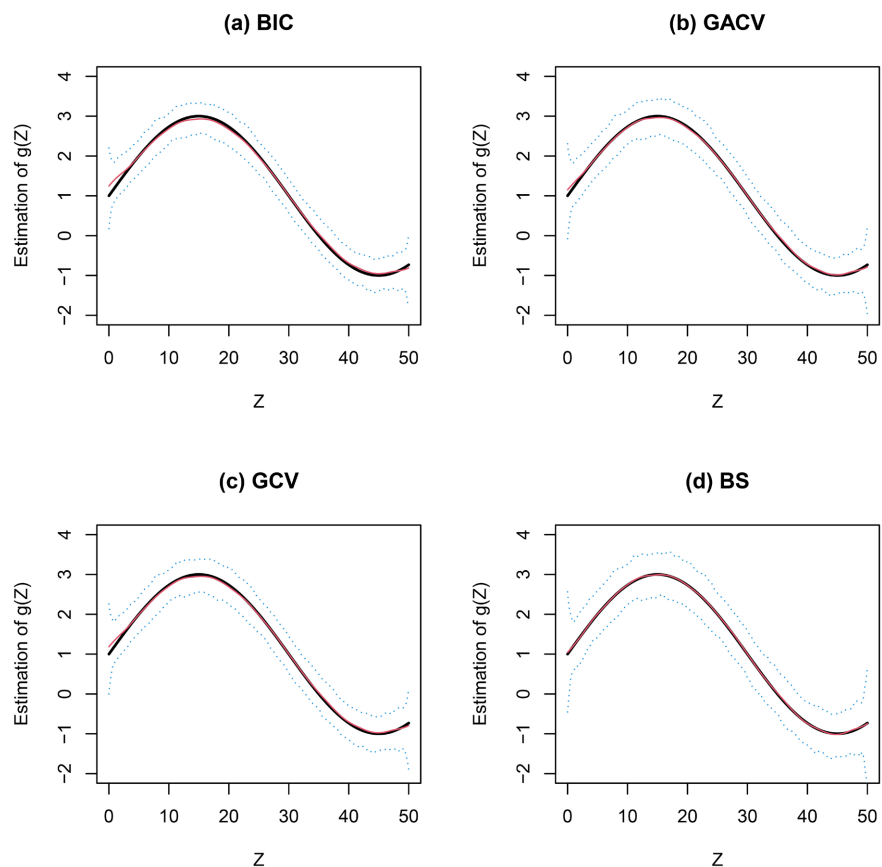


Figure 1. Estimate of $g(Z)$ for Model A under case (i) with sample size 300. The solid thick line (black) and solid thin line (red) correspond to the true and estimated curves, respectively. The dashed lines (blue) are the 95% point-wise confidence bands.

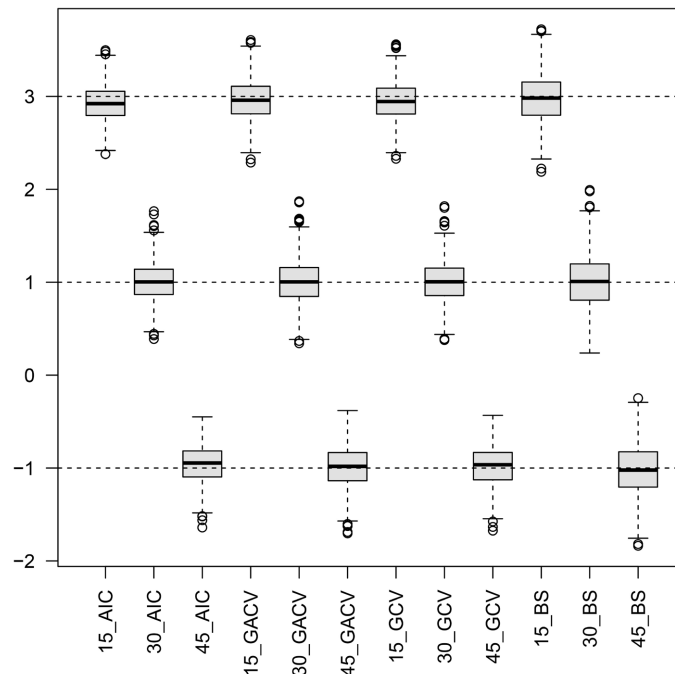


Figure 2. Box-plots of $g(Z)$ at $Z=15,30,45$, respectively, for Model A under case (i) with sample size 300.

3.2. Model B and Results

We generated (Y, Z) from the following model

$$Y = \exp\left(\frac{|Z - 25|^3}{5000}\right)\varepsilon, \tag{11}$$

where $Z \sim Unif(0,50)$ and is independent with the random error ε . Others are the same as defined in Model A. We reported the results in **Table 2** and displayed the average estimated curves and their 95% point-wise confidence band under cases (i) with sample size $n = 300$ for model B in **Figure 4**. Again, results are similar to those presented in **Table 1** and **Figures 1-3** are found. More related figures can be available in the supplementary materials.

4. Real Data Analysis

In this section, we will analyze the ethanol data to illustrate our proposed method. This data is available in the R package *lattice*, and aims to investigate the relationship between the emissions of nitrogen oxides, denoted by Y , and various settings of the engine compression ratio and equivalence ratio, denoted by Z , when ethanol fuel is burned in a single-cylinder engine. In this data set, 88 observations are collected. [13] analyzed it in the framework of varying coefficient multiplicative regression model based on the LPRE criterion and local linear approximation technique. His results demonstrate that various settings of the engine compression ratio have little effect in Y , as illustrated in page 280 therein. Therefore, here we only consider the covariate Z in the following model.

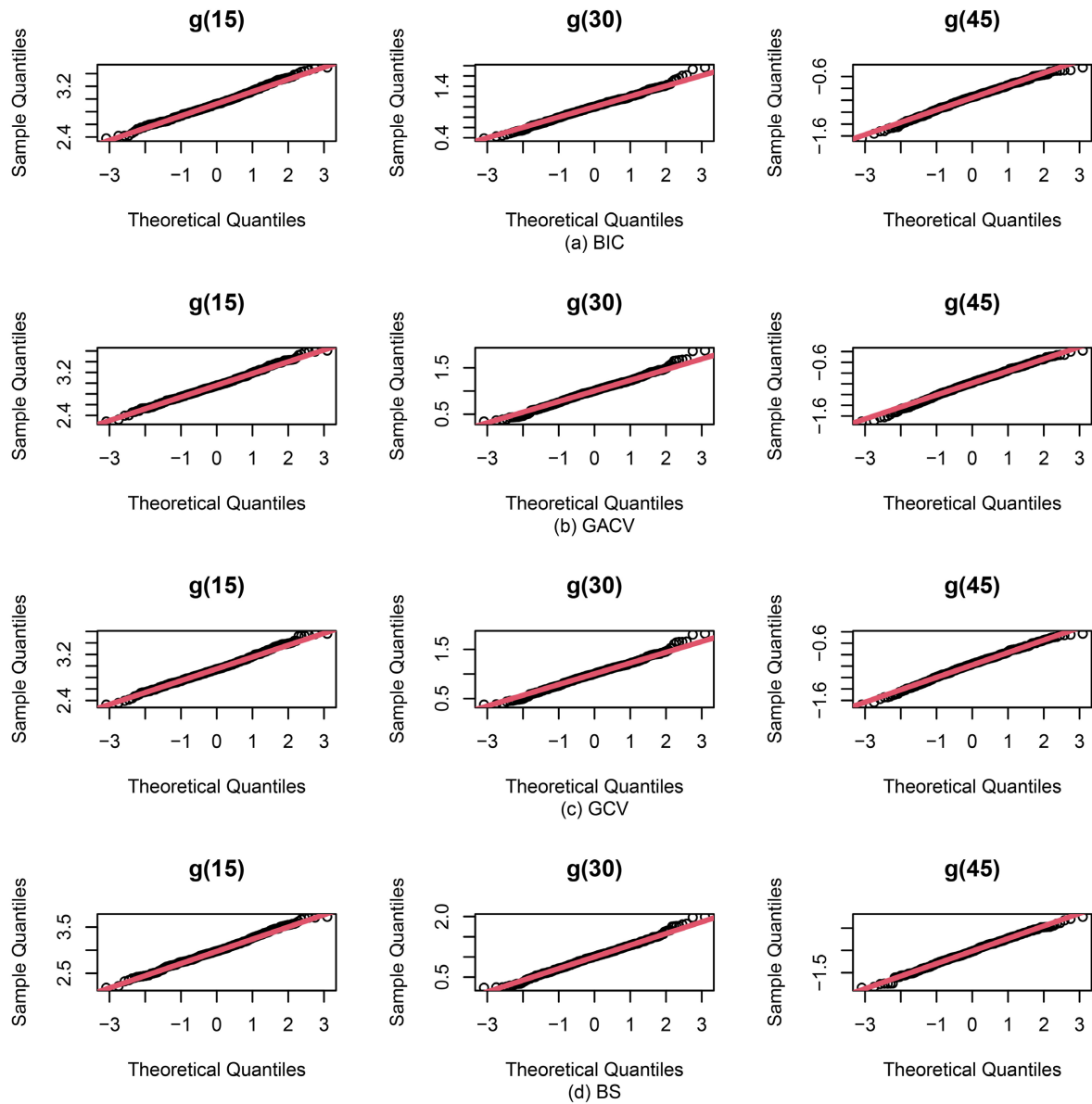


Figure 3. QQ-plots of $g(Z)$ at $Z = 15, 30, 45$, respectively, for Model A under case (i) with sample size 300.

Table 2. Results of Model B under cases (i)-(iii) with different sample sizes and criterions ($\times 10^{-2}$).

	Case (i)				Case (ii)				Case (iii)			
	BIC	GACV	GCV	BS	BIC	GACV	GCV	BS	BIC	GACV	GCV	BS
<i>n</i> = 100												
IABIAS	30.954	34.347	31.649	43.095	31.263	35.429	31.520	46.884	50.598	54.612	52.022	63.973
	6.074	6.498	6.170	8.037	6.626	7.611	6.604	11.695	9.431	12.712	9.376	14.377
MISE	16.093	20.245	16.762	37.521	16.510	21.848	16.727	60.375	38.319	59.115	40.487	89.844
	5.979	14.659	6.219	56.611	6.612	18.404	6.645	185.554	21.209	334.487	23.024	328.872
DF	1253.243	1429.974	1292.463	-	1210.361	1395.702	1228.656	-	1281.472	1445.200	1338.260	-
	99.732	100.901	95.411	-	108.539	110.258	98.756	-	106.142	108.857	104.281	-

Continued

<i>n</i> = 300												
IABIAS	18.762	20.381	19.606	23.643	16.979	18.632	17.519	22.294	34.039	36.581	36.376	40.696
	3.338	3.458	3.426	3.644	3.176	3.521	3.287	3.838	6.588	6.406	6.537	6.289
MISE	5.888	6.823	6.358	9.179	4.972	5.794	5.183	8.317	19.121	21.600	21.393	26.320
	1.962	2.180	2.087	2.752	1.830	2.143	1.893	2.961	9.197	9.268	9.422	9.832
DF	1258.010	1457.920	1363.208	-	1205.782	1408.397	1284.380	-	1325.434	1518.928	1496.542	-
	65.165	70.682	71.458	-	67.726	71.847	63.174	-	70.470	71.405	87.412	-
<i>n</i> = 500												
IABIAS	15.087	16.093	15.705	18.229	13.501	14.360	13.814	16.558	29.910	31.766	31.988	34.599
	2.561	2.696	2.665	2.910	2.225	2.417	2.284	2.715	7.169	6.858	6.965	6.741
MISE	3.816	4.257	4.075	5.438	3.160	3.452	3.231	4.568	16.721	18.253	18.463	20.914
	1.230	1.357	1.301	1.651	1.041	1.137	1.052	1.470	11.018	10.847	10.949	10.939
DF	1281.520	1484.408	1406.453	-	1233.752	1437.042	1329.486	-	1360.986	1560.749	1586.328	-
	56.717	62.400	68.205	-	54.702	65.378	57.779	-	70.218	67.571	92.849	-

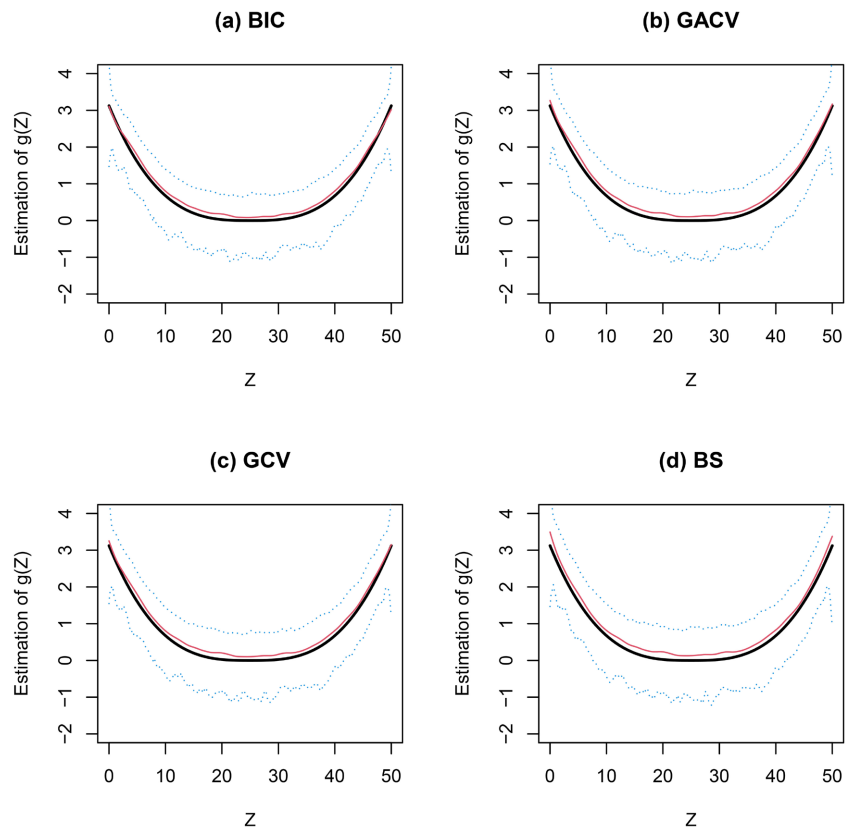


Figure 4. Estimate of $g(Z)$ for Model B under case (i) with sample size 300. The solid thick line (black) and solid thin line (red) correspond to the true and estimated curves, respectively. The dashed lines (blue) are the 95% point-wise confidence bands.

$$Y = \exp(g(Z))\varepsilon.$$

The estimated curves of g with 8 and 20 internal knots quadratic P-splines are

plotted in **Figure 5**, which have similar trends to the plot in [13] except a vertical shift. We can see that all the four curves are almost identical, which maybe caused by the data's own well fitness to the above model. Scatted plot of Y and the corresponding estimated curves of $\exp(g(Z))$ are presented in **Figure 6**, which indicate the proposed method fits the data well enough. The number of knots has some effects on the estimates of curves, but doesn't affect the performance substantially, which is also supported by the mean and median of absolute prediction errors $|Y_i - \hat{Y}_i|$ and squared prediction errors $(Y_i - \hat{Y}_i)^2$ given in the **Table 3**, where $\hat{Y}_i = \exp(\hat{g}(Z_i))$.

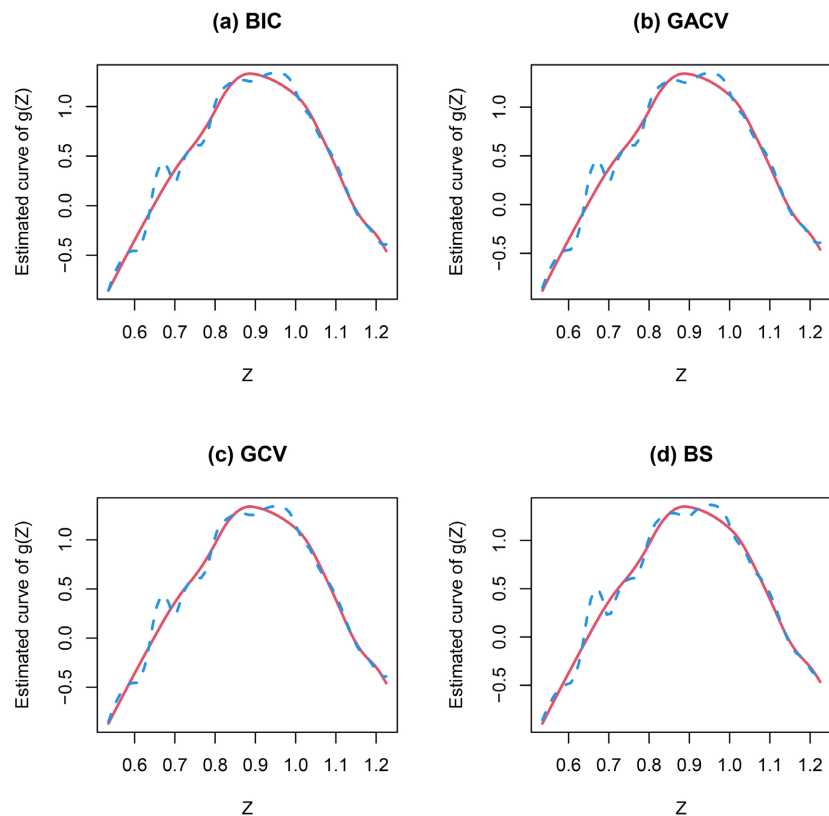


Figure 5. Estimated curves of $g(Z)$ for the ethanol data. The solid (red) and dashed (blue) curves correspond to the estimated function curves with 8 and 20 internal knots, respectively.

Table 3. Prediction errors and degrees of freedom of the ethanol data.

	$k_n = 8$					$k_n = 20$				
	MAPE	MDAPE	MSPE	MDSPE	DF	MAPE	MDAPE	MSPE	MDSPE	DF
BIC	0.240	0.188	0.105	0.035	7.116	0.187	0.128	0.069	0.016	12.488
GACV	0.242	0.188	0.105	0.035	8.208	0.186	0.126	0.068	0.016	13.613
GCV	0.241	0.188	0.105	0.035	7.658	0.187	0.128	0.069	0.016	12.488
BS	0.243	0.197	0.105	0.038	-	0.183	0.1282	0.067	0.016	-

Note: k_n : the number of interior knots; MAPE: mean absolute prediction errors; MDAPE: median absolute prediction errors; MSPE: mean squared prediction errors; MASPE: median squared prediction errors.

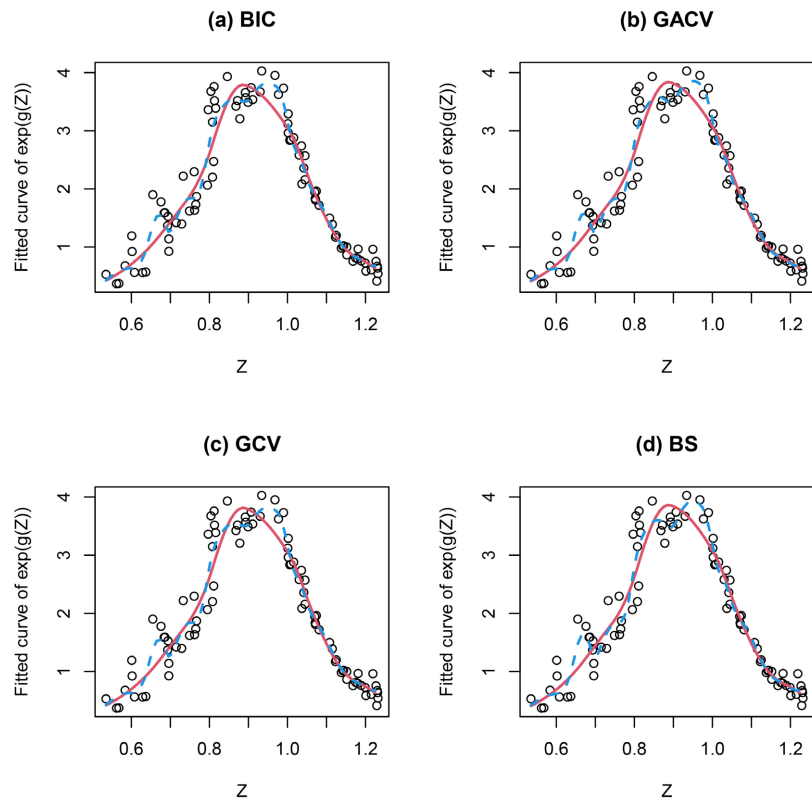


Figure 6. Scatted plot of Y and estimated curves of $\exp(g(Z))$ for the ethanol data. The solid (red) and dashed (blue) curves correspond to the estimated function curves with 8 and 20 internal knots, respectively.

5. Conclusion and Discussions

In this study, we used the penalized spline method to estimate the nonparametric function in model (1) based on the LPRE loss function. Inspired by the iteratively reweighted least squares algorithm, we develop the effective model degree of freedom and propose three smoothing parameter selection methods. Some asymptotic results are established. Furthermore, numerical simulation studies and real data analysis found that the proposed approach works well in several settings. As indicated in Section 1, the approaches proposed in this paper may be adapted to the partially linear, even additive, single index or varying coefficients multiplicative regression models. Our future work will also consider extensions to fields with covariates measurement errors, censored data or longitudinal data analysis, which are meaningful for practitioners.

Data Availability Statement

The data set used in the real data analysis can be available publicly from the R package *lattice*.

Acknowledgements

This work was partly supported by a grant from Natural Science Foundation of

Jiangsu Province under project ID BK20210889 and Start-up fund for doctoral research of Jiangsu University of Science and technology. The authors also thank the lecturer Feng-Ling Ren, School of Computer and Engineering, Xinjiang University of Finance & Economics, for helpful discussions.

Conflicts of Interest

The authors declare no conflicts of interest.

References

- [1] Chen, K., Guo, S., Lin, Y. and Ying, Z. (2010) Least Absolute Relative Error Estimation. *Journal of the American Statistical Association*, **105**, 1104-1112. <https://doi.org/10.1198/jasa.2010.tm09307>
- [2] Khoshgoftaar, T.M., Bhattacharyya, B.B. and Richardson, G.D. (1992) Predicting Software Errors, during Development, Using Nonlinear Regression Models: A Comparative Study. *IEEE Transactions on Reliability*, **41**, 390-395. <https://doi.org/10.1109/24.159804>
- [3] Narula, S.C. and Wellington, J.F. (1977) Prediction, Linear Regression and the Minimum Sum of Relative Errors. *Technometrics*, **19**, 185-190. <https://doi.org/10.1080/00401706.1977.10489526>
- [4] Park, H. and Stefanski, L.A. (1998) Relative-Error Prediction. *Statistics and Probability Letters*, **40**, 227-236. [https://doi.org/10.1016/S0167-7152\(98\)00088-1](https://doi.org/10.1016/S0167-7152(98)00088-1)
- [5] Chen, K., Lin, Y., Wang, Z. and Ying, Z. (2016) Least Product Relative Error Estimation. *Journal of Multivariate Analysis*, **144**, 91-98. <https://doi.org/10.1016/j.jmva.2015.10.017>
- [6] Zhang, Q. and Wang, Q. (2013) Local Least Absolute Relative Error Estimating Approach for Partially Linear Multiplicative Model. *Statistica Sinica*, **23**, 1091-1116. <https://doi.org/10.5705/ss.2012.133>
- [7] Hirose, K. and Masuda, H. (2018) Robust Relative Error Estimation. *Entropy*, **20**, Article 632. <https://doi.org/10.3390/e20090632>
- [8] Zhang, J., Feng, Z. and Peng, H. (2018) Estimation and Hypothesis Test for Partial Linear Multiplicative Models. *Computational Statistics & Data Analysis*, **128**, 87-103. <https://doi.org/10.1016/j.csda.2018.06.017>
- [9] Chen, Y. and Liu, H. (2021) A New Relative Error Estimation for Partially Linear Multiplicative Model. *Communications in Statistics-Simulation and Computation*, **52**, 4962-4980.
- [10] Liu, H. and Xia, X. (2018) Estimation and Empirical Likelihood for Single-Index Multiplicative Models. *Journal of Statistical Planning and Inference*, **193**, 70-88. <https://doi.org/10.1016/j.jspi.2017.08.003>
- [11] Zhang, J., Zhu, J. and Feng, Z. (2019) Estimation and Hypothesis Test for Single-Index Multiplicative Models. *Test*, **28**, 242-268. <https://doi.org/10.1007/s11749-018-0586-2>
- [12] Zhang, J., Cui, X. and Peng, H. (2020) Estimation and Hypothesis Test for Partial Linear Single-Index Multiplicative Models. *Annals of the Institute of Statistical Mathematics*, **72**, 699-740. <https://doi.org/10.1007/s10463-019-00706-6>
- [13] Hu, D.H. (2019) Local Least Product Relative Error Estimation for Varying Coefficient Multiplicative Regression Model. *Acta Mathematicae Applicatae Sinica, English Series*, **35**, 274-286. <https://doi.org/10.1007/s10255-018-0794-2>

-
- [14] Chen, Y., Liu, H. and Ma, J. (2022) Local Least Product Relative Error Estimation for Single-Index Varying-Coefficient Multiplicative Model with Positive Responses. *Journal of Computational and Applied Mathematics*, **415**, Article ID: 114478. <https://doi.org/10.1016/j.cam.2022.114478>
- [15] Ming, H., Liu, H. and Yang, H. (2022) Least Product Relative Error Estimation for Identification in Multiplicative Additive Models. *Journal of Computational and Applied Mathematics*, **404**, Article ID: 113886. <https://doi.org/10.1016/j.cam.2021.113886>
- [16] Kalogridis, I. and Van Aelst, S. (2021) M-Type Penalized Splines with Auxiliary Scale Estimation. *Journal of Statistical Planning and Inference*, **212**, 97-113. <https://doi.org/10.1016/j.jspi.2020.09.004>
- [17] Kalogridis, I. and Van Aelst, S. (2021) Robust Penalized Spline Estimation with Difference Penalties. *Econometrics and Statistics*, **29**, 169-188.
- [18] Ruppert, D., Wand, M.P. and Carroll, R.J. (2003) *Semiparametric Regression*. Cambridge University Press, New York. <https://doi.org/10.1017/CBO9780511755453>
- [19] Schumaker, L. (2007) *Spline Functions: Basic Theory*. Cambridge University Press, New York. <https://doi.org/10.1017/CBO9780511618994>
- [20] Wu, C. and Yu, Y. (2014) Partially Linear Modeling of Conditional Quantiles Using Penalized Splines. *Computational Statistics & Data Analysis*, **77**, 170-187. <https://doi.org/10.1016/j.csda.2014.02.020>