



Adaptive Partially Linear Regression Models by Mixing Different Estimates

Magda Mohamed Mohamed Haggag

Department of Statistics, Mathematics, and Insurance, Faculty of Commerce, Damanhour University, Damanhour, Egypt

Email address:

magmhag@yahoo.com, magda.haggag@com.dmu.edu.eg

To cite this article:

Magda Mohamed Mohamed Haggag. Adaptive Partially Linear Regression Models by Mixing Different Estimates. *American Journal of Theoretical and Applied Statistics*. Vol. 8, No. 5, 2019, pp. 157-168. doi: 10.11648/j.ajtas.20190805.11

Received: June 24, 2019; **Accepted:** July 26, 2019; **Published:** September 4, 2019

Abstract: This paper proposes adapting the semiparametric partial model (PLM) by mixing different estimation procedures defined under different conditions. Choosing an estimation method of PLM, from several estimation methods, is an important issue, which depends on the performance of the method and the properties of the resulting estimators. Practically, it is difficult to assign the conditions which give the best estimation procedure for the data at hand, so adaptive procedure is needed. Kernel smoothing, spline smoothing, and difference based methods are different estimation procedures used to estimate the partially linear model. Some of these methods will be used in adapting the PLM by mixing. The adapted proposed estimator is found to be a square root-consistent and has asymptotic normal distribution for the parametric component of the model. Simulation studies with different settings, and real data are used to evaluate the proposed adaptive estimator. Correlated and non-correlated regressors are used for the parametric components of the semiparametric partial model (PLM). Best results are obtained in the case of correlated regressors than in the non-correlated ones. The proposed adaptive estimator is compared to the candidate model estimators used in mixing. Best results are obtained in the form of less risk error and less convergence rate for the proposed adaptive partial linear model (PLM).

Keywords: Backfitting Method, Combining Regression Procedures, Difference Based Method, Partially Linear Models, Profile Likelihood Method, Semiparametric Regression, Spline Smoothing

1. Introduction

Different methods are used for combining regression models. Most combinations are imposed to parametric or nonparametric candidate models (see for example, [14, 29, 30, 3, 20, 31]). Yang (2001) proposed a method for combining nonparametric regression procedures, this method is called adaptive regression procedures by mixing (ARM). This method worked under Gaussian errors, and can be used where there are multiple candidate error distributions.

Partially linear models have been first considered by these researchers [7, 4, 23, 27]. A partial linear model (PLM) is a semiparametric regression model which contains two components, one is parametric and the other is nonparametric. Parametric estimation methods are used to estimate the parametric component, and the nonparametric estimation methods are used to estimate the nonparametric one [9, 19].

This paper proposes adapting the partial linear model

(PLM) by combining different estimation procedures and the resulting regression model called adaptive partial linear model (APLM). Different estimation schemes are used in estimating PLM, such as kernel smoothing, spline smoothing, and difference-based method. Robinson, Speckman, profile likelihood, backfitting are different estimation procedures based on kernel regression [23, 27]. In this work, different estimation procedures based on using kernel regression, spline smoothing regression methods are used in adapting the PLM.

The rest of the paper is considered as follows section (2) presents the partially linear model (PLM), its estimation methods and their statistical properties. Section (3) introduces the proposed adaptive APLM and its theoretical properties. Section (4) considers numerical studies using simulation studies under different settings and real data example. The conclusions of this work are presented in section (5).

2. The Partially Linear Regression Model (PLM)

2.1. The Model

A partially linear model (PLM) takes the following form:

$$Y = X^T \beta + m(Z) + \varepsilon, \tag{1}$$

where X is an $(n \times p)$ matrix of regressors in the parametric component of the model, Z is an $(n \times q)$ matrix of regressors in the nonparametric component, β is a $(p \times 1)$ vector of unknown parameters, m is an unknown function (a nonparametric function) from \mathbb{R}^q to \mathbb{R} , ε is an independent vector of random errors with mean zero and finite variance σ^2 . A PLM in (1) is a semiparametric model since it contains parametric and nonparametric components. PLM is a preferred regression model than a fully parametric model and a fully nonparametric one, since PLM is a more flexible than the first one and combat the curse of dimensionality which is a well-known problem in nonparametric regression models.

2.2. Estimation Methods of PLM

Estimation of the PLM in (1) will be first started by estimating the parametric component, i.e., estimating the unknown parameter vector β . The resulting estimator of β will be then used to estimate the nonparametric function $m(Z)$. Several methods are used to estimate PLM, which may be divided into three sections according to the method of estimating the nonparametric component:

- a) Methods based on kernel regression [23, 27, 2, 11, 25, 19]. These methods will be presented in sections (2-2-1): (2-2-3).
- b) Methods based on regression splines [5, 21, 26-28, 34]. The smoothing spline method will be presented in section (2-2-4).
- c) Methods based on differences [32, 33, 17, 29].

Choosing an estimation method of PLM, from several estimation methods, is an important issue, which depends on the performance of the method and the properties of the resulting estimators. The interest in this work will be on presenting some estimation methods that will be used to form APLM.

The following assumptions are needed to hold through the paper.

Assumptions (2-2)

- 1. Assume that the set of $\{X_i, Y_i\}$, $i=1, 2, \dots, n$ are i. i. d. design inputs.
- 2. Assume that the random errors ε_i 's are independent of (X_i, Z_i) , i.e., $E(\varepsilon \setminus X, Z) = 0$, and $E(\varepsilon^2 \setminus X, Z) < \infty$.
- 3. Assume that $E(X \setminus Z = 0)$, and the covariance matrix of X given Z , i.e., $\text{cov}(X \setminus Z) = E(\tilde{X}\tilde{X}^T)$ is a p. d. matrix, where $\tilde{X} = (X - E(X \setminus Z))$.

- 4. It is assumed that the two expectations, $E(X \setminus Z)$ and $E(Y \setminus Z)$ have bounded and continuous second derivatives.
- 5. It is assumed that both $E(X \setminus Y, Z)$ and $E(XX^T \setminus Y, Z)$ have bounded first derivatives.
- 6. Assume that the first two derivatives of $m(z)$ are Lipschitz continuous of order one.
- 7. If $K_i(z)$ is a weight function defined as:

$$K_i(z) = \frac{k\left(\frac{z-z_i}{h_n}\right)}{\sum_{j=1}^n k\left(\frac{z-z_j}{h_n}\right)}, \quad i=1, 2, \dots, n,$$

is a constant, then the following conditions are satisfied by $K_i(z)$:

$$\max_{1 \leq i \leq n} \sum_{j=1}^n K_i(z_j) = O$$

$$\max_{1 \leq j \leq n} \sum_{i=1}^n K_i(z_j) = O$$

The kernel function $k(z)$ is a symmetric density function with compact support and satisfies:

$$\int k(t) dt = 1,$$

$$\int tk(t) dt = 0,$$

$$\int t^2 k(t) dt = 1.$$

- 1. The density function of Z and (Y, Z) are bounded away from zero and have bounded continuous second derivatives.

2.2.1. Least Squares Method (Robinson's Estimator)

Robinson (1988) [23], proposed a feasible least squares estimator of β using a Nadaraya-Watson kernel estimator of the nonparametric function $m(Z)$. Consider the conditional expectation of PLM in (1) given Z ,

$$E(Y \setminus Z) = E(X \setminus Z)\beta + m(Z) \tag{2}$$

Since $E(m(Z) \setminus Z) = m(Z)$, and $E(\varepsilon \setminus X, Z) = 0$. Subtracting (2) from (1) result in:

$$Y - E(Y \setminus Z) = [X - E(X \setminus Z)]^T \beta + \varepsilon \tag{3}$$

IF $[Y - E(Y \setminus Z)]$ and $[X - E(X \setminus Z)]$ are replaced by \tilde{Y} and \tilde{X} , respectively, such that:

$$\tilde{Y} = [Y - E(Y \setminus Z)], \tilde{X} = [X - E(X \setminus Z)]. \quad (4)$$

Equation (3) is simply a version of the standard linear model. Robinson (1988) proposed replacing the unknown conditional expectations by their kernel estimates as follows:

$$\hat{Y}_i = \hat{E}(Y_i \setminus Z_i) = \frac{1}{n} \sum_{j=1}^n Y_j K_h(Z_j - Z_i) / \hat{m}(Z_i), \quad (5)$$

$$\hat{X}_i = \hat{E}(X_i \setminus Z_i) = \frac{1}{n} \sum_{j=1}^n X_j K_h(Z_j - Z_i) / \hat{m}(Z_i), \quad (6)$$

where,

$$\hat{m}(Z_i) = \frac{1}{n} \sum_{j=1}^n K_h(Z_j - Z_i), \quad (7)$$

and $K_h(\cdot)$ is a kernel function, with a bandwidth=h, defined as follows:

$$K_h(Z_j - Z_i) = \prod_{l=1}^q \frac{1}{h_l} K\left(\frac{Z_j - Z_i}{h_l}\right). \quad (8)$$

So, β can be estimated by the standard linear regression. By subtracting $E(X \setminus Z)\beta$ from both sides of (2) and getting:

$$E[(Y - X^T \beta) \setminus Z] = m(Z), \quad (9)$$

which means that β can be estimated by least squares of Y on X, plugging $X^T \hat{\beta}$ in (9) a nonparametric regression estimate can be obtained for $m(Z)$. The proposed least squares estimator of β will be as follows:

$$\hat{\beta}_R = \left[\frac{1}{n} \sum_{i=1}^n (X_i - \hat{X}_i)(X_i - \hat{X}_i)^T 1_i \right]^{-1} \left[\frac{1}{n} \sum_{i=1}^n (X_i - \hat{X}_i)(Y_i - \hat{Y}_i)^T 1_i \right] \quad (10)$$

where $1_i = 1(\hat{m}(Z_i \geq b))$, and b is a trimming parameter, $b > 0$ satisfies $b \rightarrow 0$ as $n \rightarrow \infty$. The least squares estimator, $\hat{\beta}$ is used in estimating the nonparametric component $m(Z)$ as follows:

$$m(Z) = E[(Y - X^T \hat{\beta}) \setminus Z],$$

then,

$$\hat{m}(Z) = \frac{\frac{1}{n} \sum_{i=1}^n (Y_i - X_i^T \hat{\beta}) K_h(Z_i - Z)}{\frac{1}{n} \sum_{i=1}^n K_h(Z_i - Z)}. \quad (11)$$

Theorem (1)

Under assumptions (2-2), the Robinson estimator $\hat{\beta}_R$ defined in (10) has the following properties:

- a) $\hat{\beta}_R$ is a \sqrt{n} -consistent estimator of β .
- b) $\sqrt{n}(\hat{\beta}_R - \beta) \xrightarrow{d} N(0, \Omega_0)$,

where,

$$\Omega_0 = \hat{\Phi}_0^{-1} \hat{\Psi}_0 \hat{\Phi}_0^{-1},$$

$$\hat{\Phi}_0 = \frac{1}{n} \sum_{i=1}^n (X_i - \hat{X}_i)(X_i - \hat{X}_i)^T 1_i,$$

$$\hat{\Psi}_0 = \frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_i^2 (X_i - \hat{X}_i)(X_i - \hat{X}_i)^T 1_i,$$

and,

$$\hat{\varepsilon}_i = (Y_i - \hat{Y}_i) - (X_i - \hat{X}_i)^T \hat{\beta}_R. \quad (12)$$

The proof of this theorem is given in Appendix (A).

It is found that the asymptotic distribution of $\hat{\beta}_R$ does not depend on the bandwidth h. So, $\hat{\beta}_R$ does not provide a method for choosing h in practice [23, 16, 15].

2.2.2. The Speckman Estimator

Speckman (1988) [27] derived the parametric and nonparametric components estimators of PLM based on the modified variables of X and Y in (4). Using the sample values of (Y_i, X_i, T_i) , then:

$$\hat{\beta}, X = \begin{bmatrix} X_{11} & \dots & X_{1p} \\ \vdots & \ddots & \vdots \\ X_{n1} & \dots & X_{np} \end{bmatrix}, m(T) = \begin{bmatrix} m(T_1) \\ m(T_2) \\ \vdots \\ m(T_n) \end{bmatrix}$$

Speckman considered the partial regression plots to form the estimators of the parametric and nonparametric components of (1). The algorithm will be as follows:

- a) Estimating the parametric component β :

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

where,

$$\tilde{X} = (I - S)X, \tilde{Y} = (I - S)Y,$$

And S is a smoother matrix defined by its elements as follows:

$$S_{ij} = \frac{K_H(Z_i - Z_j)}{\sum_{i=1}^n K_H(Z_i - Z_j)} \quad (14)$$

Estimating the nonparametric component $m(T)$ as follows:

$$\hat{m} = S(Y - X\hat{\beta}_{Speck}) \quad (15)$$

An updating step for both β and m can be used as an iteration step up to convergence for the estimation [27, 18].

Speckman (1988) [27] showed that the estimator $\hat{\beta}_{Speck}$ in (13) has asymptotic normality as Robinson estimator.

Theorem (2)

Under assumptions (2-2), the least squares estimator $\hat{\beta}$ defined in (10) has the following properties:

a) $\hat{\beta}_{Speck}$ is a \sqrt{n} -consistent estimator of β .

b) $\sqrt{n}(\hat{\beta}_{Speck} - \beta) \xrightarrow{d} N(0, \Omega_1)$,

where,

$$\Omega_1 = \hat{\Phi}_1^{-1} \hat{\Psi}_1 \hat{\Phi}_1^{-1},$$

$$\hat{\Phi}_1 = \tilde{X}^T \tilde{X},$$

$$\hat{\Psi}_1 = \sigma^2 \tilde{X}^T (I - S)(I - S)^T \tilde{X},$$

The proof of this theorem is in Appendix (A).

2.2.3. The Profile-Likelihood Method

Severini and Wong (1992) [25] proposed estimating the PLM based on the conditional distribution of Y given X and Z . They found that this conditional distribution is parametric. They started by fixing the parameter vector β to estimate the nonparametric function $m_\beta(Z)$ which depends on the fixed β . The resulting estimator $\hat{m}_{\hat{\beta}_{Prof}}(Z)$ is then used to construct the profile likelihood for β . They found that their estimator $\hat{\beta}_{Prof}$ is estimated at \sqrt{n} -rate and has an asymptotic normal distribution and is asymptotically efficient. Also, the resulting estimator $\hat{m}_{\hat{\beta}_{Prof}}(Z)$ is a consistent estimator of $m(Z)$. The procedure of the profile likelihood algorithm is abbreviated be as follows:

a) Estimating the parametric component β :

$$\hat{\beta}_{Prof} = (\tilde{X}^T \tilde{X})^{-1} \tilde{X}^T \tilde{Y}, \quad (16)$$

where,

$$\tilde{X} = (I - S)X, \quad \tilde{Y} = (I - S)Y,$$

And S is a smoother matrix defined by its elements as follows:

$$S_{ij} = \frac{K_H(Z_i - Z_j)}{\sum_{i=1}^n K_H(Z_i - Z_j)} \quad (17)$$

b) Estimating the nonparametric component $m(T)$ as follows:

$$\hat{m} = S(Y - X\hat{\beta}_{Prof}) \quad (18)$$

The Speckman and Profile-likelihood methods are coinciding for the estimation of PLM [2, 11, 18].

2.2.4. The Backfitting Method

The backfitting method is referred to the studies [2, 11] as an iterative algorithm for estimating an additive model. The procedure of the backfitting algorithm will be as follows:

a) Estimating the parametric component β :

$$\hat{\beta}_{back} = (X^T \tilde{X})^{-1} X^T \tilde{Y}, \quad (19)$$

where,

$$\tilde{X} = (I - S)X, \quad \tilde{Y} = (I - S)Y. \quad (20)$$

b) Estimating the nonparametric component $m(T)$ as follows:

$$\hat{m} = S(Y - X\hat{\beta}_{back}) \quad (21)$$

where S is a smoother matrix as defined in (14).

Opsomer, and Ruppert, 1999 showed that $\hat{\beta}_{back}$ is a \sqrt{n} -consistent estimator of β for the right choice of the bandwidth. They proposed a method based the Empirical bias bandwidth selection of [24].

Theorem (3)

Under assumptions (2-2), the backfitting estimator $\hat{\beta}_{back}$ defined in (19) has the following properties:

a) $\hat{\beta}_{back}$ is a \sqrt{n} -consistent estimator of β . (The proof is in Theorem 2.2 of the study [22]).

b) $\sqrt{n}(\hat{\beta}_{back} - \beta) \xrightarrow{d} N(0, \Omega_3)$. (The proof is in Corollary 2.1 of study [22]).

where,

$$\Omega_3 = \hat{\Phi}_3^{-1} [\sigma^2 I + (X^T SX - X^T SS^T X) \hat{\Phi}_3^{-1}],$$

$$\hat{\Phi}_3 = (X^T \tilde{X}), \text{ and } \tilde{X} \text{ is as defined in (20).}$$

and,

$$\hat{\epsilon}_i = (Y_i - \hat{Y}_i) - (X_i - \hat{X}_i)^T \hat{\beta}.$$

The proof of this theorem is in the Appendix (A).

2.2.4. Smoothing Spline Based-Method

The estimation of the components of the PLM, based on smoothing spline, is performed by minimizing the following sum of square criterion (Q):

$$Q(\beta, m) = \sum_{i=1}^n (Y_i - X_i^T \beta - m(z_i))^2 + \lambda \int_a^b (m''(z))^2 dz, \quad (22)$$

The objective of Equation (22) is to estimate the parameter vector β and the nonparametric smooth function $m(z)$, where $m''(\cdot)$ is the second derivative of $m(\cdot)$, and λ is a smoothing parameter which controls the trade-off between having a linear function when $\lambda \rightarrow 0$ or having a wiggly function when $\lambda \rightarrow \infty$ [6, 8, 24].

Applying the estimation scheme of [27], the spline smoothing algorithm will be as follows:

- a) Given a smoothing parameter λ , find a smoothing matrix S_λ which depend on λ .
- b) Estimate the parametric component β as:

$$\hat{\beta}_{Spline} = (\tilde{X}^T \tilde{X})^{-1} \tilde{X}^T \tilde{Y}, \quad (23)$$

where,

$\tilde{X} = (I - S_\lambda)X$, $\tilde{Y} = (I - S_\lambda)Y$, are the residuals of both X and Y, respectively.

- a) Estimate the nonparametric component $m(z)$ as follows:

$$\hat{m} = S_\lambda (Y - X \hat{\beta}_{Spline}) \quad (24)$$

- b) Choosing different values of λ until minimizing function Q in (22). The studies [6, 4] suggested using a generalized cross-validation method as a way of choosing λ .

Theorem (4)

Under assumptions (2-2), smoothing spline estimator $\hat{\beta}_{Spline}$ defined in (23) has the following properties:

- a) $\hat{\beta}_{Spline}$ is a \sqrt{n} -consistent estimator of β .
- b) $\sqrt{n}(\hat{\beta}_{Spline} - \beta) \xrightarrow{d} N(0, \Omega_3)$.

where,

$$\Omega_3 = \hat{\Phi}_3^{-1} \hat{\Psi}_3 \hat{\Phi}_3^{-1},$$

and,

$$\hat{\Phi}_3 = \frac{1}{n} \tilde{X}^T \tilde{X},$$

$$\hat{\Psi}_3 = \frac{1}{n} \tilde{X}^T (I - S_\lambda) \hat{\epsilon}^2 (I - S_\lambda)^T \tilde{X},$$

and,

$$\hat{\epsilon}^2 = \text{diag}(\hat{\epsilon}_1^2, \dots, \hat{\epsilon}_n^2).$$

The proof of this theorem is in Appendix (A).

3. The Proposed Adaptive Partially Linear Model (APLM)

3.1. The Model

Estimation methods of the PLM perform well under different conditions. Therefore, adaptive partial linear model (APLM) is proposed to handle the practical problems under any condition. Consider the i th copies (Y_i, X_i, Z_i) for the PLM in (1), where

$$Y_i = X_i^T \beta + m(Z_i) + \epsilon_i, \quad i=1, 2, \dots, n$$

where X_i is p-vector of covariates of the parametric part, Z_i is a q-vector covariates of the nonparametric part, and the error terms ϵ_i 's are assumed to be independent with a conditional mean zero given the covariates X and Z.

3.2. APLM Algorithm

Let δ_j , $j=1, 2, 3$ denotes the regression estimation procedures used in this work such that:

- δ_1 : is the Speckman estimation procedure,
- δ_2 : is the backfitting estimation procedure, and
- δ_3 : the spline smoothing procedure.

The proposed APLM algorithm for mixing the three procedures is determined as follows:

1. The used data are splitted into two equal sections, the first section is used for estimation and the second is used for prediction evaluation.
2. The estimators of β and $m(Z)$ are obtained for each method i.e., $\hat{\beta}_j$ and $\hat{m}_j(Z)$, $j=1, 2, 3$, using the first section of data ($i=1, 2, \dots, n/2$).
3. The error distribution is computed for each method j using the second section of data ($i=n/2+1, \dots, n$):

$$Q_j = \prod_{i=\frac{n}{2}+1}^n f\left(\frac{Y_i - \hat{Y}_{ij}}{\hat{\sigma}_j}\right), \quad (25)$$

where $\hat{Y}_{ij} = x_i^T \hat{\beta}_j + \hat{m}_j(Z_i)$, $i = \frac{n}{2} + 1, \dots, n$, and $j=1, 2, 3$. If the function f is normal, then the function Q_j will be as follows:

$$Q_j = (2\pi)^{-n/4} \hat{\sigma}_j^{-n/2} \exp\left(-\sum_{i=\frac{n}{2}+1}^n \frac{(Y_i - \hat{Y}_{ij})^2}{2\hat{\sigma}_j^2}\right) \quad (26)$$

4. For mixing the three estimation procedures, the following quantity is computed as a weight used for mixing:

$$W_j = \frac{Q_j}{\sum_{j=1}^3 Q_j} \tag{27}$$

5. The above four steps will be repeated many times to obtain the average of W_j , i.e. \bar{W}_j . If \bar{W}_j is the average weight for procedure δ_j , then the adaptive prediction of PLM will be obtained as:

$$\hat{Y}_{APLM} = \sum_{j=1}^3 \bar{W}_j \hat{Y}_j = X^T \hat{\beta}_{APLM} + \hat{m}_{APLM}, \tag{28}$$

where,

$$\hat{Y}_j = X^T \hat{\beta}_j + \hat{m}_j(Z), \tag{29}$$

$$\hat{\beta}_{APLM} = \sum_{j=1}^n \bar{W}_j \hat{\beta}_j, \tag{30}$$

and

$$\hat{m}_{APLM} = \sum_{j=1}^3 \bar{W}_j \hat{m}_j \tag{31}$$

Theorem (5)

Under assumptions (2-2), the APLM estimator $\hat{\beta}_{APLM}$ of β defined in (30) has the following asymptotic properties:

- a) $\hat{\beta}_{APLM}$ is a \sqrt{n} -consistent estimator of β .
- b) $\sqrt{n}(\hat{\beta}_{APLM} - \beta) \xrightarrow{d} N(0, \Omega_n)$,

The proof of this theorem is in Appendix (A).

Corollary (1)

Let $\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k$ be a sequence of independent estimators for the parametric component β in PLM. If the sequence of $\bar{W}_j \hat{\beta}_j$, for $j=1, 2, \dots, k$, converges in distribution to

$$N(\bar{W}_j \beta, \Omega_j), \text{ where } \bar{W}_j \text{ is a quantity such that } \sum_{j=1}^k \bar{W}_j = 1,$$

and Ω_j is a covariance matrix of $\bar{W}_j \hat{\beta}_j$, then the linear combination

$$\sum_{j=1}^k \bar{W}_j \hat{\beta}_j \text{ converges in distribution to}$$

$$N(\beta, \Omega_n), \text{ where } \Omega_n = \sum_{j=1}^k \Omega_j.$$

The proof of Corollary (1) is omitted, since this corollary is a generalization of Theorem (4).

4. Numerical Experiments

Simulation and real data will be used for comparing the proposed adaptive APLM with individual regression

procedures. The performance of APLM will be evaluated using the squared loss between in prediction between Y and \hat{Y}_j as will be shown in (32).

4.1. Simulation Studies

4.1.1. Simulation Assumptions

Three regression estimation procedures δ_j 's, $j=1, 2, 3$ are defined as follows:

- δ_1 : for Speckman's estimator.
- δ_2 : for backfitting estimator.
- δ_3 : for smoothing spline estimator.

The following assumptions are used:

1. It is assumed that there are two regressors in the parametric component, X_1 and X_2 , and one regressor, Z in the nonparametric component of the model.
2. The bandwidth parameter, h , used in computing the Speckman and backfitting estimators is assumed equal 0.5.
3. The smoothing spline parameter, λ , used in computing the smoothing spline estimator is assumed equal 0.95.
4. $X = (X_1, X_2)$, $X_1 \text{ uniform}(0,1)$, $X_2 \text{ uniform}(0,1)$, $\beta = (1, -1.5)$, $Z \text{ uniform}(0,1)$, and $\varepsilon \sim N(0,0.5)$.
5. Two cases with five true models will be used in the simulation with different sample sizes: $n=20, 50, 100$, and 200 as follows:

Case One: Dependent Regressors:

In this case, it is assumed that there is a relation between X_1, X_2 , and Z as follows:

$$X_j = \rho Z + (1 - \rho)U, \text{ for } j=1, 2,$$

where, $\rho = 0.75$, and $U \text{ uniform}(0,1)$.

Case Two: Independent Regressors:

In this case, it is assumed that there is no relation between X_1, X_2 , and Z .

The Models:

Five models will be used in this study as follows:

Model (1): $Y = X^T \beta + 2\text{Sin}(3Z) + \varepsilon$

Model (2): $Y = X^T \beta + (Z - 0.5)^2 + \varepsilon$

Model (3): $Y = X^T \beta + \exp(2Z - 0.5) + \varepsilon$

Model (4): $Y = X^T \beta + (Z - 0.5)^2 + \exp(2Z - 0.5) + \varepsilon$.

Model (5): $Y = X^T \beta + 2Z^2 + \exp(2Z - 0.5) + \varepsilon$.

In each case the three regression estimation methods δ_1 , δ_2 , and δ_3 are computed and compared with APLM using different sample sizes $n=10, 20, 50, 100$, and 200. The average squared loss (ASL) in prediction is used as an evaluation criterion computed over 1000 replications as follows:

$$ASL = \frac{1}{1000} (Y - \hat{E}(Y \setminus X, Z))^T (Y - \hat{E}(Y \setminus X, Z)), \tag{32}$$

where, $\hat{E}(Y \setminus X, Z) = X^T \hat{\beta} + \hat{m}(Z)$.

4.1.2. Simulation Results

a) Case One: (dependent regressors):

Tables 1-5 shows the average squared loss (ASL) in (32) for the five models. Best results (bold and italic numbers) are obtained for all model in the form of less ASL. In Table 3 for model (3), the ASL for smoothing spline estimator equals that of APLM when the sample size n=200.

Table 1. The average squared loss (ASL) for Model (1)-dependent case.

N	Speckman' estimator	Backfitting Estimator	Smoothing Spline estimator	APLM Estimator
10	0.0032530	0.0313660	0.0050805	<i>0.0032259</i>
20	0.0064236	0.0313868	0.0077545	<i>0.0063990</i>
50	0.0144990	0.0913750	0.0094870	<i>0.0044983</i>
100	0.0374200	0.2010470	0.0216320	<i>0.0374200</i>
200	0.0673550	0.4504880	0.0389500	<i>0.0373550</i>

Table 2. The average squared loss (ASL) for Model (2)-dependent case.

N	Speckman' estimator	Backfitting Estimator	Smoothing Spline estimator	APLM Estimator
10	0.001376	0.001378	0.001444	<i>0.001366</i>
20	0.002481	0.002798	0.002562	<i>0.002454</i>
50	0.009949	0.019527	0.009439	<i>0.009429</i>
100	0.022103	0.023630	0.021530	<i>0.021529</i>
200	0.041432	0.041650	0.038870	<i>0.038317</i>

Table 3. The average squared loss (ASL) for Model (3)-dependent case.

N	Speckman' estimator	Backfitting Estimator	Smoothing Spline estimator	APLM Estimator
10	0.001622	0.002988	0.001740	<i>0.001621</i>
20	0.001522	0.001773	0.001484	<i>0.001422</i>
50	0.013325	0.013379	0.011973	<i>0.011435</i>
100	0.024988	0.025734	0.024012	<i>0.024011</i>
200	0.051800	0.055376	<i>0.051038</i>	<i>0.051038</i>

Table 4. The average squared loss (ASL) for Model (4)-dependent case.

N	Speckman' estimator	Backfitting Estimator	Smoothing Spline estimator	APLM Estimator
10	0.003179	0.0035487	0.003090	<i>0.003063</i>
20	0.004414	0.0061213	0.004483	<i>0.004410</i>
50	0.011674	0.0220990	0.011160	<i>0.011151</i>
100	0.025925	0.0290460	0.055310	<i>0.025755</i>
200	0.053093	0.0554160	0.050940	<i>0.050660</i>

Table 5. The average squared loss (ASL) for Model (5)-dependent case.

N	Speckman' estimator	Backfitting Estimator	Smoothing Spline estimator	APLM Estimator
10	0.020710	0.223490	0.012223	<i>0.020600</i>
20	0.088107	0.814354	0.037381	<i>0.037101</i>
50	0.210667	3.237850	0.015437	<i>0.012106</i>
100	0.618449	5.150950	0.033938	<i>0.031844</i>
200	1.106305	9.119636	0.040787	<i>0.000400</i>

b) Case Two: (independent regressors):

Tables 6-10 shows the average squared loss (ASL) in (32) for the five models. Best results (bold and italic numbers) are obtained for all model in the form of less ASL. Some best results are obtained for smoothing spline estimator, in Table 6 for model (1) when n=50, in Table 7 for model (2) when

n=200, in Table 9 for model (4) when n=100, and in Table 10 for model (5) when n=100 and n=200.

Table 6. The average squared loss (ASL) for Model (1)-Independent case.

n	Speckman' estimator	Backfitting Estimator	Smoothing Spline estimator	APLM Estimator
10	0.003253	0.031366	0.005080	<i>0.003250</i>
20	0.006423	0.031386	0.007754	<i>0.006399</i>
50	0.014449	0.091375	<i>0.009487</i>	0.014498
100	0.037420	0.201047	0.021632	<i>0.037420</i>
200	0.067355	0.450488	0.038950	<i>0.067355</i>

Table 7. The average squared loss (ASL) for Model (2)-Independent case.

n	Speckman' estimator	Backfitting Estimator	Smoothing Spline estimator	APLM Estimator
10	0.001376	0.001378	0.001444	<i>0.001366</i>
20	0.002481	0.002798	0.002562	<i>0.002445</i>
50	0.009949	0.019527	0.009439	<i>0.009390</i>
100	0.022103	0.023630	0.021530	<i>0.021499</i>
200	0.041432	0.041650	<i>0.038870</i>	0.040317

Table 8. The average squared loss (ASL) for Model (3)-Independent case.

n	Speckman' estimator	Backfitting Estimator	Smoothing Spline estimator	APLM Estimator
10	0.006857	0.001601	0.007170	<i>0.000685</i>
20	0.002290	0.005615	0.002289	<i>0.002235</i>
50	0.016869	0.016980	0.016083	<i>0.016868</i>
100	0.023635	0.023631	0.021648	<i>0.021635</i>
200	0.052850	0.054238	0.053467	<i>0.051751</i>

Table 9. The average squared loss (ASL) for Model (4)-Independent case.

n	Speckman' estimator	Backfitting Estimator	Smoothing Spline estimator	APLM Estimator
	0.001621	0.002988	0.001740	<i>0.001620</i>
	0.001522	0.001773	0.001484	<i>0.001512</i>
	0.013325	0.013379	0.011973	<i>0.011325</i>
	0.024988	0.025734	<i>0.024012</i>	0.024988
	0.051800	0.055376	0.050609	<i>0.050606</i>

Table 10. The average squared loss (ASL) for Model (5)-Independent case.

n	Speckman' estimator	Backfitting Estimator	Smoothing Spline estimator	APLM Estimator
10	0.051370	0.086246	0.038444	<i>0.006900</i>
20	0.032373	0.697480	0.013787	<i>0.013684</i>
50	0.034683	4.805047	0.013186	<i>0.013146</i>
100	0.400107	5.507278	<i>0.022834</i>	5.507278
200	0.961307	9.199111	<i>0.056496</i>	9.199111

4.2. Real Data Examples

The APLM is illustrated using a real data example called the current population survey (CPS). The CPS data are taken from a population survey in 1985 in USA. (See Berndt, 1991). The CPS data contains 534 observations on 11 variables to study the determinants of wages. The variables are wage, education, experience, age, ethnicity, region, gender, occupation, sector, union, and married. In this work, the interest will be in the effect of gender and education variables (parametrically), and experience variable (nonparametrically) on the person's wage.

Table 11 shows the results of APLM for CPS data. From Table 11, best results (bold and italic number) are obtained

for APLM in the form of less ASL compared to the other methods. Figures 1-3 shows the estimated nonparametric functions using Speckman, backfitting, and smoothing spline estimator for the variable experience of the partial linear model (PLM).

Table 11. The average squared loss (ASL) for CPS data.

N	Speckman' estimator	Backfitting Estimator	Smoothing Spline estimator	APLM estimator
534	103.3909	913.8942	104.4412	103.0753

Speckman Kernel Estimator of Partially linear Model for Experience (CPS Data)

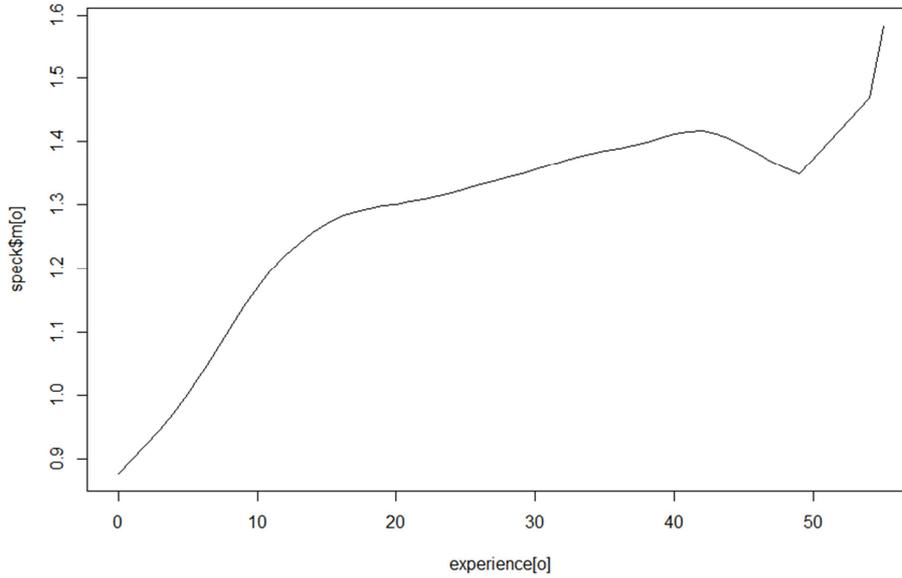


Figure 1. Speckman estimator of PLM for CPS data.

Backfitting Kernel Estimator of PLM for Experience (CPS Data)

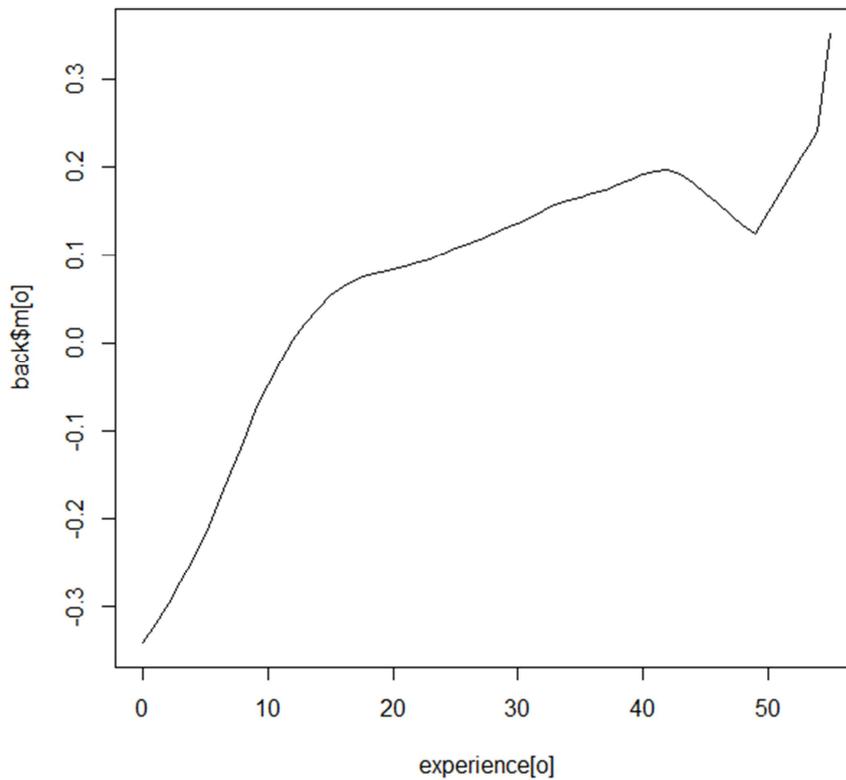


Figure 2. Backfitting estimator of PLM for CPS data.

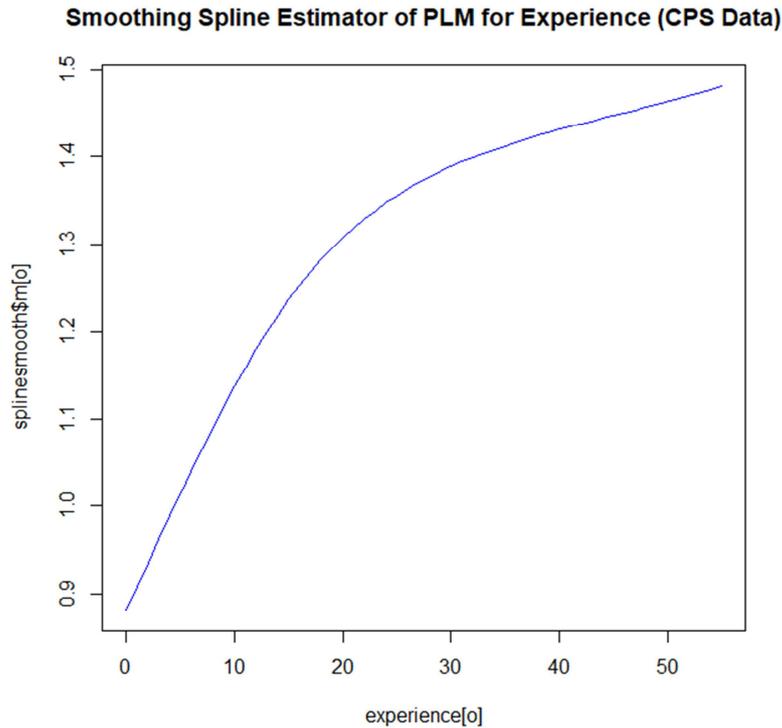


Figure 3. Smoothing Spline estimator for CPS data.

5. Conclusions

In this work, the semiparametric partial linear model (PLM) is adapted by mixing different estimation procedures defined under different conditions. Kernel smoothing, spline smoothing, and backfitting methods are different estimation procedures used to estimate the partially linear model. These methods are used in adapting the PLM by mixing. Theoretically, the adapted proposed estimator APLM is found to be a \sqrt{n} -consistent and has asymptotic normal distribution for the parametric component of the model. Also practically, best results are obtained for APLM in the form of less average squared loss (ASL) in prediction. In the simulation studies, best results are obtained when the regressors in both the parametric and nonparametric components are related (dependent) than that in the case of independency. This means that the proposed APLM is more appropriate in the case of dependent regressors.

Appendix

Proof of Theorem (1)

To prove that $\hat{\beta}_R$ is a \sqrt{n} -consistent estimator of

$$\begin{aligned} \text{plim } \hat{\beta}_R &= \beta + \text{plim} \left(\frac{1}{n} \sum_{i=1}^n X_i^* X_i^{*T} \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^n X_i^* m(z_i) \right) + \text{plim} \left(\frac{1}{n} \sum_{i=1}^n X_i^* X_i^{*T} \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^n X_i^* \varepsilon_i \right) \\ &= \beta + \hat{\Phi}_0^{-1} \text{plim} \left(\frac{1}{n} \sum_{i=1}^n X_i^* m(z_i) \right) + \hat{\Phi}_0^{-1} \text{plim} \left(\frac{1}{n} \sum_{i=1}^n X_i^* \varepsilon_i \right) \\ &= \beta \end{aligned}$$

β : Suppose that: $X_i^* = x_i - \hat{x}_i$, and $Y_i^* = y_i - \hat{y}_i$. Consider $\hat{\beta}_R$ as defined in (10):

$$\begin{aligned} \hat{\beta}_R &= \left[\frac{1}{n} \sum_{i=1}^n (X_i - \hat{X}_i)(X_i - \hat{X}_i)^T 1_i \right]^{-1} \left[\frac{1}{n} \sum_{i=1}^n (X_i - \hat{X}_i)(Y_i - \hat{Y}_i)^T 1_i \right] \\ &= \left[\frac{1}{n} \sum_{i=1}^n X_i^* X_i^{*T} 1_i \right]^{-1} \left[\frac{1}{n} \sum_{i=1}^n X_i^* (X_i^{*T} \beta + m(z_i) + \varepsilon_i) 1_i \right] \end{aligned}$$

For simplicity, the indicator 1_i will be removed and $\hat{\beta}_R$ will be written as follows:

$$\begin{aligned} \hat{\beta}_R &= \left[\frac{1}{n} \sum_{i=1}^n X_i^* X_i^{*T} \right]^{-1} \left[\frac{1}{n} \sum_{i=1}^n X_i^* X_i^{*T} \beta + \frac{1}{n} \sum_{i=1}^n X_i^* m(z_i) + \frac{1}{n} \sum_{i=1}^n X_i^* \varepsilon_i \right] \\ &= \beta + \left(\frac{1}{n} \sum_{i=1}^n X_i^* X_i^{*T} \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^n X_i^* m(z_i) \right) + \left(\frac{1}{n} \sum_{i=1}^n X_i^* X_i^{*T} \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^n X_i^* \varepsilon_i \right) \end{aligned} \tag{A1}$$

If the matrix $\left(\frac{1}{n} \sum_{i=1}^n X_i^* X_i^{*T} \right)$ has an inverse (nonsingular),

then the limit of $\hat{\beta}_R$ as $n \rightarrow \infty$ (plim) will be as:

$\therefore \hat{\beta}_R$ is a consistent estimator of β .

Also, it can be shown that $\hat{\beta}_R$ is a \sqrt{n} -consistent estimator of β as follows:

$$\begin{aligned} \therefore \text{cov}(\hat{\beta}_R) &= \text{cov} \left\{ \left[\frac{1}{n} \sum_{i=1}^n (X_i - \hat{X}_i)(X_i - \hat{X}_i)^T \mathbf{1}_i \right]^{-1} \left[\frac{1}{n} \sum_{i=1}^n (X_i - \hat{X}_i)(Y_i - \hat{Y}_i)^T \mathbf{1}_i \right] \right\} \\ &= \text{cov} \left\{ \left[\frac{1}{n} \sum_{i=1}^n X_i^* X_i^{*T} \mathbf{1}_i \right]^{-1} \left[\frac{1}{n} \sum_{i=1}^n X_i^* Y_i^* \mathbf{1}_i \right] \right\} \end{aligned}$$

For simplicity, the indicator $\mathbf{1}_i$ will be removed and $\text{cov}(\hat{\beta}_R)$ will be written as follows:

$$\begin{aligned} \text{cov}(\hat{\beta}_R) &= \text{cov} \left\{ \left[\frac{1}{n} \sum_{i=1}^n X_i^* X_i^{*T} \right]^{-1} \left[\frac{1}{n} \sum_{i=1}^n X_i^* Y_i^* \right] \right\} \\ &= \text{cov} \left\{ \left[\frac{1}{n} \sum_{i=1}^n X_i^* X_i^{*T} \right]^{-1} \left[\frac{1}{n} \sum_{i=1}^n X_i^* (m(z_i) + \varepsilon_i) \right] \right\} \quad (A2) \\ &= \left[\frac{1}{n} \sum_{i=1}^n X_i^* X_i^{*T} \right]^{-1} \left[\frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_i^2 X_i^* X_i^{*T} \right] \left[\frac{1}{n} \sum_{i=1}^n X_i^* X_i^{*T} \right]^{-1} \end{aligned}$$

Since the covariance of $\hat{\beta}_R$ is $O(1/n)$, then the convergence rate of $\hat{\beta}_R$ is $n^{-1/2}$, i.e. $\hat{\beta}_R$ is \sqrt{n} -consistent estimator of β .

To prove that $\sqrt{n}(\hat{\beta}_R - \beta) \xrightarrow{d} N(0, \hat{\Phi}_0^{-1} \hat{\Psi}_0 \hat{\Phi}_0^{-1})$:

From (A-1),

$$\begin{aligned} \hat{\beta}_R &= \left[\frac{1}{n} \sum_{i=1}^n X_i^* X_i^{*T} \right]^{-1} \left[\frac{1}{n} \sum_{i=1}^n X_i^* X_i^{*T} \beta + \frac{1}{n} \sum_{i=1}^n X_i^* m(z_i) + \frac{1}{n} \sum_{i=1}^n X_i^* \varepsilon_i \right] \\ &= \beta + \left(\frac{1}{n} \sum_{i=1}^n X_i^* X_i^{*T} \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^n X_i^* m(z_i) \right) + \left(\frac{1}{n} \sum_{i=1}^n X_i^* X_i^{*T} \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^n X_i^* \varepsilon_i \right) \\ \sqrt{n}(\hat{\beta}_R - \beta) &= \sqrt{n} \left[\frac{1}{n} \sum_{i=1}^n X_i^* X_i^{*T} \right]^{-1} \left[\frac{1}{n} \sum_{i=1}^n X_i^* m(z_i) + \frac{1}{n} \sum_{i=1}^n X_i^* \varepsilon_i \right] \\ &= \left(\frac{1}{n} \sum_{i=1}^n X_i^* X_i^{*T} \right)^{-1} \left(\frac{1}{\sqrt{n}} \left(\sum_{i=1}^n X_i^* m(z_i) + \sum_{i=1}^n X_i^* \varepsilon_i \right) \right) \end{aligned}$$

Since $\hat{\beta}_R \xrightarrow{p} \beta$, and consider (A2), then:

$$\frac{1}{n} \sum_{i=1}^n X_i^* X_i^{*T} \xrightarrow{p} E[X_i^* X_i^{*T}], \text{ and } \frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_i^2 X_i^* X_i^{*T} \xrightarrow{p} E[\hat{\varepsilon}_i^2 X_i^* X_i^{*T}].$$

The following inequality holds from the Cauchy-schwarz inequality:

$$E \left\| \hat{\varepsilon}_i^2 X_i^* X_i^{*T} \right\| \leq E \left\| X_i^* X_i^{*T} \right\|^2 \right)^{1/2} E(\hat{\varepsilon}_i^4)^{1/2}.$$

Also, from the Schwarz matrix inequality, it can be shown that:

$$E \left[\left\| X_i^* X_i^{*T} \right\|^2 \right]^{1/2} E(\hat{\varepsilon}_i^4)^{1/2} \leq E \left(\left\| X_i^* \right\|^4 \right)^{1/2} E(\hat{\varepsilon}_i^4)^{1/2},$$

and, $E \left(\left\| X_i^* \right\|^4 \right) E(\hat{\varepsilon}_i^4)^{1/2} < \infty$.

Using the central limit theorem, it is found that:

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i^* \varepsilon_i \xrightarrow{d} N(0, \Omega_0).$$

Given that: $\frac{1}{n} \sum_{i=1}^n X_i^* X_i^{*T} \xrightarrow{p} \Phi_0$, and $\frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_i^2 X_i^* X_i^{*T} \xrightarrow{p} \Psi_0$

then using Slutsky's theorem, we have:

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \Omega_0),$$

where,

$$\Omega_0 = (\hat{\Phi}_0^{-1} \hat{\Psi}_0 \hat{\Phi}_0^{-1})$$

Proof of Theorem (2)

a) To prove that $\hat{\beta}_{Speck}$ is a \sqrt{n} -consistent estimator of β :

the proof is similar to that $\hat{\beta}_R$ in Theorem (1).

b) To prove that $\sqrt{n}(\hat{\beta}_{Speck} - \beta) \xrightarrow{d} N(0, \Omega_1)$: See

Theorem (4) of Speckman (1988).

Proof of Theorem (3)

a) To prove that $\hat{\beta}_{back}$ is a \sqrt{n} -consistent estimator of β :

the proof is in Theorem 2.2 of Opsomer and Ruppert (1999).

b) To prove that $\sqrt{n}(\hat{\beta}_{back} - \beta) \xrightarrow{d} N(0, \Omega_2)$: See

Corollary 2.1 of Opsomer and Ruppert (1999).

Proof of Theorem (4)

a) To prove that $\hat{\beta}_{Spline}$ is a \sqrt{n} -consistent estimator of

β : the proof is similar to that $\hat{\beta}_R$ in Theorem (1).

b) To prove that $\sqrt{n}(\hat{\beta}_{Spline} - \beta) \xrightarrow{d} N(0, \Omega_3)$: the

proof is similar to that $\hat{\beta}_R$ in Theorem (1). (See also, Holland 2017).

Proof of Theorem (5)

a) To prove that $\hat{\beta}_{APLM}$ is a \sqrt{n} -consistent estimator of

β :

$$\hat{\beta}_{APLM} = \sum_{j=1}^3 \bar{W}_j \hat{\beta}_j = \bar{W}_1 \hat{\beta}_1 + \bar{W}_2 \hat{\beta}_2 + \bar{W}_3 \hat{\beta}_3, \quad (A3)$$

where $\hat{\beta}_1 = \hat{\beta}_{Speck}$ (Speckman estimator), $\hat{\beta}_2 = \hat{\beta}_{back}$

(backfitting estimator), and $\hat{\beta}_3 = \hat{\beta}_{Spline}$ (spline smoothing estimator). In Theorems (1:4), it is shown that the Robinson estimator ($\hat{\beta}_R$), the Speckman estimator ($\hat{\beta}_{Speck}$), backfitting estimator ($\hat{\beta}_{back}$), and the spline smoothing estimator ($\hat{\beta}_{Spline}$) are all \sqrt{n} -consistent estimators of β . Since $\hat{\beta}_{APLM}$ is a linear combination of three independent estimators $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$, then $\hat{\beta}_{APLM}$ in (A3) can be proved as a consistent estimator as follows:

$$\text{cov}(\hat{\beta}_{APLM}) = \text{cov}(\hat{\beta}_R + \hat{\beta}_{Speck} + \hat{\beta}_{Spline}) = \text{cov}(\hat{\beta}_R) + \text{cov}(\hat{\beta}_{Speck}) + \text{cov}(\hat{\beta}_{Spline})$$

since the three estimators are independent. As it is proved in theorems (1:4), that each of the three estimators is a \sqrt{n} -consistent estimator of β , then from (A2) it is found that the covariance of $\hat{\beta}_R$ is $O(1/n)$, and the convergence rate of $\hat{\beta}_R$ is $n^{-1/2}$, i.e. $\hat{\beta}_R$ is \sqrt{n} -consistent estimator of β . Also, the same is true for $\hat{\beta}_{Speck}$, $\hat{\beta}_{back}$, and $\hat{\beta}_{Spline}$, and so the $\text{cov}(\hat{\beta}_{APLM})$ is $O(1/n)$, with convergence rate of $n^{-1/2}$, and $\hat{\beta}_{APLM}$ is a \sqrt{n} -consistent estimator of β .

b) To prove that $\sqrt{n}(\hat{\beta}_{APLM} - \beta) \xrightarrow{d} N(0, \Omega_n)$:

From Theorem (2), it is shown that: $\sqrt{n}(\hat{\beta}_{Speck} - \beta) \xrightarrow{d} N(0, \Omega_1)$,

from Theorem (3), it is shown that: $\sqrt{n}(\hat{\beta}_{back} - \beta) \xrightarrow{d} N(0, \Omega_2)$, and from Theorem (4), it is

shown that: $\sqrt{n}(\hat{\beta}_{Spline} - \beta) \xrightarrow{d} N(0, \Omega_3)$,

Since $\hat{\beta}_{APLM} = \bar{W}_1 \hat{\beta}_1 + \bar{W}_2 \hat{\beta}_2 + \bar{W}_3 \hat{\beta}_3$ from (A3),

then $\sqrt{n}(\hat{\beta}_{APLM} - \beta) \xrightarrow{d} N(0, \Omega_n)$,

where $\Omega_n = \sum_{j=1}^3 \bar{W}_j^2 \Omega_j$.

References

[1] Berndt, E. R. (1991). The practice of econometrics, The classic and contemporary, Addison-Wesley Pub. Co.
 [2] Buja, A., Hastie, T. J., and Tibshirani, R. j. (1989). "Linear smoother and additive models (with discussion), *Annals of Statistics*, 17, 453-555.
 [3] Castillo, E.; Castillo, C.; and Hadi, A. S. (2009). "Combining estimates in regression and classification", *Journal of the American Statistical Association*, 91, 1641-1650.
 [4] Denby, L. (1986). "Smooth regression function", *Statistical Research Report 26*. AT&T, Bell Laboratories, Princeton,

$$\begin{aligned} \text{plim } \hat{\beta}_{APLM} &= \text{plim} \sum_{j=1}^3 \bar{W}_j \hat{\beta}_j \\ &= \bar{W}_1 \text{plim } \hat{\beta}_1 + \bar{W}_2 \text{plim } \hat{\beta}_2 + \bar{W}_3 \text{plim } \hat{\beta}_3 \\ &= \beta \sum_{j=1}^3 \bar{W}_j = \beta. \end{aligned}$$

Then $\hat{\beta}_{APLM}$ is a consistent estimator of β . Also, It can be proved that $\hat{\beta}_{APLM}$ is a \sqrt{n} -consistent estimator of β as follows.

New Jersey.

[5] Eilers, P. H. C.; and Marx, B. D. 0 (1996). "Flexible smoothing with b-splines and penalties", *Statistical Science*, 11, 89-121.
 [6] Engle, R.; Granger, C.; Rice, J.; and Weiss, A. (1986). "Nonparametric estimates of the relation between weather and electricity sales", *Journal of the American Statistical Association*, 81, 310-320.
 [7] Green, P., and Yandell, B. S. (1985). *Semi-parametric Generalized Linear Models*. Part of the *Lecture Notes in Statistics* book series (LNS, volume 32).
 [8] Green, P., Jennison, C., and Seheult, A. (1985). "Analysis of field experiments by least squares smoothing", *Journal of the Royal Statistical Society, Series B.*, 47, 299-315.
 [9] Hardle, W., Liang, H., and Gao, J. (2000). *Partially linear models*, Springer Verlag.
 [10] Hardle, W., Muller, M., Sperlich, S., and Werwatz, A. (2004). *Nonparametric and Semiparametric Modeling: An Introduction*, Springer, New York.
 [11] Hastie, T. J., and Tibshirani, R. j. (1990). *Generalized additive models*, Vol. 43 of *Monographs on statistics and applied probability*, Chapman and Hall, London.
 [12] Heckman, N. E., (1986). "Spline smoothing in a partly linear model", *Journal of the Royal Statistical Society, Series B.*, 48, 244-248.
 [13] Holland, A. (2017). "Penalized spline estimation in the partially linear model", *Journal of Multivariate Analysis*, 153, 211-235.
 [14] LeBlanc, M., and Tibshirani, R. (1996). "Combining estimates in regression and classification", *Journal of the American Statistical Association*, 91, 1641-1650.
 [15] Li, Q. (1996). "Semiparametric estimation of partially linear panel data models", *Journal of Econometrics*, Volume 71, Issues 1-2, 389-397.
 [16] Linton, O. B. (1995). "estimation in semiparametric models: A review". In: Phillips, P. C. B., Maddala, G. S. (Eds.), a Volume in Honor of C. R. Raw. Blackwell.
 [17] Liu, Q. (2010). "Asymptotic Theory for Difference-based Estimator of Partially Linear Models", *Journal of the Japan Statistical Society*, 39, 393-406.

- [18] Muller, M. (2000). Generalized Partial Linear Models, XploRe –Application Guid, 145-170.
- [19] Muller, M. (2001). “Estimation and testing in generalized partial linear models- a comparative study, Statistics and computing, 11, 299-309.
- [20] Nkurnziza, S. (2015). “On combining estimation problems under quadratic loss: A generalization”, Windsor Mathematics Statistics Report, University of Windsor, Ontario, Canada.
- [21] O’Sullivan, F. (1986). “A statistical perspective on ill-posed inverse problems (with discussion)”, Statistical Science, 1, 505–527.
- [22] Opsomer, and Ruppert, 1999. “A root-n consistent backfitting estimator for semiparametric additive modelling”.
- [23] Robinson, P. M. (1988). “Root n-consistent semiparametric regression”, Econometrica, 56, 931-954.
- [24] Ruppert, D. Wand M. P.; Carroll, R. J. (2003). Semiparametric regression, New York, Cambridge University press.
- [25] Severini, T.; and Wong, W (1992). “Generalized profile likelihood and conditional parametric models”, Annals of Statistics, 20, 1768-1802.
- [26] Silverman, B. W. (1985). “Some Aspects of the Spline Smoothing Approach to Non-Parametric Regression Curve Fitting”, Journal of the Royal Statistical Society. Series B Vol. 47, No. 1, 1-52.
- [27] Speckman, P. (1988). “Kernel smoothing in partial linear models”, Journal of the Royal Statistical Society, Series B., 50, 413-436.
- [28] Wahba, G. (1990). Spline models for observational data, Society for Industrial and applied mathematics (Siam), Philadelphia, Pennsylvania.
- [29] Wang L.; Brown, L.; and Cai, T. (2011). “A difference based approach to the semiparametric partial linear model”, Electronic Journal of Statistics, 5, 619-641.
- [30] Yang, Y. (1999). “Regression with multiple candidate models: selecting or mixing?” Technical report no. 8, Department of Statistics, Iowa state University.
- [31] Yang, Y. (2001). “Adaptive regression by mixing”, Journal of the American Statistical Association, 96, 574-588.
- [32] Yatchew, A. (1997). “An Elementary Estimator of the Partial Linear Model, Economics Letters, 57, 135-43.
- [33] Yatchew, A. (2003). Semiparametric regression for the applied econometrician, New York, Cambridge University press.
- [34] Zhou, S.; Chen, X.; Wolfe, D. A. (1998). “Local asymptotics for regression splines and confidence regions”, *Annals of Statistics*, 26, 1760-1782.