



B-spline Speckman Estimator of Partially Linear Model

Sayed Meshaal El-sayed, Mohamed Reda Abonazel*, Mohamed Metwally Seliem

Department of Applied Statistics and Econometrics, Faculty of Graduate Studies for Statistical Research, Cairo University, Giza, Egypt

Email address:

mabonazel@cu.edu.eg (M. R. Abonazel), mabonazel@hotmail.com (M. R. Abonazel)

*Corresponding author

To cite this article:

Sayed Meshaal El-sayed, Mohamed Reda Abonazel, Mohamed Metwally Seliem. B-spline Speckman Estimator of Partially Linear Model. *International Journal of Systems Science and Applied Mathematics*. Vol. 4, No. 4, 2019, pp. 53-59. doi: 10.11648/j.ijssam.20190404.12

Received: August 25, 2019; **Accepted:** December 31, 2019; **Published:** February 3, 2020

Abstract: The partially linear model (PLM) is one of semiparametric regression models; since it has both parametric (more than one) and nonparametric (only one) components in the same model, so this model is more flexible than the linear regression models containing only parametric components. In the literature, there are several estimators are proposed for this model; where the main difference between these estimators is the estimation method used to estimate the nonparametric component, since the parametric component is estimated by least squares method mostly. The Speckman estimator is one of the commonly used for estimating the parameters of the PLM, this estimator based on kernel smoothing approach to estimate nonparametric component in the model. According to the papers in nonparametric regression, in general, the spline smoothing approach is more efficient than kernel smoothing approach. Therefore, we suggested, in this paper, using the basis spline (B-spline) smoothing approach to estimate nonparametric component in the model instead of the kernel smoothing approach. To study the performance of the new estimator and compare it with other estimators, we conducted a Monte Carlo simulation study. The results of our simulation study confirmed that the proposed estimator was the best, because it has the lowest mean squared error.

Keywords: Kernel Smoothing, Monte Carlo Simulation, Penalized B-spline Estimation, Semiparametric Regression, Spline Smoothing

1. Introduction

Linear regression modelling is a good form for linking variables because in general the parameters have some kind of meaning or interpretation. Nevertheless, it is known that the main drawback of the linear regression models is lacking flexibility. In practice, this fact causes that some interesting relationships cannot be modelled by means of this class of models. A way to avoid that drawback is to add to the linear regression model a nonparametric component. The resulting model known as partially linear model (PLM) was introduced by Engle *et al.* [1] to study the effect of weather on electricity demand. The PLM is defined by:

$$y_i = x_i\beta + f(t_i) + \varepsilon_i; \quad i = 1, 2, \dots, n, \quad (1)$$

where y_i denotes the dependent variable, $x_i = (x_{i1}, \dots, x_{iq})$ and t_i are the independent variables, $f(\cdot)$ is the nonparametric part in the model, $\beta = (\beta_1, \dots, \beta_q)'$ is the vector of regression parameters of the parametric part in the

model, and the random errors $(\varepsilon_1, \dots, \varepsilon_n)$ are independent and identically distributed, and $E(\varepsilon_i|x_i, t_i) = 0$; $E(u_i^2|x_i, t_i) = \sigma^2$. Note that the intercept term has been omitted from the parametric component without loss of any generality, since the first point on a nonparametric regression line plays the role of an intercept. This model has gained great popularity since it was first introduced by Engle *et al.* [1] and has been widely applied in economics, social, biological sciences, and so on.

The PLM can be reduced to the classical linear regression model if the nonparametric component is equal zero. The goal is to estimate the unknown parameter vector β and nonparametric function f from the data $\{y_i, x_i, t_i\}$.

For estimating the nonparametric component in the PLM, the spline smoothing approach is used in many studies, such as [1-6]. And the kernel approach is used by, for example, [7-9]. Also, there are other estimators of this model are discussed, such as [10-16].

Recently, Abonazel *et al.* [6] modified the Speckman estimator by using the spline smoothing approach, and they

showed that their estimator is more efficient than traditional Speckman estimator that based on the kernel smoothing approach.

In this paper, we will suggest using the basis spline (B-spline) to estimate a nonparametric component in the model. And we will compare the new estimator with the traditional Speckman's [8] and Abonazel's [6] estimators.

The rest of the paper is organized as follows. In the next two sections, we introduce the traditional Speckman's [8] and Abonazel's [6] estimators respectively. Our proposed estimator is presented in section 4. While in section 5, the Monte Carlo simulation study is conducted to compare the performance of the three estimators. The concluding remarks are included in section 6.

2. Speckman Estimator

The PLM in equation (1) can be rewritten as

$$Y = X\beta + f(t) + \varepsilon, \quad (2)$$

where $Y = (y_1, \dots, y_n)'$, $t = (t_1, \dots, t_n)'$, $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)'$, and $X = [x_1 \dots x_q]$; $x_r = (x_{1r}, \dots, x_{nr})'$ for $r = 1, \dots, q$. Taking the conditional expectation of both sides of equation (2) given t :

$$\Phi_Y(t) = \Phi_X(t)\beta + f(t), \quad (3)$$

where $\Phi_Y(t) = E(Y|t)$ and $\Phi_X(t) = E(X|t)$. Subtracting equation (3) from equation (2) yields

$$\tilde{Y} = \tilde{X}\beta + \varepsilon, \quad (4)$$

where $\tilde{Y} = Y - \Phi_Y(t)$ and $\tilde{X} = X - \Phi_X(t)$. In practice, the conditional means are unknown and should be estimated. Speckman estimator uses the kernel smoothing approach to estimate the nonparametric component in PLM. Therefore, to get the classical version of Speckman estimator should be used the Nadaraya-Watson smoother matrix:

$$S = \{s_{ij}\}_{n \times n}; s_{ij} = \mathcal{K}\left(\frac{t_i - t_j}{h}\right) / \sum_{m=1}^n \mathcal{K}\left(\frac{t_i - t_m}{h}\right), \quad (5)$$

where $\mathcal{K}(\cdot)$ denote a kernel function and h is a bandwidth value. Based on the smoother matrix in equation (5), the equation (4) can be rewritten:

$$\tilde{Y}_\kappa = \tilde{X}_\kappa \beta + \varepsilon,$$

where $\tilde{Y}_\kappa = Y - \hat{\phi}_Y^\kappa(t)$; with $\hat{\phi}_Y^\kappa(y) = SY$, and $\tilde{X}_\kappa = X - \hat{\phi}_X^\kappa(t)$; with $\hat{\phi}_X^\kappa(t) = SX$.

Now, the vector of parametric component can be estimated by

$$\hat{\beta}_\kappa = (\tilde{X}'_\kappa \tilde{X}_\kappa)^{-1} \tilde{X}'_\kappa \tilde{Y}_\kappa, \quad (6)$$

while the nonparametric component can be estimated by $\hat{f}_\kappa = S(\tilde{Y}_\kappa - \tilde{X}_\kappa \hat{\beta}_\kappa)$.

3. Spline Smoothing Estimator

Abonazel *et al.* [6] modified the traditional Speckman

estimator by using the spline smoothing approach to estimate the nonparametric component in PLM. According to their approach, the fitted values of f are given by minimizing the penalized sum of squares (PSS):

$$PSS = \sum_{i=1}^n [(y_i - x_i \hat{\beta}) - f(t_i)]^2 + \lambda \int_a^b [f''(t)]^2 dt, \quad (7)$$

where $\hat{\beta}$ is the estimated parametric component that based on the smoother spline matrix (S_λ), depending on smoothing parameter (λ) and the knots points. And the nonparametric component can be estimated by

$$\hat{f}_\lambda = \left(\hat{f}_\lambda(\xi_1), \dots, \hat{f}_\lambda(\xi_K) \right)' = S_\lambda(Y - X\hat{\beta}),$$

where \hat{f}_λ is a natural cubic spline with knots at ξ_1, \dots, ξ_K for a fixed $\lambda > 0$, and S_λ is a well-known positive-definite smoother matrix which depends on λ and the knot points. To solve equation (7), an iterative algorithm is required.

It note that the second term ($\int_a^b [f''(t)]^2 dt$) is the penalty (or regularization) term, where $f''(\cdot)$ refers to second derivatives, and a and b are the minimum and maximum values of t respectively. The first term in equation (7) measures the closeness to the data. While the second term in same equation penalized the curvature in the function, but it is difficult to solve mathematically, so Green and Silverman [17] solved it with the assumption that the function f is a natural cubic spline. This might seem an over-parameterized model, however, the penalty term ensures that the coefficients are shrunk towards linearity, limiting the number of degrees of freedom used.

Let (a, b) be an interval and let (ξ_1, \dots, ξ_K) be K points such that $a < \xi_1 < \dots < \xi_K < b$. A continuous function f on (a, b) is a cubic spline with knots $\{\xi_1, \dots, \xi_K\}$ if

- a) f is a cubic polynomial over the intervals $\{(\xi_1, \xi_2), \dots\}$.
- b) f has continuous first and second derivatives at the knots.

Based on the above the natural cubic spline can be defined as a polynomial spline $f: [a, b] \rightarrow \mathbb{R}$ of degree three if $f''(a) = f''(b) = 0$. Natural cubic spline are cubic spline with the constraint that they are linear in their tails beyond the boundary knots (a, ξ_1) and (ξ_K, b) .

In general, the placement of the knots and the determination of the penalty are very important for a spline smoothing where the number of knots is equal the number of observations. While in penalized spline be the number of knots is less than the number of observations, see [18, 19] for more details.

4. Proposed Estimator

Now we suggest using the B-spline approach instead of the kernel approach to fit the nonparametric component in this model. A B-spline is a spline function that has minimal support with respect to a given degree, smoothness, and domain partition. De Boor [20] shows that every spline function of a given degree, smoothness, and domain partition can be uniquely represented as a linear combination of B-splines of that same degree and smoothness, and over that

same partition.

Beginning, let $\xi_0 = a$ and $\xi_{K+1} = b$ where the knots from ξ_1 to ξ_K is called inner knots while a and b are called boundary knots. Define new knots $\kappa_1 < \dots < \kappa_M$ such that $\kappa_1 \leq \kappa_2 \leq \dots \leq \kappa_M \leq \xi_0, \kappa_{s+M} = \xi_s$ for $s = 1, \dots, K$, and $\xi_{K+1} \leq \kappa_{K+M+1} \leq \dots \leq \kappa_{K+2M}$. The choice of extra knots is arbitrary; usually one takes $\kappa_1 = \kappa_2 = \dots = \kappa_M = \xi_0$ and $\xi_{K+1} = \kappa_{K+M+1} = \dots = \kappa_{K+2M}$ (see Wasserman [19]). Given a set of K knots, the B-spline basis function recursively can be defined by

$$B_s^0(t, \kappa) = \begin{cases} 1 & \kappa_s < t \leq \kappa_{s+1}, \\ 0 & \text{otherwise.} \end{cases} \quad (8)$$

De Boor [20] introduced an algorithm to compute B-spline of any degree from B-spline of lower degree. Because a B-spline basis function in equation (8) is just a constant on one interval between two knots, it is simple to compute B-spline of any degree. His algorithm also works for any placement of knots (i.e. equidistant knots or not equidistant knots). The general B-spline of degree $p \leq M$ with K knots is

$$B_s^p(t, \kappa) = \frac{t - \kappa_s}{\kappa_{s+p} - \kappa_s} B_s^{p-1}(t) + \frac{\kappa_{s+p+1} - t}{\kappa_{s+p+1} - \kappa_{s+1}} B_{s+1}^{p-1}(t), \quad (9)$$

where $s = 1, \dots, K + p + 1$ and $B_s^p(t, \kappa)$ denotes s^{th} B-spline of degree p . It is understood that if the denominator in equation (9) is equal zero, then the function is defined to be zero. Note that additional $2p + 2$ knots are necessary for constructing the full B-spline basis of degree p .

Eilers and Marx [21] introduced the combination of B-splines and difference penalties which they called penalized B-spline (P-spline). The P-spline estimation is given by minimizes the following function:

$$PPSS = \|Y - B\tau\|^2 + \lambda \|D_d \tau\|^2,$$

where B is a $n \times (K + p + 1)$ matrix contains B-splines that given by (9), $\tau = (\tau_1, \dots, \tau_{K+p+1})'$ is a parameter vector of the spline function, and D_d has dimension $(n - d) \times n$. In practice, $d = 2$ or 3 is commonly used, see [22-24]. For example, if $d = 1$ or $d = 2$, the D_1 and D_2 matrices can be designed as follows:

$$D_1 = \begin{pmatrix} -1 & 1 & \dots & 0 \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & -1 & 1 \end{pmatrix}; D_2 = \begin{pmatrix} 1 & -2 & 1 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 1 & -2 & 1 \end{pmatrix}.$$

Minimizing PPSS leads to

$$\hat{\tau} = [B' B + \lambda (D_d' D_d)]^{-1} B' Y = H Y, \quad (10)$$

where $H = [B' B + \lambda (D_d' D_d)]^{-1} B'$ is the hat matrix. Using equation (10), we can estimate the conditional expectation $\hat{\Phi}_Y(t)$ and $\hat{\Phi}_X(t)$:

$$\hat{\Phi}_Y^{Bs}(t) = B H Y, \quad \hat{\Phi}_X^{Bs}(t) = B H X. \quad (11)$$

The modified variables \tilde{X}_{Bs} and \tilde{Y}_{Bs} are:

$$\tilde{X}_{Bs} = X - \hat{\Phi}_X^{Bs}(t), \quad \tilde{Y}_{Bs} = Y - \hat{\Phi}_Y^{Bs}(t). \quad (12)$$

Now we can estimate the parametric and nonparametric component:

$$\hat{\beta}_{Bs} = (\tilde{X}'_{Bs} \tilde{X}_{Bs})^{-1} \tilde{X}'_{Bs} \tilde{Y}_{Bs}, \quad (13)$$

$$\hat{f}_{Bs}(t) = \hat{\Phi}_Y^{Bs}(t) - \hat{\Phi}_X^{Bs}(t) \hat{\beta}_{Bs}. \quad (14)$$

We can summarize our proposed estimator in the following algorithm:

- Step 1: Construct of B-splines as in (9).
- Step 2: Estimate the parameter vector of the spline function τ by (10).
- Step 3: Estimate $\Phi_Y(t)$ and $\Phi_X(t)$ according to (11).
- Step 4: Estimate \tilde{X}_{Bs} and \tilde{Y}_{Bs} according to (12).
- Step 5: Estimate the parametric and nonparametric components using (13) and (14) respectively.

5. Monto Carlo Simulation Study

In this section, we will investigate the performance of the three estimators: the spline smoothing estimator, kernel estimator (the used kernel function is biweight), and proposed (B-spline) estimator, by conduct a Monte Carlo simulation study. R software is used to perform our simulation study. For information about how to make Monte Carlo simulation studies using R, see [25, 26].

In our simulation study, Monte Carlo experiments were carried out based on equation (1). The simulated model is generated as follows:

1. The number of parametric coefficients is $q = 2, 4$, and 8 ; where $\beta_r = 1 \forall r = 1, \dots, q$.
2. Three functions have been used for the nonparametric component in the generated model, as in Abonazel *et al.* [6]: $f_1 = 1.5 \sin(\pi t)$, $f_2 = 1.5 \sin(\pi t^2)$, and $f_3 = 3 \sin(\pi t^3)$.
3. The explanatory variables ($x_r \forall r = 1, \dots, q$) are generated from the standard uniform distribution. While the variable t is generated from uniform from -1 to 1 .
4. The errors are generated from normal distribution with mean zero and standard deviation (σ) equals 0.5 and 1 .
5. The different sample sizes have been used as: $n = 100, 150, 300$, and 400 .
6. All Monte Carlo experiments involved 1000 replications and all the results of all separate experiments are obtained by precisely the same series of random numbers.

The goodness of fit of \hat{f} and $\hat{\beta}$ can be quantified by computing the average of mean squared error values (MSEs) for \hat{f} and $\hat{\beta}$ each iteration run $l = 1, \dots, 1000$. The MSEs of \hat{f} and $\hat{\beta}$ are calculated as:

$$MSE_l(\hat{f}) = \frac{1}{n} \sum_{i=1}^n [\hat{f}(t_i) - f(t_i)]^2,$$

$$MSE_l(\hat{\beta}) = \frac{1}{q} \sum_{r=1}^q (\hat{\beta}_r - \beta_r)^2,$$

where $\hat{f}(t_i)$ and $\hat{\beta}_r$ are the estimated values of $f(t_i)$ and β_r , respectively.

The results of simulation are recorded in tables 1-9. These

tables present the average of MSE (AMSE) for \hat{f} and $\hat{\beta}$ using error terms with different standard deviations, different sample sizes, different shapes of the nonparametric component, and different number of explanatory variables.

From tables 1-9, we can summarize some effects for the kernel, spline, and B-spline estimators in the following points:

- a) As n increases, the AMSEs decrease.
- b) As σ increases, the AMSEs increase.
- c) As q_j increases, the AMSEs increase.

In general, we can conclude that the AMSEs of B-spline estimator are smaller than the AMSEs of kernel and spline estimators in all simulation situations. But it notes that, for

the parametric component, the AMSEs of the kernel, spline, and B-spline estimators are relatively close.

Graphically, we illustrate the degree of goodness of fit of the kernel, spline, and B-spline estimators for the three nonparametric functions via different simulated PLMs. These models are generated based on different n and q_j , while $\sigma = 0.5$. The fitted curves of the estimators for the three nonparametric functions are shown in figures 1-3, respectively. From figure 1, we find that the fitted curve of B-spline estimator is closer to the true curve than kernel and spline estimators. The same results can be concluded from other figures. This means that B-spline estimator performs better regardless of the form of nonparametric function.

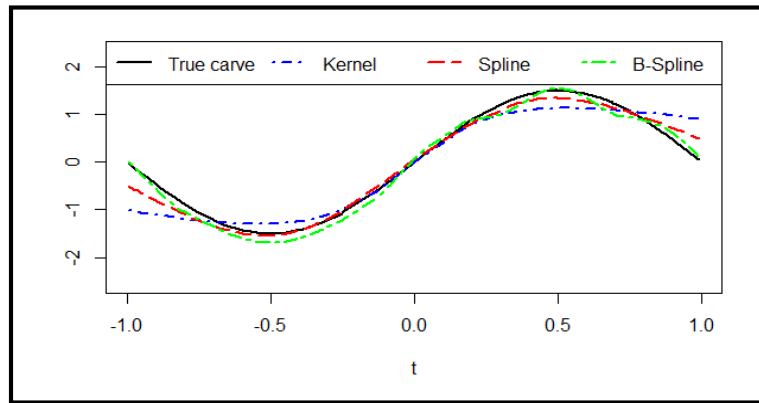


Figure 1. Fitted values for the estimators of f_1 when $n = 100$ and $q_j = 4$.

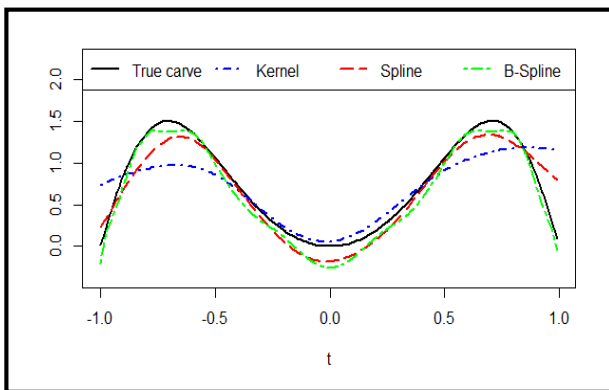


Figure 2. Fitted values for the estimators of f_2 when $n = 200$ and $q_j = 4$.

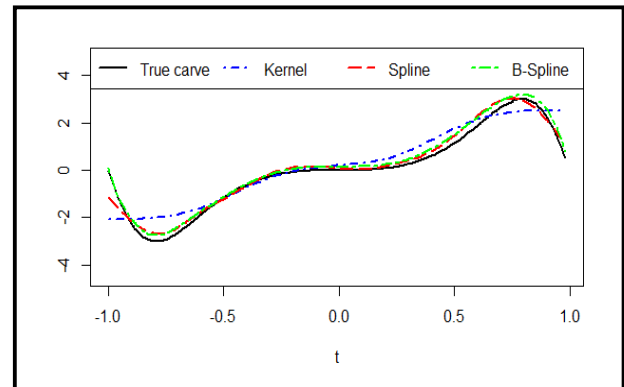


Figure 3. Fitted values for the estimators of f_3 when $n = 300$ and $q_j = 8$.

Table 1. AMSEs of the estimators when $q_j = 2$ and using f_1 .

n	Parametric component			Nonparametric component		
	Kernel	Spline	B-Spline	Kernel	Spline	B-Spline
$\sigma = 0.5$						
100	0.0340	0.0310	0.0307	0.1152	0.0345	0.0318
150	0.0236	0.0213	0.0212	0.0906	0.0229	0.0210
300	0.0110	0.0105	0.0105	0.0603	0.0117	0.0108
400	0.0076	0.0074	0.0073	0.0502	0.0086	0.0076
$\sigma = 1$						
100	0.1340	0.1348	0.1332	0.1842	0.1348	0.1281
150	0.0863	0.0849	0.0844	0.1333	0.0846	0.0814
300	0.0407	0.0406	0.0405	0.0820	0.0421	0.0390
400	0.0315	0.0312	0.0309	0.0683	0.0327	0.0300

Table 2. AMSEs of the estimators when $q_1 = 4$ and using f_1 .

n	Parametric component			Nonparametric component		
	Kernel	Spline	B-Spline	Kernel	Spline	B-Spline
$\sigma = 0.5$						
100	0.03615	0.03329	0.03273	0.13272	0.05133	0.04830
150	0.02343	0.02139	0.02135	0.10241	0.03334	0.03140
300	0.01100	0.01052	0.01045	0.06612	0.01762	0.01613
400	0.00807	0.00771	0.00766	0.05532	0.01308	0.01212
$\sigma = 1$						
100	0.13485	0.13189	0.13121	0.26194	0.20947	0.19795
150	0.19795	0.08450	0.08467	0.17358	0.12266	0.11863
300	0.04300	0.04278	0.04286	0.10344	0.06254	0.05985
400	0.03061	0.03046	0.03040	0.08347	0.04723	0.04505

Table 3. AMSEs of the estimators when $q_1 = 8$ and using f_1 .

n	Parametric component			Nonparametric component		
	Kernel	Spline	B-Spline	Kernel	Spline	B-Spline
$\sigma = 0.5$						
100	0.03878	0.03459	0.03459	0.03442	0.17919	0.09066
150	0.02403	0.02185	0.02175	0.12622	0.05517	0.05212
300	0.01139	0.01078	0.01073	0.07751	0.02846	0.02646
400	0.00801	0.00772	0.00771	0.06252	0.02073	0.01932
$\sigma = 1$						
100	0.13989	0.13736	0.13774	0.40771	0.35397	0.34801
150	0.08849	0.08748	0.08768	0.28368	0.23585	0.22915
300	0.04270	0.04245	0.04237	0.15443	0.11435	0.11180
400	0.03182	0.03156	0.03148	0.11953	0.08230	0.08025

Table 4. AMSEs of the estimators when $q_1 = 2$ and using f_2 .

n	Parametric component			Nonparametric component		
	Kernel	Spline	B-Spline	Kernel	Spline	B-Spline
$\sigma = 0.5$						
100	0.03044	0.03030	0.03067	0.03067	0.02920	0.02920
150	0.02108	0.02081	0.02107	0.02897	0.01927	0.02064
300	0.01042	0.01040	0.01049	0.01729	0.01729	0.01057
400	0.00730	0.00730	0.00730	0.01339	0.00705	0.00743
$\sigma = 1$						
100	0.12976	0.13125	0.13303	0.11063	0.12020	0.12770
150	0.08412	0.08412	0.08412	0.07328	0.074282	0.08113
300	0.04062	0.04024	0.04048	0.03941	0.03570	0.03877
400	0.03096	0.03101	0.03088	0.03165	0.03165	0.02980

Table 5. AMSEs of the estimators when $q_1 = 4$ and using f_2 .

n	Parametric component			Nonparametric component		
	Kernel	Spline	B-Spline	Kernel	Spline	B-Spline
$\sigma = 0.5$						
100	0.03256	0.03383	0.03299	0.06079	0.06079	0.04499
150	0.02064	0.02103	0.02078	0.04494	0.03674	0.03218
300	0.01042	0.01050	0.01042	0.02454	0.01878	0.01575
400	0.00779	0.00784	0.00780	0.01934	0.01469	0.01220
$\sigma = 1$						
100	0.13144	0.13327	0.13420	0.18275	0.19818	0.19119
150	0.08682	0.08721	0.08743	0.12871	0.13173	0.13099
300	0.04198	0.04219	0.04230	0.06620	0.06775	0.06443
400	0.03103	0.03124	0.03107	0.05020	0.05114	0.04766

Table 6. AMSEs of the estimators when $q_1 = 8$ and using f_2 .

n	Parametric component			Nonparametric component		
	Kernel	Spline	B-Spline	Kernel	Spline	B-Spline
$\sigma = 0.5$						
100	0.04326	0.03587	0.03447	0.19081	0.10514	0.08371
150	0.02721	0.02232	0.02181	0.14280	0.06277	0.05271
300	0.01245	0.01085	0.01076	0.09716	0.03225	0.02662
400	0.00854	0.00775	0.00772	0.08229	0.02391	0.01949
$\sigma = 1$						
100	0.14403	0.14010	0.13787	0.41425	0.38937	0.34886
150	0.09125	0.08875	0.08774	0.29883	0.25662	0.22931
300	0.04392	0.04274	0.04239	0.17429	0.12316	0.11194
400	0.03257	0.03173	0.03150	0.13960	0.08910	0.08055

Table 7. AMSEs of the estimators when $q_1 = 2$ and using f_3 .

n	Parametric component			Nonparametric component		
	Kernel	Spline	B-Spline	Kernel	Spline	B-Spline
$\sigma = 0.5$						
100	0.0358	0.0308	0.0307	0.1214	0.0334	0.0316
150	0.0232	0.0211	0.0211	0.0961	0.0223	0.0208
300	0.0114	0.0104	0.0105	0.0656	0.0116	0.0108
400						
$\sigma = 1$						
100	0.0079	0.0074	0.0073	0.0551	0.0084	0.0077
150	0.0850	0.0842	0.0844	0.1415	0.0840	0.0813
300	0.0415	0.0405	0.0405	0.0876	0.0409	0.0390
400	0.0312	0.0311	0.0309	0.0734	0.0317	0.0300

Table 8. AMSEs of the estimators when $q_1 = 4$ and using f_3 .

n	Parametric component			Nonparametric component		
	Kernel	Spline	B-Spline	Kernel	Spline	B-Spline
$\sigma = 0.5$						
100	0.06769	0.03614	0.03809	0.41000	0.07400	0.09201
150	0.04231	0.02227	0.02474	0.35542	0.04780	0.07353
300	0.01885	0.01087	0.01209	0.29061	0.02615	0.05764
400	0.01409	0.00789	0.00914	0.26622	0.01952	0.05209
$\sigma = 1$						
100	0.17044	0.13868	0.13683	0.53868	0.26659	0.24120
150	0.10396	0.08759	0.08776	0.43256	0.16530	0.16299
300	0.05035	0.04383	0.044289	0.32406	0.08537	0.10064
400	0.03646	0.03099	0.03179	0.29560	0.06565	0.08537

Table 9. AMSEs of the estimators when $q_1 = 8$ and using f_3 .

n	Parametric component			Nonparametric component		
	Kernel	Spline	B-Spline	Kernel	Spline	B-Spline
$\sigma = 0.5$						
100	0.07252	0.03723	0.04007	0.49354	0.11981	0.13605
150	0.04432	0.02309	0.02552	0.40289	0.07489	0.09920
300	0.02002	0.01106	0.01259	0.30973	0.03777	0.07000
400	0.01346	0.00786	0.00895	0.00893	0.02816	0.06105
$\sigma = 1$						
100	0.17287	0.14603	0.14343	0.72437	0.43057	0.39545
150	0.10833	0.09146	0.09124	0.55722	0.28829	0.27511
300	0.05078	0.04359	0.04425	0.38420	0.38420	0.15408
400	0.03769	0.03214	0.03292	0.33964	0.10125	0.12194

6. Conclusion

In this paper, we proposed new estimator for the PLM based on B-spline approach. Moreover, the performance of our estimator and Speckman's [8] and Abonazel's [6] estimators are investigated by a Monte Carlo simulation study. The simulation study is conducted to evaluate and compare the performance of these estimators (that based on kernel smoothing, spline smoothing, and B-spline smoothing) under different situations (such as: different shapes of the nonparametric component, different number of parametric variables, different sample sizes, and different standard deviations of error term). The simulation results confirm that our proposed estimator is more efficient than other estimators.

References

- [1] Engle, R. F., Granger, C. W., Rice, J., & Weiss, A. (1986). Semiparametric estimates of the relation between weather and electricity sales. *Journal of the American Statistical Association*, 81 (394), 310-320.
- [2] Heckman, N. E. (1986). Spline smoothing in a partly linear model. *Journal of the Royal Statistical Society: Series B (Methodological)*, 48 (2), 244-248.
- [3] Rice, J. (1986). Convergence rates for partially splined models. *Statistics & probability letters*, 4 (4), 203-208.
- [4] Chen, H., & Shiau, J. J. H. (1991). A two-stage spline smoothing method for partially linear models. *Journal of Statistical Planning and Inference*, 27 (2), 187-201.
- [5] Abonazel, M. R., & Gad, A. A. E. (2018). Robust partial residuals estimation in semiparametric partially linear model. *Communications in Statistics-Simulation and Computation*, 1-14.
- [6] Abonazel, M. R., Helmy, N. & Azazy, A. (2019). The Performance of Speckman Estimation for Partially Linear Model using Kernel and Spline Smoothing Approaches. *International Journal of Mathematical Archive*, 10 (6):10-18.
- [7] Robinson, P. M. (1988), Root-N-consistent semiparametric regression. *Econometrica*, 56 (4), 931-54.
- [8] Speckman, P. (1988). Kernel smoothing in partial linear models. *Journal of the Royal Statistical Society: Series B (Methodological)*, 50 (3), 413-436.
- [9] Hamilton, S. A., & Truong, Y. K. (1997). Local linear estimation in partly linear models. *Journal of Multivariate Analysis*, 60 (1), 1-19.
- [10] Carroll, R. J., Fan, J., Gijbels, I., & Wand, M. P. (1997). Generalized partially linear single-index models. *Journal of the American Statistical Association*, 92 (438), 477-489.
- [11] Yatchew, A. (1997). An elementary estimator of the partial linear model. *Economics Letters*, 57 (2), 135-143.
- [12] Yatchew, A. (2000). Scale economies in electricity distribution: a semiparametric analysis. *Journal of Applied Econometrics*, 15 (2), 187-210.
- [13] Yatchew, A. (2003). *Semiparametric regression for the applied econometrician*. Cambridge University Press.
- [14] Wang, L., Brown, L. D., & Cai, T. T. (2011). A difference based approach to the semiparametric partial linear model. *Electronic Journal of Statistics*, 5, 619-641.
- [15] Henderson, D. J., & Parmeter, C. F. (2015). Single-step estimation of a partially linear model. Working Papers, University of Miami, Department of Economics. Available at: https://www.bus.miami.edu/_assets/files/repec/WP2015-01.pdf
- [16] Elgohary, M. M., Abonazel, M. R., Helmy, N. M., & Azazy, A. R. (2019). New robust-ridge estimators for partially linear model. *International Journal of Applied Mathematical Research*, 8 (2): 46-52.
- [17] Green, P. J., & Silverman, B. W. (1994). *Nonparametric regression and generalized linear models: a roughness penalty approach*. Chapman and Hall/CRC.
- [18] Ruppert, D., Wand, M. P., & Carroll, R. J. (2003). *Semiparametric regression (No. 12)*. Cambridge university press.
- [19] Wasserman, L. (2006). *All of nonparametric statistics*. Springer Science & Business Media.
- [20] De Boor, C. (1978). *A practical guide to splines. (Vol. 27, p. 325)*. New York: springer-verlag.
- [21] Eilers, P. H., & Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science*, 11 (2), 89-102.
- [22] Eilers, P. H., & Marx, B. D. (2010). *Splines, knots, and penalties. Wiley Interdisciplinary Reviews: Computational Statistics*, 2 (6), 637-653.
- [23] Eilers, P. H., Marx, B. D., & Durbán, M. (2015). Twenty years of P-splines. *Statistics and Operations Research Transactions*, 39 (2), 0149-186.
- [24] Marx, B. D., & Eilers, P. H. (1999). Generalized linear regression on sampled signals and curves: a P-spline approach. *Technometrics*, 41 (1), 1-13.
- [25] Abonazel, M. R. (2018). A practical guide for creating Monte Carlo simulation studies using R. *International Journal of Mathematics and Computational Science*, 4 (1), 18-33.
- [26] Abonazel, M. R. (2019). Advanced statistical techniques using R: outliers and missing data, *Annual Conference on Statistics, Computer Sciences and Operations Research, Vol. 54, Dec. 2019, Faculty of Graduate Studies for Statistical Research, Cairo University, Egypt*.