Partially Linear Generalized Single Index Models for Functional Data (PLGSIMF)

Mohamed Alahiane 1,*, Idir Ouassou 1,*, Mustapha Rachdi 2,* and Philippe Vieu 3

1 Ecole Nationale des Sciences Appliquées, Université Cadi Ayyad, Marrakech 40 001, Morocco; alahianemed@gmail.com (M.A.); i.ouassou@uca.ac.ma (I.O.)
2 Laboratoire AGEIS, UFR SHS, Université Grenoble Alpes, BP 47, CEDEX 09, 38040 Grenoble, France
3 Institut de Mathématiques de Toulouse, Université Paul Sabatier, CEDEX 9, 31062 Toulouse, France; vieu@math.univ-toulouse.fr

* Correspondence: mustapha.rachdi@univ-grenoble-alpes.fr

Abstract: Single-index models are potentially important tools for multivariate non-parametric regression analysis. They generalize linear regression models by replacing the linear combination $a_0^\top X$ with a non-parametric component $\eta_0(a_0^\top X)$, where $\eta_0(\cdot)$ is an unknown univariate link function. In this article, we generalize these models to have a functional component, replacing the generalized partially linear single index models $\eta_0(a_0^\top X) + \beta_0^\top Z$, where $a$ is a vector in $\mathbb{R}^d$, $\eta_0(\cdot)$ and $\beta_0(\cdot)$ are unknown functions that are to be estimated. We propose estimates of the unknown parameter $a_0$, the unknown functions $\beta_0(\cdot)$ and $\eta_0(\cdot)$ and establish their asymptotic distributions, and furthermore, a simulation study is carried out to evaluate the models and the effectiveness of the proposed estimation methodology.

Keywords: asymptotic normality; functional data analysis (FDA); polynomial splines; quasi-likelihood; semi-parametric regression; single-index model

1. Introduction

Generalized linear models are proposed by Nelder and Wedderburn [1], $g(\mu(X)) = \beta^\top X$; for a detailed review, we refer the readers to McCullagh and Nelder [2]; it consists of a random component and systematic component. GLMs assume the responses come from the exponential dispersion model family. They extend linear models to allow the relationship between the predictors and the function of the mean of continuous or discrete response through a canonical link function. These models encounter problems such as the canonical link function is sometimes unknown, the link between response and predictors can be complex as well as the plague of dimension reduction. To address these problems, several approaches have been developed. Hastie and Tibshirani [3] propose the GAMs models, in which the linear predictor depends linearly on smooth of predictor variables, one of the criticisms of these models is that they do not take into consideration the interactions between covariates. The manuscripts of Wood [4] and Dunn, Peter, Smyth, Gordon [5] are the latest references dealing with these two models.

The single index model had been employed to reduce the dimensionality of data, and avoid the “curse of dimensionality” while maintaining the advantages of non-parametric smoothing in multivariate regression cases over the last few decades, see for example the work of Lai et al. [6].

The single index $a^\top X$ aggregates the influence of the observed values $X = (X_1, \cdots, X_d)$ of the explanatory variables into one number.

Examples of economic index include the following: a stock index, inflation index, cost-of-living index, and price index.

Furthermore, this idea had first been extended to the functional setting by Ferratry, Vieu et al. [7] for functional regression problems, which led to the functional single index
regression model (FSIRM). The functional index acts as a filter permitting the extraction of
the part of explaining the scalar response $Y$, and plays an important role in a such model.

The predictor is not generally linear, but is complex, which prompted Carroll et al. [8]
to propose their model GPLSIM by applying the local-quasi-likelihood function and the
kernel type of smoothing by approximating the function $\eta_0$ using local linear methods
$g(\mu(X,Z)) = \eta_0(a_0^\top X) + \beta_0^\top Z$. A GPLSIM was proposed by Chin-Shang Li et al. [9], in
which the unknown smooth function of single index was approximated by a spline function
that can be expressed as a linear combination of B-spline basis function considered as
follows $g(\mu(X,Z)) = \beta^\top X + q(a^\top Z)$ using a modified Fisher-scoring method. Moreover,
Wang and Cao [10], have studied the GPLSIM model by applying the quasi-likelihood and
polynomial spline smoothing $g(\mu(X,Z)) = \eta_0(a_0^\top X) + \beta_0^\top Z$.

In recent years, the analysis of functional data has made considerable progress in
several areas, including image processing, biomedical studies, environmental sciences,
public health, etc. Several researchers have focused their efforts on studying this type of
data. We mention the work of Aneiros-Pérez and Vieu [11], and for more details, we refer
to the books of Horváth Kokoszka [12], Ferraty and Vieu [7], Aneiros-Pérez and Vieu [13],
and Ramsay and Silverman [14]. Yu, Du and Zhang [15] have proposed the SIPFLM model
combining the single-index model (SIM) and the FLM model by optimizing the sum of the
least squares using the B-Spline basis $Y = g(a^\top X) + \int_0^1 \beta(t)Z(t)dt$. Jiang Du et al. [16]
proposed the GFPLM model using the functional principal component analysis $g(\mu(X,Z)) = a^\top X + \int_0^1 \beta(t)Z(t)dt$. Rachdi, Alahiane, Ouassou and Vieu presented a book chapter on
the generalization of the GPLSIM model in the Iwfos 2020 conference, see Rachdi et al. [17].
Our objective is to combine the GPLSIM with the SIPFLM and consider the following gen-
eralized partially functional single-index model called PLGSIMF using B-Spline expansion
and the quasi-likelihood function $g(\mu(X,Z)) = \eta_0(a^\top X) + \int_0^1 \beta(t)Z(t)dt$ in order to rem-
ey the interaction effects, the dimension scourge and to take into account the functional
random variables.

The paper is organized as follows. In Sections 1 and 2, we localize our model in the
literature, and we present the Fisher-scoring update algorithm used to estimate our single-
index vector, the parametric function and the slope function. In Section 3, we investigate
an asymptotic study of the estimators presented in the paper. Numerical simulation in the
Gaussian case as in the logistic case is presented in Section 4. The proofs of the results are
developed in Section 6 and in the Appendix A for the different technical lemmas necessary
to develop our asymptotic study both for the non-parametric function, for the single-index
vector and for the slope function.

Let $H$ be a separable Hilbert space, which is endowed with the scalar product $\langle \cdot, \cdot \rangle_H$
and the norm $|| \cdot ||_H$. Let $Y$ be a scalar response variable and $(X,Z) \in \mathbb{R}^d \times H$ be the
predictor vector where $X = (X_1, \ldots, X_d)$ and $Z$ to be a functional random variable that is
valued in $H$. For a fixed $(x,z) \in \mathbb{R}^d \times H$, we assume that the conditional density function
of the response $Y$ given $(X,Z) = (x,z)$ belongs to the following canonical exponential
family:

$$f_{Y|X=x,Z=z}(y) = \exp\left(y \xi(x,z) - B(\xi(x,z)) + C(y)\right), \quad (1)$$

where $B$ and $C$ are two known functions that are defined from $\mathbb{R}$ into $\mathbb{R}$, and
$\xi : \mathbb{R}^d \times H \rightarrow \mathbb{R}$ is the parameter in the generalized parametric linear model, which is
linked to the dependent variable

$$\mu(x,z) = E\left[Y|X=x, Z=z\right] = B'(\xi(x,z)), \quad (2)$$

where $B'$ denotes the first derivative of the function $B$. In what follows, we consider the
function $g(\mu(x,z))$ as a generalized single-index partially functional linear model:

$$g(\mu(x,z)) = \eta_0(a^\top x) + \int_0^1 \beta(t)z(t)dt, \quad (3)$$
where \( \alpha = (a_1, a_2, \ldots, a_d) \in \mathbb{R}^d \) is the \( d \)-dimensional single-index coefficient vector, \( \beta \) is the coefficient function in the functional component, and \( \eta_0 \) is the unknown single-index link function which will be assumed to be sufficiently smooth.

If the conditional variance \( \text{Var}(Y | X = x, Z = z) = \sigma^2 V(\mu(x, z)) \), where \( V \) is an unknown positive function, then the estimation of the mean function \( \mu(x, z) \) may be obtained by replacing the log-likelihood \( f_{Y|X=x,Z=z} \) given by (1), by the quasi-likelihood \( Q(u, v) \) given by

\[
\frac{\partial Q(u, v)}{\partial u} = \frac{v - u}{\sigma^2 V(u)},
\]

for any real numbers \( u \) and \( v \), which may be written as \( Q(u, v) = \int_0^u \frac{v - t}{\sigma^2 V(t)} \, dt. \)

2. Estimation Methodology

Let \( (X_i, Y_i, Z_i)_{i=1}^n \) be a sequence of independent and identically distributed (i.i.d.) as \( (X, Y, Z) \) and, for each \( i = 1, \ldots, n \),

\[
g(\mu(X_i, Z_i)) = \eta_0(\alpha^\top X_i) + \int_0^1 \beta(t) Z_i(t) \, dt.
\]

(4)

We assume that the function \( \eta_0 \) is supported within the interval \([a, b]\) where \( a = \inf(\alpha^\top X) \) and \( b = \sup(\alpha^\top X) \).

We introduce a sequence of knots \( (k_m) \) in the interval \([a, b]\), with \( J \) interior knots, such that \( k_{r+1} = \cdots = k_1 = k_0 = a < k_1 < \cdots < k_J = k_{J+1} = \cdots = k_{J+r} \), where \( J = J_n \) is a sequence of integers which increases with the sample size \( n \). Now, let \( N_n = J_n + r \) be the number of knots, \( (B_i(u))_{i=1}^n \) be the B-spline functions of order \( r \), and \( h = (b - a) / (J_n + 1) \) be the distance between the neighbors knots.

Let \( S_n \) be the space of polynomial splines on \([a, b]\) of order \( r \geq 1 \). By De Boor [18], we can approximate \( \eta_0 \), assumed in \( H(p) \) (which will be defined in section 3) by a function \( \hat{\eta} \in S_n \). So, we can write \( \hat{\eta}(u) = \hat{\gamma}^\top B(u) \) where \( B(u) \) is the spline basis and \( \hat{\gamma} \in \mathbb{R}^{N_n} \) is the spline coefficient vector.

We introduce a new knots sequence \( 0 = t_0 < t_1 < \cdots < t_{k+1} = 1 \) of \([0, 1]\). Then, there exists \( N' = k + r + 1 \) functions in the B-splines basis which are normalized and of order \( r \), such that

\[
\beta(\cdot) \approx \delta^\top B_2(\cdot) \quad \text{where} \quad B_2(\cdot) = (B_{21}(\cdot), B_{22}(\cdot), \ldots, B_{2N'}(\cdot))^\top \quad \text{and} \quad \delta \in \mathbb{R}^{N'}.
\]

By setting

\[
W = \left( \int_0^1 Z(t) B_{21}(t) \, dt, \ldots, \int_0^1 Z(t) B_{2N'}(t) \, dt \right),
\]

(5)

and \( w \) and \( W_i \) are defined accordingly to (5), the mean function estimator \( \hat{\mu}(x, z) \) is then given by the evaluation of the parameter \( \theta = (\alpha^\top, \gamma^\top, \delta^\top)^\top \) and by inverting the following equation \( g(\hat{\mu}(x, z)) = \hat{\gamma}^\top B_1(\hat{\alpha}^\top x) + \hat{\delta}^\top w \). Notice that the parameter \( \theta = (\alpha^\top, \gamma^\top, \delta^\top)^\top \) is determined by maximizing the following quasi-likelihood rule

\[
\hat{\theta} = \left( \hat{\alpha}^\top, \hat{\gamma}^\top, \hat{\delta}^\top \right)^\top = \arg\max_{\theta=(\alpha, \gamma, \delta) \in \mathbb{R}^d \times \mathbb{R}^{N_n} \times \mathbb{R}^{N'}} l(\theta),
\]

where \( l(\theta) := l(\alpha, \gamma, \delta) = \frac{1}{n} \sum_{i=1}^n Q(g^{-1}(m_i), Y_i) \), with

\[
m(x, z) = \gamma^\top B_1(\alpha^\top x) + \delta^\top \omega,
\]

\[
m_i := \gamma^\top B_1(\alpha^\top X_i) + \delta^\top W_i \quad \text{and} \quad m_{0i} = \gamma^\top B_1(U_{0i}) + \delta^\top W_i,
\]
where \( U_0 = a_0^\top X_1 \) with \( a_0, \gamma_0, \delta_0, \eta_0, \beta_0 \) denoting the true values, respectively, of \( a, \gamma, \delta, \eta \) and \( \beta \).

To overcome the constraint \( \|a\| = 1 \) and \( a_1 > 0 \) of the \( d \)-dimensional index \( a \), we proceed by a re-parameterization, which is similar to Yu and Ruppert [19]

\[
\alpha(\tau) = \left( \sqrt{1 - \|\tau\|^2}, \tau \right)^\top \text{ for } \tau \in \mathbb{R}^{d-1}.
\]

The true value \( \tau_0 \) of \( \tau \), must satisfy \( \|\tau_0\| \leq 1 \). Then, we assume that \( \|\tau_0\| < 1 \). The Jacobian matrix of \( \alpha : \tau \rightarrow \alpha(\tau) \) of dimension \( d \times (d - 1) \) is \( J(\tau) \). Notice that \( \tau \) is unconstrained and is one dimension lower than \( \alpha \).

Finally, let

\[
R(\tau) = \begin{pmatrix} f(\tau) & 0 \\ 0 & I_{N' \times N'} \end{pmatrix}
\]

the Jacobian matrix of \((\alpha(\tau)^\top, \delta^\top)^\top\), which is of dimension \((d + N') \times (d + N' - 1)\). Let

\[
(\tilde{\alpha}, \tilde{\delta}) = \arg\max_{(\alpha, \delta) \in \mathbb{R}^d \times \mathbb{R}^N, \tau \in \mathbb{R}^{d-1}} \frac{1}{n} \sum_{i=1}^n Q\left( \gamma^{-1} \left\{ \tilde{\eta}(\alpha(\tau)^\top X_i) + \delta^\top W_i \right\}, Y_i \right) \text{ and } T_i = \left( X_i^\top, W_i^\top \right)^\top,
\]

Denote

\[
m_i = \gamma B_1(\alpha^\top X_i) + \delta^\top W_i, \quad T_i = \left( X_i^\top, W_i^\top \right)^\top,
\]

\[
m_{0i} = m_0(X_i, W_i) = \gamma B_1(a_0^\top X_i) + \delta_0^\top W_i = \gamma B_1(U_0) + \delta_0^\top W_i \text{ with } U_0 = a_0^\top X_i,
\]

\[
m_0(T) = \gamma B_1(a_0^\top X) + \delta_0^\top W = \gamma B_1(U_0) + \delta_0^\top W \text{ with } U_0 = a_0^\top X.
\]

and

\[
(\tilde{\tau}, \tilde{\delta}) = \arg\max_{\tau, \delta} \tilde{l}(\tau, \delta) \text{ where } \tilde{l}(\tau, \delta) = \frac{1}{n} \sum_{i=1}^n Q\left( \gamma^{-1} \left\{ \tilde{\eta}(\alpha(\tau)^\top X_i) + \delta^\top W_i \right\}, Y_i \right).
\]

Note that \( \theta_\tau = (\tau^\top, \gamma^\top, \delta^\top)^\top \) is a \((d - 1) \times N_n \times N'\)-dimensional parameter, while \( \theta \) is a \( d \times N_n \times N'\)-dimensional one. Let

\[
\rho_1(m) = \frac{1}{\sigma^2 V(\gamma^{-1}(m))} \left[ \frac{d}{dm}(\gamma^{-1}(m)) \right]^l
\]

and denote

\[
q_l(m, y) = \frac{\partial^l}{\partial m} Q(\gamma^{-1}(m), y), \text{ for } l = 1, 2.
\]

Then,

\[
q_1(m, y) = (y - \gamma^{-1}(m)) \rho_1(m) \quad \text{and} \quad q_2(m, y) = (y - \gamma^{-1}(m)) \rho_1'(m) - \rho_2(m).
\]

So, \( l(\theta_\tau) \) becomes

\[
l(\theta_\tau) = \frac{1}{n} \sum_{i=1}^n Q\left( \gamma^{-1} \left\{ \gamma B_1(\alpha(\tau)^\top X_i) + \delta^\top W_i \right\}, Y_i \right) = \frac{1}{n} \sum_{i=1}^n Q\left( \gamma^{-1} \left\{ m_i \right\}, Y_i \right)
\]

The score vector is then

\[
S(\theta_\tau) = \frac{\partial l}{\partial \theta_\tau}(\theta_\tau) = \frac{1}{n} \sum_{i=1}^n q_1(m, Y_i) \xi_i(\tau, \gamma, \delta),
\]

where
\begin{align*}
\xi_i(\tau, \gamma, \delta) &= \left( \gamma^\top B_1 \begin{pmatrix} n^\top(\tau) X_i \\ B_1(\alpha^\top(\tau) X_i) \\ W_i \end{pmatrix} \right). 
\end{align*}

The expectation of the Hessian matrix is
\begin{align*}
H(\theta_\tau) &= \mathbb{E} \left[ \frac{\partial^2}{\partial \theta_\tau \partial \theta_\tau} S(\theta_\tau) \right] \\
&= -\frac{1}{n} \sum_{i=1}^{n} \rho_2(m_i) \xi_i(\tau, \gamma, \delta) \xi_i^\top(\tau, \gamma, \delta),
\end{align*}

The Fisher Scoring update equations $\theta_\tau^{(k+1)} = \theta_\tau^{(k)} - H(\theta_\tau^{(k)})^{-1} S(\theta_\tau^{(k)})$, becomes
\begin{align*}
\theta_\tau^{(k+1)} &= \theta_\tau^{(k)} + \left[ \sum_{i=1}^{n} \rho_2(m_i^{(k)}) \xi_i(\tau^{(k)}, \gamma^{(k)}, \delta^{(k)}) \xi_i^\top(\tau^{(k)}, \gamma^{(k)}, \delta^{(k)}) \right]^{-1} \\
&\quad \times \left[ \sum_{i=1}^{n} \rho_1(m_i) \xi_i(\tau^{(k)}, \gamma^{(k)}, \delta^{(k)}) \right],
\end{align*}

where $m_i^{(k)} = \gamma^{(k)^\top} B_1 \left( \alpha^{(k)^\top}(\tau^{(k)} X_i) + \delta^{(k)^\top} W_i \right)$ for $1 \leq i \leq n$ and $m_i^{(k)} = S^{-1}(m_i^{(k)})$.

It follows that
\begin{align*}
\hat{\delta}(t) &= \hat{\delta}^\top B_2(t) = \delta^{(k)^\top} B_2(t), \\
\hat{\gamma}(t) &= \hat{\gamma}^\top B_1(t) = \gamma^{(k)^\top} B_1(t), \\
\hat{m}_i &= \gamma^{(k)^\top} B_1 \left( \alpha^{(k)^\top}(\tau^{(k)} X_i) + \delta^{(k)^\top} W_i \right) = \gamma^{(k)^\top} B_1 \left( \alpha^{(k)^\top}(\tau^{(k)} X_i) \right) + \delta^{(k)^\top} W_i,
\end{align*}

where $\hat{\delta}_i = S^{-1}(\hat{\delta}_i)$, and $\hat{\alpha} = \hat{\alpha}(\tau^{(k)})$ is the estimator of the single-index coefficient vector of the PLGSIMF model.

### 3. Some Asymptotics
We present asymptotic properties of the estimators for the non-parametric components, the functional component, the single-index coefficient vector and the slope function of the PLGSIMF model. For this aim, we will need some assumptions.

#### 3.1. Some Additional Notions and Assumptions

Let $\varphi$, $\varphi_1$ and $\varphi_2$ be measurable functions on $[a, b]$. We define the empirical inner product $\langle \varphi_1, \varphi_2 \rangle_n$ and its corresponding norm $\| \varphi \|_n$ as follows
\begin{align*}
\langle \varphi_1, \varphi_2 \rangle_n &= \frac{1}{n} \sum_{i=1}^{n} \varphi_1(U_i) \varphi_2(U_i) \quad \text{and} \quad \| \varphi \|_n^2 = \frac{1}{n} \sum_{i=1}^{n} \varphi^2(U_i) \quad \text{where} \quad U_i = \alpha^\top X_i.
\end{align*}

If $\varphi$, $\varphi_1$ and $\varphi_2$ are $L^2$-integrable, we define the theoretical inner product and its corresponding norm as follows
\begin{align*}
\langle \varphi_1, \varphi_2 \rangle &= \mathbb{E} \left[ \varphi_1(U) \varphi_2(U) \right] \quad \text{and} \quad \| \varphi \|_2^2 = \mathbb{E} \left[ \varphi^2(U) \right] = \int_a^b \varphi^2(u) f(u) du.
\end{align*}

Let $v \in \mathbb{N}^*$ and $\varepsilon \in (0, 1]$ such that $p = v + \varepsilon > 1.5$. We denote by $\mathcal{H}(p)$ the collection of functions $g$, which are defined on $[a, b]$ whose $v$-th order derivative, $g^{(v)}$, exists and satisfies the following $\varepsilon$-th order Lipschitz condition
\begin{align*}
\left| g^{(v)}(m') - g^{(v)}(m) \right| \leq C|m' - m|^\varepsilon, \quad \text{for all} \quad a \leq m, m' \leq b.
\end{align*}
Let $\varepsilon = Y - g^{-1}(m_0(T))$ where $T = (X^T, W^T)^T$.

(C1) The single-index link function $\eta_0 \in \mathcal{H}(p)$, where $\mathcal{H}(p)$ is defined as above.

(C2) For all $m \in \mathbb{R}$ and for all $y$ in the range of the response variable $Y$, the function $q_2(m, y)$ is strictly negative, and for $k = 1, 2$, there exist some positive constants $c_\eta$ and $C_\eta$ such that $c_\eta < \left| q_2^k(m, y) \right| < C_\eta$.

(C3) The marginal density function of $\alpha^\top X$ is continuous and bounded away from zero and is infinite on its support $[a, b]$. The $v$-th order partial derivatives of the joint density function of $X$ satisfy the Lipschitz condition of order $\alpha$ ($\alpha \in (0, 1])$.

(C4) For any vector $\tau$, there exist positive constants $c_\tau$ and $C_\tau$, such that

$$c_\tau I_{t \times t} \leq \mathbb{E} \left[ \left( \frac{1}{T} \right)^\top \left( \frac{1}{T} \right)^T \alpha^\top (\tau) X = \alpha^\top (\tau) X \right] \leq C_\tau I_{t \times t},$$

where $t = 1 + N_n + N'$ and $T = (X^T, W^T)^T$.

(C5) The number of knots $N_n$ satisfy $n^{-\frac{1}{p}} \ll N_n \ll n^\frac{1}{p}$, for $p > 3$.

(C6) The fourth order moment of the random variable $Z$ is finite, i.e., $\mathbb{E} \| Z(.) \|^4 \leq C$, where $C$ denotes a generic positive constant.

(C7) The covariance function $K(t, s) = \text{Cov}(Z(t), Z(s))$ is positive definite.

(C8) The slope function $\beta$ is a $r$-th order continuously differentiable function, i.e., $\beta \in C^r([0, 1])$.

(C9) For some finite positive constants $C_\rho, C^*_\rho$ and $M_0$

$$|\rho_1(m_0)| \leq C_\rho \text{ and } |\rho_1(m) - \rho_1(m_0)| \leq C^*_\rho |m - m_0| \text{ for all } |m - m_0| \leq M_0.$$

(C10) For some finite positive constants $C_\delta^\gamma, C^*_\delta^\gamma$ and $M_1$, the link function $g$, in the model (3), satisfies:

$$\left| \frac{d}{dm} g(m) \right|_{m = m_0} \leq C_\delta^\gamma \text{ and, for all } |m - m_0| \leq M_1,$$

$$\left| \frac{d}{dm} g^{-1}(m) - \frac{d}{dm} g^{-1}(m) \right|_{m = m_0} \leq C^*_\delta^\gamma |m - m_0|.$$

(C11) It exists a positive constant $C_0$, such that $\mathbb{E}(\varepsilon^2 | U_{r,0} \leq C_0$, where $\varepsilon = Y - g^{-1}(m_0(T))$.

3.2. Estimators Consistencies

Next we formulate several assertions on the considered estimators.

3.3. Estimation of the Nonparametric Component

The following theorem states the convergence, with rates, of the estimator $\hat{\eta}$.

**Theorem 1.** Under assumptions (C6) - (C8), we have

$$\| \hat{\eta} - \eta_0 \|_2 = O_p \left\{ \sqrt{N_n \left( \frac{1}{\sqrt{nh} + h^p} \right)} \right\} \text{ and } \| \hat{\eta} - \eta_0 \|_n = O_p \left\{ \sqrt{N_n \left( \frac{1}{\sqrt{nh} + h^p} \right)} \right\},$$

where $O_p$ denotes a “grand O of Landau” in probability.

**Proof of Theorem 1.** The proof of the previous theorem is given in the Appendix A. □

3.4. Estimation of the Slope Function

**Theorem 2.** Under assumptions (C1) - (C8), and $k \sim n^{1/(2r+1)}$, we have
\[ \| \hat{\beta}(\cdot) - \beta_0(\cdot) \|^2 = \mathcal{O}_P \left( N_2^2 \left( h^p + \frac{1}{\sqrt{nh}} \right)^2 \right) + \mathcal{O}_P (n^{-2r/(2r+1)}). \]

**Proof of Theorem 2.** The proof of the previous theorem is given in the Appendix A. \hfill \Box

### 3.5. Estimation of the Parametric Components

The next theorem shows that the maximum quasi-likelihood estimator is root-\(n\) consistent and is asymptotically normal, although the convergence rate of the non-parametric component \(\hat{\eta}\) is slower than root-\(n\). Before enouncing the theorem, let us denote

\[
\begin{align*}
Y(u_{\tau,0}) &= \mathbb{E} \left[ X \rho_2(m_0(T)) | U_{\tau,0} = u_{\tau,0} \right] \quad \text{and} \quad \Gamma(u_{\tau,0}) = \frac{\mathbb{E} \left[ W \rho_2(m_0(T)) | U_{\tau,0} = u_{\tau,0} \right]}{\mathbb{E} \left[ \rho_2(m_0(T)) | U_{\tau,0} = u_{\tau,0} \right]}, \\
\Phi(x) &= \Phi(U_{\tau,0}, x) = x - Y(u_{\tau,0}) \quad \text{and} \quad \Psi(w) = \Psi(U_{\tau,0}, w) = w - \Gamma(u_{\tau,0}).
\end{align*}
\]

**Theorem 3.** Under assumptions (C_1) – (C_11), the constrained quasi-likelihood estimators \(\hat{\alpha}\) and \(\hat{\delta}\) with \(\|\hat{\alpha}\|_d = 1\) are jointly asymptotically normally distributed, i.e.,

\[
\sqrt{n} \left( \begin{array}{c} \hat{\alpha} - \alpha_0 \\ \hat{\delta} - \delta_0 \end{array} \right) \overset{D}{\rightarrow} \mathcal{N} \left( 0, R(\tau_0) D^{-1} R^\top(\tau_0) \right),
\]

where \(\overset{D}{\rightarrow}\) denotes the convergence in distribution, and

\[
D = \mathbb{E} \left[ \rho_2(m_0(T)) \left( \begin{array}{c} \eta_0(U_{\tau,0}) \Gamma(\tau_0) \Phi(X) \\ \Psi(W) \end{array} \right) \right] \left( \begin{array}{c} \eta_0(U_{\tau,0}) \Gamma(\tau_0) \Phi(X) \\ \Psi(W) \end{array} \right)^\top,
\]

and

\[
R(\tau) = \left( \begin{array}{cc} f(\tau) & 0 \\ 0 & I_{N' \times N'} \end{array} \right).
\]

**Proof of Theorem 3.** The proof of the previous theorem is given in the Appendix A. \hfill \Box

### Comments on the Assumptions

The smoothness condition in (C1) describes that the single-index function \(\eta_0(\cdot)\) can be approximated by functions in the B-spline space with a normalized basis. On the other hand, the condition (C2) ensures the uniqueness of the solution, where the condition (C3) is a smoothness condition on the joint and marginal density functions of \(\alpha^\top X\) and \(X\). The condition (C5) allows to obtain the rate of growth of the dimension of the spline spaces relative to the sample size. Conditions (C6) and (C7) are required for covariates function \(Z\), and (C8) is a smoothness condition for slope function. Conditions, (C4) and (C9)–(C11) are technical lemmas that will be used to prove the cited theorems in this article.

Then, in this paper, we introduce a new generalized functional partially linear single-index model based on a combination of polynomial smoothing. The asymptotic properties of the resulting estimators under certain regularity assumptions are established for this model and hence the non-parametric component \(\eta\) and the slope function \(\beta\) are evaluated by the B-spline functions. Finally, we give some simulations to illustrate our results.

### 4. A Numerical Study

We conduct a simulation study in order to show our results’ effectiveness. We will treat two main cases of link functions: the identity and the logit link functions. Recall that if the density of \(Y\) given \(X = x\) and \(Z = z\) is

\[
f_{Y|X=x,Z=z}(y) = \exp \left( y \xi(x,z) - B(\xi(x,z)) + C(y) \right),
\]
then the link function
\[ g(\mu(x,z)) = \mathbb{E}[Y|X = x, Z = z] = B'(\xi(x,z)) \text{ and } V(\mu(x,z)) = \frac{B''(\xi(x,z))}{c^2}. \]

4.1. Case 1: Identity Link Function

We consider the case where the link function is the identity and the model
\[ Y_i = \sin\left\{ \pi \left( \frac{\alpha^\top X_i - A}{B - A} \right) \right\} + \int_0^1 \beta(t) Z_i(t) dt + \epsilon_i \text{ for } i = 1, \ldots, n. \] (6)

The responses \( Y_i \) are simulated according to the Equation (6), \( X_i \) are taken uniformly over the interval \([-0.5, 0.5]\), whereas the errors are normally distributed with mean 0 and variance 0.01, \( \epsilon_i \sim \mathcal{N}(0, 0.01) \). Moreover, we take the following coefficients
\[ \alpha = \frac{1}{\sqrt{3}} (1, 1, 1)^\top, \quad A = \frac{\sqrt{3}}{2} - \frac{1.645}{\sqrt{12}} \text{ and } B = \frac{\sqrt{3}}{2} + \frac{1.645}{\sqrt{12}}. \]

The function \( \beta(\cdot) \) and \( Z_i(\cdot) \) are given by
\[ \beta(t) = \sqrt{2} \sin\left( \frac{\pi t}{2} \right) + 3\sqrt{2} \sin\left( \frac{3\pi t}{2} \right) \text{ and } Z_i(t) = \sum_{j=1}^{50} \xi_j v_j(t) \]

where \( v_j(t) = \sqrt{2} \sin((j - 0.5) \pi t), \xi_j = \mathcal{N}(0, \lambda_j) \) and \( \lambda_j = ((j - 0.5) \pi)^{-2} \).

The knots are selected according to the formula \( Cn^{\frac{1}{2}} \log(n) \) where \( C \in [0.3, 1] \) (Like in Wang and Cao [10]). We chose \( C = 0.6 \) and we made 300 replications with samples of sizes \( n = 500 \) and \( n = 1000 \).

Computations of the bias, the standard deviation (SD) and the Mean Squared Error (MSE) with respect to (i) the parameter \( \tau \), (ii) the parameter \( \gamma \) and (iii) the parameter \( \delta \) are summarized, for \( n = 500 \) (respectively, \( n = 1000 \)), in the following Tables 1–3 (respectively, in the Tables 4–6).

### Table 1. Bias, SD and MSE according to the parameter \( \tau \) for PLGSIMF with the identity link function and \( n = 500 \).

| \( \tau \)  | Bias   | SD     | MSE          |
|----------|--------|--------|--------------|
| \( \tau_1 \) | 0.004  | 0.0031 | 2.6565 \times 10^{-5} |
| \( \tau_2 \) | -0.0009 | 0.0025 | 7.1819 \times 10^{-6} |

### Table 2. Bias, SD and MSE evolutions with respect to the parameter \( \gamma \) variation for PLGSIMF with the identity link function and \( n = 500 \).

| \( \gamma \) | Bias   | SD     | MSE          |
|---------------|--------|--------|--------------|
| \( \gamma_1 \) | -0.0256 | 0.1834 | 0.0343       |
| \( \gamma_2 \) | -0.0575 | 0.2430 | 0.0624       |
| \( \gamma_3 \) | 0.0526  | 0.3037 | 0.0950       |
| \( \gamma_4 \) | 0.1230  | 0.3621 | 0.1463       |
| \( \gamma_5 \) | 0.0471  | 0.2049 | 0.0442       |
| \( \gamma_6 \) | -0.0015 | 0.3756 | 0.1411       |
| \( \gamma_7 \) | -0.0148 | 0.3101 | 0.0964       |
| \( \gamma_8 \) | 0.0241  | 0.2281 | 0.0526       |
| \( \gamma_9 \) | 0.0005  | 0.1710 | 0.0292       |
Table 3. Bias, SD and MSE evolutions with respect to the parameter $\delta$ variation for PLGSIMF with the identity link function and $n = 500$.

| $\delta$   | Bias    | SD       | MSE       |
|------------|---------|----------|-----------|
| $\delta_1$ | -0.3444 | 0.9457   | 1.0131    |
| $\delta_2$ | -0.5969 | 0.5569   | 0.6664    |
| $\delta_3$ | -0.2527 | 0.2565   | 0.1296    |
| $\delta_4$ | 0.0696  | 0.0916   | 0.0132    |
| $\delta_5$ | 0.1093  | 0.1059   | 0.0231    |
| $\delta_6$ | 0.0746  | 0.1150   | 0.0188    |
| $\delta_7$ | 0.2909  | 0.1669   | 0.1125    |

Table 4. Bias, SD and MSE according to the parameter $\tau$ for PLGSIMF with the identity link function and $n = 1000$.

| $\tau$   | Bias    | SD       | MSE       |
|----------|---------|----------|-----------|
| $\tau_1$ | 0.0054  | 0.0023   | $3.5152 \times 10^{-5}$ |
| $\tau_2$ | 0.0021  | 0.0021   | $9.5050 \times 10^{-6}$ |

Table 5. Bias, SD and MSE evolutions with respect to the parameter $\gamma$ variation for PLGSIMF with the identity link function and $n = 1000$.

| $\gamma$ | Bias    | SD       | MSE       |
|----------|---------|----------|-----------|
| $\gamma_1$ | -0.0969 | 0.1583   | 0.0344    |
| $\gamma_2$ | -0.2220 | 0.2194   | 0.0974    |
| $\gamma_3$ | -0.1260 | 0.2478   | 0.0773    |
| $\gamma_4$ | -0.0815 | 0.3041   | 0.0991    |
| $\gamma_5$ | -0.0806 | 0.2098   | 0.0505    |
| $\gamma_6$ | -0.0135 | 0.3092   | 0.0958    |
| $\gamma_7$ | 0.2509  | 0.2509   | 0.1259    |
| $\gamma_8$ | 0.2234  | 0.1636   | 0.0998    |
| $\gamma_9$ | 0.1636  | 0.0535   |           |

Table 6. Bias, SD and MSE evolutions with respect to the parameter $\delta$ variation for PLGSIMF with the identity link function and $n = 1000$.

| $\delta$ | Bias    | SD       | MSE       |
|----------|---------|----------|-----------|
| $\delta_1$ | -0.2189 | 1.1558   | 1.3839    |
| $\delta_2$ | 0.0665  | 0.3477   | 0.1253    |
| $\delta_3$ | -0.0113 | 0.1447   | 0.0210    |
| $\delta_4$ | -0.0025 | 0.0644   | 0.0041    |
| $\delta_5$ | -0.0171 | 0.0712   | 0.0053    |
| $\delta_6$ | -0.0006 | 0.0716   | 0.0051    |
| $\delta_7$ | -0.0231 | 0.1296   | 0.0173    |

4.2. Case 2: Logit Link Function

By taking a logit link function, data are generated from the model

$$\text{logit}\{P[Y_i = 1|X_i, Z_i]\} = \sin\left\{\frac{\pi (a^\top X_i - A)}{B - A}\right\} + \int_0^1 \beta(t)Z_i(t)dt + \varepsilon_i, \quad i = 1, \ldots, n.$$

for which we have kept the same parameters and the variables as for the identity link function. Then, similarly to the identity link function case, computations of the bias, SD and the MSE with respect to the parameters $\tau$, $\gamma$ and then $\delta$ are summarized, for $n = 500$ (respectively, $n = 1000$), in the Tables 7–9 (respectively, in the Tables 10–12).

Table 7. Bias, SD and MSE evolutions with respect to the parameter $\tau$ variation for PLGSIMF with the logit link function and $n = 500$.

| $\tau$ | Bias    | SD       | MSE       |
|--------|---------|----------|-----------|
| $\tau_1$ | -0.0094 | 0.0102   | 0.0002    |
| $\tau_2$ | 0.0059  | 0.0107   | 0.0001    |
It is obviously seen that the quality of the estimators are illustrated via simulations. The method performs quite well. The Bias, SD and MSE are reasonably small in general. The parametric and nonparametric components, the single-index and also the slope function are computed by the procedure given in this paper.

Both tables correspondingly indicate the consistency of $\hat{\alpha}$ and $\hat{\delta}$ as the bias, SD and MSE decrease as the sample size increasing. The knots selection with formula $Cn^{1/2}\log(n)$ by using $C \in [0.3, 1]$ like in Li Wang and Guanqun CAO [10], we have chosen $C = 0.6$. We developed our algorithm in both cases: the identity link function and the logistic link function. Simulations show that the PLGSIMF algorithm works well in both cases.

| Table 8. Bias, SD and MSE evolutions with respect to the parameter $\gamma$ variation for PLGSIMF with the logit link function and $n = 500$. |
|----------------------------------|
| $\gamma$ | $\gamma$ | $\gamma$ | $\gamma$ | $\gamma$ | $\gamma$ | $\gamma$ | $\gamma$ | $\gamma$ |
| Bias     | 0.6075   | -0.1033  | 0.2749   | 0.1901   | 0.1506   | 0.1697   | 0.6662   | 0.5948   | 0.4662   |
| SD       | 0.2720   | 0.1529   | 0.1832   | 0.1055   | 0.0850   | 0.1266   | 0.4510   | 0.3224   | 0.4455   |
| MSE      | 0.4431   | 0.0340   | 0.1092   | 0.0473   | 0.0299   | 0.0448   | 0.6472   | 0.4577   | 0.4159   |

| Table 9. Bias, SD and MSE evolutions with respect to the parameter $\delta$ variation for PLGSIMF with the logit link function and $n = 500$. |
|----------------------------------|
| $\delta$ | $\delta$ | $\delta$ | $\delta$ | $\delta$ | $\delta$ | $\delta$ | $\delta$ |
| Bias     | -0.2977  | 0.7625   | -0.3420  | -0.2347  | -0.0243  | 0.1246   | 0.0325   |
| SD       | 0.9025   | 0.2310   | 0.5342   | 0.2008   | 0.2160   | 0.2149   | 0.3286   |
| MSE      | 0.9031   | 0.6348   | 0.4024   | 0.0954   | 0.0472   | 0.0617   | 0.1090   |

| Table 10. Bias, SD and MSE evolutions with respect to the parameter $\tau$ variation for PLGSIMF with the logit link function and $n = 1000$. |
|----------------------------------|
| $\tau$ | $\tau$ |
| Bias | 0.0054 | 0.0021 |
| SD | 0.0023 | 0.0021 |
| MSE | $3.5152 \times 10^{-5}$ | $9.5050 \times 10^{-6}$ |

| Table 11. Bias, SD and MSE evolutions with respect to the parameter $\gamma$ variation for PLGSIMF with the logit link function and $n = 1000$. |
|----------------------------------|
| $\gamma$ | $\gamma$ | $\gamma$ | $\gamma$ | $\gamma$ | $\gamma$ | $\gamma$ | $\gamma$ | $\gamma$ |
| Bias     | -0.0969  | -0.2220  | -0.1260  | -0.0815  | -0.0806  | -0.0135  | 0.2509   | 0.2234   | 0.1636   |
| SD       | 0.1583   | 0.2194   | 0.2478   | 0.3041   | 0.2098   | 0.3092   | 0.2509   | 0.2234   | 0.1636   |
| MSE      | 0.0344   | 0.0974   | 0.0773   | 0.0991   | 0.0505   | 0.0958   | 0.1259   | 0.0998   | 0.0535   |

| Table 12. Bias, SD and MSE evolutions with respect to the parameter $\delta$ variation for PLGSIMF with the logit link function and $n = 1000$. |
|----------------------------------|
| $\delta$ | $\delta$ | $\delta$ | $\delta$ | $\delta$ | $\delta$ | $\delta$ |
| Bias     | -0.2189  | 0.0665   | -0.0113  | -0.0025  | -0.0171  | -0.0006  | -0.0231  |
| SD       | 1.1558   | 0.3477   | 0.1447   | 0.0644   | 0.0712   | 0.0716   | 0.1296   |
| MSE      | 1.3839   | 0.1253   | 0.0210   | 0.0041   | 0.0053   | 0.0051   | 0.0173   |
In the figure below (Figure 1), we illustrate 500 realizations of the functional random variable $Z$.

![Figure 1](image1.png)

**Figure 1.** A sample of curves $\{Z_i(t), \ t \in [0, 1]\}_{i=1,...,500}$.

In the following figure (Figure 2), we observe the almost linearity of the single-index $u = \alpha^\top(\tau)X$ and its estimate $\hat{u} = \hat{\alpha}^\top(\hat{\tau})X$.

![Figure 2](image2.png)

**Figure 2.** Single index $u = \alpha^\top X$ versus predicted single index $\hat{u} = \hat{\alpha}^\top X$.

In the figure below (Figure 3), we plot the slope function $\beta(.)$ and its estimator $\hat{\beta}(.)$
Our model approximates well the slope function $\beta(\cdot)$.

The following figure (Figure 4) shows us the comparison between the non-parametric function $\eta(\cdot)$ and its estimator $\hat{\eta}(\cdot)$.

We consider that our model approximated to the best the non-parametric function $\eta(\cdot)$. 
To study the performance of our estimation for non-parametric function $\eta(.)$ and slope function, respectively, we will use the square root of average square errors criterion (RASE, see Peng et al. [20]):

$$RASE_1 = \left( \frac{1}{n} \sum_{i=1}^{n} (\hat{\eta}(u_i) - \eta(u_i))^2 \right)^{1/2}$$

$$RASE_2 = \left( \frac{1}{n} \sum_{i=1}^{n} (\hat{\beta}(t_i) - \beta(t_i))^2 \right)^{1/2}$$

The following tables (Tables 13 and 14) summarize the sample means, medians and variances of the $RASE_i$ ($i = 1, 2$) with different sample sizes in the Gaussian case.

For the case $n = 500$, we get

| Gaussian Cases | Mean  | Median | Var  |
|----------------|-------|--------|------|
| RASE$_1$       | 0.039 | 0.038  | 0.003|
| RASE$_2$       | 0.123 | 0.122  | 0.020|

For the case $n = 1000$, we get

| Gaussian Cases | Mean  | Median | Var  |
|----------------|-------|--------|------|
| RASE$_1$       | 0.016 | 0.016  | 0.001|
| RASE$_2$       | 0.027 | 0.125  | 0.006|

The following tables (Tables 15 and 16) summarize the sample means, medians and variances of the $RASE_i$ ($i = 1, 2$) with different sample sizes in the Logistic case.

For the case $n = 500$, we get

| Logistic Cases | Mean  | Median | Var  |
|----------------|-------|--------|------|
| RASE$_1$       | 0.045 | 0.044  | 0.023|
| RASE$_2$       | 0.133 | 0.103  | 0.0120|

For the case where $n = 1000$, we get

| Logistic Cases | Mean  | Median | Var  |
|----------------|-------|--------|------|
| RASE$_1$       | 0.028 | 0.026  | 0.014|
| RASE$_2$       | 0.124 | 0.121  | 0.002|

We conclude that as the sample size $n$ increases from 500 to 1000, the sample mean, median and variance of $RASE_i$ ($i = 1, 2$) decrease.
5. Application to Tecator Data

In this paragraph, we will apply the PLGSIMF model for Tecator data, popularly known in the functional data analysis. This data can be downloaded from the following link http://lib.stat.cmu.edu/datasets/tecator (accessed on 1 August 2021). For more details, see Ferraty and Vieu [7].

Given 215 finely chopped pieces of meat, Tecator’s data contain their corresponding fat contents \( Y_i, i = 1, \ldots, 215 \), near-infrared absorbance spectra \( Z_j, i = 1, \ldots, 215 \) observed on 100 equally wavelengths in the range 850–1050 nm, the protein content \( X_{1,i} \) and the moisture content \( X_{2,i} \). We are trying to predict the fat content of the finely chopped meat samples.

The following figure (Figure 5) shows the absorbance curves.

![Figure 5. Sample of 100 absorbance curves Z.](image)

We divide the sample randomly into two sub-samples: the training \( I_1 \) of size 160 and the test \( I_2 \) of size 55. The training sample is used to estimate the parameters, and the test sample is employed to verify the quality of predictions. To perform our model, we use the mean square error of prediction (MSEP) like in Aneiros-Pérez and Vieu [11] defined as follows:

\[
MSEP = \frac{1}{50} \sum_{i \in I_2} (Y_i - \hat{Y}_i)^2 / \text{var}_{I_2}(Y_i),
\]

where \( \hat{Y}_i \) is the predicted value based on the training sample and \( \text{var}_{I_2} \) is variance of response variables’ test sample.

The following table (Table 17) shows the performance of our PLGSIMF model by comparing it with other models. We can conclude that PLGSIMF is competitive one for such data.
Table 17. The MSEPs for different models.

| Functional Models | MSEP |
|-------------------|------|
| Model 1 (PLGSIMF) | $g(\mu(X_i, Z_i)) = \eta(a_1 X_{1i} + a_2 X_{2i}) + \int_{850}^{1050} \beta(t) Z_i(t) dt$ | 0.015 |
| Model 2 (GFPLM)   | $g(\mu(X_i, Z_i)) = \alpha_1 X_{1i} + \alpha_2 X_{2i} + \int_{850}^{1050} \beta(t) Z_i(t) dt$ | 0.024 |
| Model 3 (GFPLM)   | $g(\mu(X_i, Z_i)) = \eta(X_{1i}) + \int_{850}^{1050} \beta(t) Z_i(t) dt$ | 0.029 |
| Model 4 (GFPLM)   | $g(\mu(X_i, Z_i)) = \eta(X_{2i}) + \int_{850}^{1050} \beta(t) Z_i(t) dt$ | 0.046 |
| Model 5 (GFLM)    | $g(\mu(X_i, Z_i)) = \int_{850}^{1050} \beta(t) Z_i(t) dt$ | 0.051 |
| Model 6 (SIM)     | $Y_i = \eta(a_1 X_{1i} + a_2 X_{2i}) + \epsilon_i$ | 1.112 |

The following figure (Figure 6) shows us the estimator of the non-parametric function $\hat{\eta}(.)$.

![The estimated nonparametric function: $\hat{\eta}(.)$](image)

**Figure 6.** Estimated nonparametric function $\hat{\eta}(.)$.

The following figure (Figure 7) shows us the estimator of the slope function $\hat{\beta}(.)$. 

6. Proofs

In what follows, when no confusion is possible, we will denote by C a generic positive constant.

The following lemmas, [21–23] will be used to prove Theorem 1. The proof of these lemmas will be developed in the Appendix A.

Lemma 1. Under assumptions (C1)–(C4) and (C6)–(C8), we have

\[ \sqrt{n} \left( \frac{\hat{\tau} - \tau_0}{\delta - \delta_0} \right) \xrightarrow{D} \mathcal{N}\left(0, A^{-1}\Sigma_1 A^{-1}\right), \]

where \( \Sigma_1 \) and \( A \) will be defined below and in the appendix for more details. \( \xrightarrow{D} \) denotes the convergence in distribution.

\[ A = \begin{pmatrix} A_{11} & A_{12} \\ A_{12} & A_{22} \end{pmatrix}. \]

with

\[ A_{11} = \mathbb{E} \left[ \rho_2(m_0(T)) \{ \eta'_0(U_{\tau,0}) \}^2 J^\top(\tau_0) X W W^\top J(\tau_0) \right] \]

\[ A_{22} = \mathbb{E} \left[ \rho_2(m_0(T)) W W^\top \right] \]

\[ A_{12} = \mathbb{E} \left[ \rho_2(m_0(T)) \eta_0(U_{\tau,0}) J^\top(\tau_0) X W W^\top \right] \]

\[ \Sigma_1 = \mathbb{E} \left[ q_1^2(m_0(T), Y) \left( \begin{array}{c} \eta'_0(U_{\tau,0}) J^\top(\tau_0) X \\ \eta_0(U_{\tau,0}) X \end{array} \right)^\top \right] \]

By applying the \( \delta \)-method, we get the following lemma.
Lemma 2. Under the conditions of Lemma 1, we obtain
\[
\sqrt{n} \begin{pmatrix} a(\tilde{\tau}) - a(\tau_0) \\ \delta - \delta_0 \end{pmatrix} \xrightarrow{D} N\left(0, R(\tau_0)A^{-1}\Sigma_1 A^{-1}R^T(\tau_0)\right),
\]
where
\[
R(\tau) = \begin{pmatrix} J(\tau) & 0 \\ 0 & I_{N' \times N'} \end{pmatrix}.
\]
Furthermore \(a(\tilde{\tau}) - a(\tau_0) = o_p\left(\frac{1}{\sqrt{n}}\right)\) and \(\tilde{\delta} - \delta_0 = o_p\left(\frac{1}{\sqrt{n}}\right)\).

Lemma 3. Under the conditions of Lemma 1, we obtain
\[
\|\hat{\theta} - \tilde{\theta}\| = O_p\left(\sqrt{N_n} \left(h^p + \frac{1}{\sqrt{nh}}\right)\right),
\]
where \(N_n\) is number of B-splines basis functions of order \(r\).

Then, we can enunciate the following theorem.

Theorem 4. Under assumptions (C1)–(C5) and (C6)–(C8), we obtain
\[
\|\tilde{\eta} - \eta_0\|_2 = O_p\left(\sqrt{N_n} \left(h^p + \frac{1}{\sqrt{nh}}\right)\right),
\]
and
\[
\|\tilde{\eta} - \eta_0\|_n = O_p\left(\sqrt{N_n} \left(h^p + \frac{1}{\sqrt{nh}}\right)\right).
\]

Theorem 5. Under assumptions (C1)–(C8), and \(k \sim n^{1/(2r+1)}\), we obtain
\[
\|\tilde{\beta}(\cdot) - \beta_0(\cdot)\|^2 = O_p\left(N_n^2 \left(h^p + \frac{1}{\sqrt{nh}}\right)^2\right) + O_p\left(n^{-2r/(2r+1)}\right).
\]

Theorem 6. Under assumptions (C1)–(C11), the constrained quasi-likelihood estimators \(\hat{\alpha}\) and \(\hat{\delta}\) with \(\|\hat{\alpha}\|_d = 1\) is asymptotically normally distributed, i.e.,
\[
\sqrt{n} \begin{pmatrix} \hat{\alpha} - a_0 \\ \hat{\delta} - \delta_0 \end{pmatrix} \xrightarrow{D} N\left(0, R(\tau_0)D^{-1}R^T(\tau_0)\right)
\]
where
\[
D = E\left[\rho_2(m_0(T)) \left(\eta'_0(U_{\tau_0})f^T(\tau_0)\Phi(X)\right) \left(\eta'_0(U_{\tau_0})f^T(\tau_0)\Phi(X)\right)^T\right].
\]

Notice that the proof of this theorem is very long. So, in order to save space and not to make this paper more difficult to read, we opted for adding a Supplementary Materials, containing necessary details.

**Supplementary Materials:** The following are available online at https://www.mdpi.com/2571-905X/4/4/47/s1.

**Author Contributions:** Methodology, M.A., I.O., M.R., P.V.; Software, M.A., I.O.; Visualization, M.A., I.O.; Writing, M.A., I.O., M.R., P.V.; original draft, M.A., I.O.; Writing—review & editing, M.A., I.O., M.R., M.R., P.V. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.
Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

In what follows, we will present results and technical lemmas that would be used for the proof of the previous theorems.

First of all, for all probability measures $Q$, we denote by $L^2(Q)$ the space of squared integrable functions, i.e., $L^2(Q) = \left\{ f \text{ function such that } Qf^2 = \int f^2 dQ < \infty \right\}$. Then, let $F$ be a subclass of $L^2(Q)$. So, for all $f \in F$, we denote by $\|f\| = \left( \int f^2 dQ \right)^{\frac{1}{2}}$ the norm of $f$ with respect to $Q$.

We give the following definition that will be necessary to understand the results’ proofs.

Definition A1.

- The $\delta$-covering number, $N(\delta, F, L^2(Q))$, of $F$ is the smallest value $N$ for which it exists functions $f_1, f_2, \ldots, f_N$, such that for each $f \in F$, it exists $j \in \{1, \ldots, N\}$ such that $\|f - f_j\| < \delta$ or that $F \subset \bigcup_{j=1}^{N} B(f_j, \delta)$.

Notice that $f_j$s are not necessarily in $F$.

- For two functions $l$ and $u$, a bracketing $[l, u]$ is the set of functions $f$ such that $l \leq f \leq u$, i.e., $[l, u] = \{ f : l \leq f \leq u \}$.

- The $\delta$-covering number with bracketing $N_1(\delta, F, L^2(Q))$ is defined as the smallest value of $N$, necessary to cover the whole $F$, for which it exists pairs of functions $\left\{ [f_l^j, f_u^j] ; j = 1, \ldots, N \right\}$ with $\|f_u^j - f_l^j\| \leq \delta$, such that for each $f \in F$, there is a $j \in \{1, \ldots, N\}$ such that $f_l^j \leq f \leq f_u^j$.

Notice that $f_l^j$ and $f_u^j$ are not necessary in $F$.

- The $\delta$-entropy with bracketing is defined as $\log N_1(\delta, F, L^2(Q))$.

- The uniform entropy integral $J_1(\delta, F, L^2(Q))$ is defined by

$$J_1(\delta, F, L^2(Q)) = \int_0^{\delta} \left\{ 1 + \log N_1(\kappa, F, L^2(Q)) \right\}^{\frac{1}{2}} d\kappa.$$

Let $Q_n$ be the empirical measure of $Q$, i.e., $Q_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}(\cdot)$ such that

$$Q_n f = E_{Q_n}[f] = \int f dQ_n = \frac{1}{n} \sum_{i=1}^{n} f(X_i) = \frac{1}{n} \sum_{i=1}^{n} f(X_i).$$

We denote by $G_n = \sqrt{n}(Q_n - Q)$ the standardized empirical process indexed by $F$, and $\|G_n\|_F = \sup_{f \in F} |G_n f|$. Then, for all $f \in F$, we have $Qf = E[f(X)]$ and $G_n f = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (f(X_i) - E[f(X)])$.

Lemma A1. (Lemma 3.4.2. in Van Der Vaart and Wellner [23])

Let $M_0 > 0$ and $F$ be a uniformly bounded class of measurable functions such that, for all $f \in F$, $\|f\|_{\infty} < M_0$ and $Qf^2 < \delta^2$. Then

$$E^{\delta^2}(\|G_n\|_F) \leq c_0 (\delta, F, L^2(Q)) \left\{ 1 + \frac{J_1(\delta, F, L^2(Q))}{\delta^2 \sqrt{n}} M_0 \right\},$$
where $c_0$ is a finite constant, which does not depend on $n$.

**Lemma A2.** (Lemma A.1. in Huang [24])
For any $\lambda > 0$, let

$$\Theta_n = \left\{ \eta \left( \alpha^\top x \right) + \delta^\top w \text{ such that } \|\delta - \delta_0\| \leq \lambda, \quad \eta \in S_n, \|\eta - \eta_0\|_2 \leq \lambda \right\}.$$  

Then, for any $\epsilon \leq \lambda$,

$$\log N \left( \lambda, \Theta_n, L^2(\mathbb{P}) \right) \leq C N_n \log \left( \frac{\lambda}{\epsilon} \right).$$

**Lemma A3.** (Lemma A.2. in Supplement Material of Wang and Yang [25] and Lemma A.4. in Xue and Yang [26])
Let $S_n$ be the space of all polynomial spline functions of order $r$ on $[a, b]$. Under conditions (C1)–(C5), we have

$$A_n = \sup_{\eta_1, \eta_2 \in S_n} \left| \frac{\langle \eta_1, \eta_2 \rangle_n - \langle \eta_1, \eta_2 \rangle}{\|\eta_1\|_2 \|\eta_2\|_2} \right| = O_{a.s.} \left( \sqrt{\log \frac{n}{h n}} \right),$$

where a.s. means “almost surely”.

Recall that $\theta = (\alpha^\top, \gamma^\top, S^\top)^\top$. Let $D_{n,\theta} = \begin{pmatrix} \gamma^\top B' \left( \alpha^\top (\hat{\tau}) X_i \right) J^\top (\tau) & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & B \left( \alpha^\top (\hat{\tau}) X_i \right) \end{pmatrix}$
and $T_i = (X_i^\top, W_i^\top)^\top$.

Denote

$$W_{n,\theta} = \frac{1}{n} \sum_{i=1}^n D_{i,\theta} \begin{pmatrix} T_i \\ 1 \end{pmatrix} \begin{pmatrix} T_i \\ 1 \end{pmatrix}^\top D_{i,\theta}^\top$$
and

$$W_\theta = \frac{1}{n} \sum_{i=1}^n \mathbb{E} \left[ D_{i,\theta} \begin{pmatrix} T_i \\ 1 \end{pmatrix} \begin{pmatrix} T_i \\ 1 \end{pmatrix}^\top D_{i,\theta}^\top \right].$$

**Lemma A4.** (Lemma A.3 in the Supplement Material of Wang and Yang [25])
Under assumptions (C1)–(C5) and (C6)–(C8), it exists a positive constant $C$ such that

$$\sup_{\theta} \left\| W_\theta^{-1} \right\|_2 \leq C \sqrt{N_n}, \quad \text{a.s.,}$$

and

$$\sup_{\theta} \left\| W_{n,\theta}^{-1} \right\|_2 \leq C \sqrt{N_n}, \quad \text{a.s.,}$$

where

$$\|A\|_2 = \sup_{x \neq 0} \frac{\|Ax\|}{\|x\|} = \sup_{\|x\|=1} \|Ax\|.$$  

In what follows, we will enounce lemmas allowing us to prove Theorem 2.

**Lemma A5.** (Lemma 1 in Yu et al. [15])
Under assumptions (C1) and (C8), we have

$$\sup_{u \in [a, b]} \left| \eta_0(u) - \eta_0^\top B_1(u) \right| \leq C \int^r f - \beta_0(t) - \delta_0^\top B_1(t) \leq C k^{-r},$$

where $I$ is the number of inner nodes for $B_1$, and $k$ is the number of inner nodes for $B_2$.  

In what follows, we will give lemmas allowing to prove Theorem 3.

Lemma A6. Under assumptions (C1)–(C8), we have

\[ \frac{1}{n} \sum_{i=1}^{n} \rho_2(m_{0i}) \left( \hat{\eta}(U_{r,0}) - \eta_0(U_{r,0}) \right)'(U_{r,0})' \Phi(X_i) = o_P \left( \frac{1}{\sqrt{n}} \right), \]  
\[ \frac{1}{n} \sum_{i=1}^{n} \rho_2(m_{0i})\eta_0(U_{r,0}) \Phi(X_i) Y'(U_{r,0})(\tau_0)(\hat{\tau} - \tau_0) = o_P \left( \frac{1}{\sqrt{n}} \right), \]  
\[ \frac{1}{n} \sum_{i=1}^{n} \rho_2(m_{0i})\eta_0(U_{r,0}) \Phi(X_i) \Gamma'(U_{r,0})(\delta - \delta_0) = o_P \left( \frac{1}{\sqrt{n}} \right), \]

where

\[ Y(U_{r,0}) = \frac{\mathbb{E}[X \rho_2(m_0(T))|U_{r,0} = u_{r,0}]}{\mathbb{E}[\rho_2(m_0(T))|U_{r,0} = u_{r,0}]}, \]
\[ \Gamma(U_{r,0}) = \frac{\mathbb{E}[W \rho_2(m_0(T))|U_{r,0} = u_{r,0}]}{\mathbb{E}[\rho_2(m_0(T))|U_{r,0} = u_{r,0}]}, \]
\[ \Phi(x) = \Phi(U_{r,0}, x) = x - Y(U_{r,0}), \]

and

\[ \Psi(w) = \Psi(U_{r,0}, w) = w - \Gamma(U_{r,0}). \]

Lemma A7. Under assumptions (C1)–(C8), we have

\[ \frac{1}{n} \sum_{i=1}^{n} \rho_2(m_{oi}) \left( \hat{\eta}(U_{r,oi}) - \eta_0(U_{r,oi}) \right) \Psi(T_i) = o_P \left( \frac{1}{\sqrt{n}} \right), \]
\[ \frac{1}{n} \sum_{i=1}^{n} \rho_2(m_{oi})\eta_0(U_{r,oi}) \Psi(T_i) Y'(U_{r,oi})(\tau_0)(\hat{\tau} - \tau_0) = o_P \left( \frac{1}{\sqrt{n}} \right), \]
\[ \frac{1}{n} \sum_{i=1}^{n} \rho_2(m_{oi})\Psi(U_{r,oi}, W_i) \Gamma'(U_{r,oi})(\delta - \delta_0) = o_P \left( \frac{1}{\sqrt{n}} \right). \]

Summary

In this paper, we introduce estimates for the Generalized Partially Linear Single-Index Models for Functional Data (PLGSIMF). Our estimates are obtained via the Fisher Scoring update equation derived from the quasi likelihood function and the normalized B-splines basis with their derivatives.

We prove the n-consistency and asymptotic normality of our estimates and therefore, firstly, we define estimates, with rates, of the estimator \( \hat{\eta} \), which still converges at the rate to the true non-parametric function \( \eta \). Secondly, we define estimates, with rates, of the estimator \( \hat{\beta} \), which still converges at the rate to the slope function \( \beta \). Finally, we define estimates, with rates, of the estimator \( \hat{\alpha} \) and \( \hat{\delta} \), which still converge at the rate to non-parametric parameters \( \alpha \) and functional parameters \( \delta \), respectively, which still converge normally to the true parameters. A numerical study reveals that our estimation procedure performs well in higher dimensions. The quality of the estimators is illustrated via simulations.

References

1. Nelder, J.A.; Wedderburn, R.W.M. Generalized Linear Models. J. R. Stat. Soc. Ser. A 1972, 135, 370–384, https://doi.org/10.2307/2344614.
2. McCullagh, P.; Nelder, J.A. Generalized Linear Models, 2nd ed.; Chapman and Hall: London, UK, 1972; http://dx.doi.org/10.1007/978-1-4899-3242-6.
3. Hastie, T.J.; Tibshirani, R.J. Generalized Additive Models; Chapman and Hall: London, UK, 1990; https://doi.org/10.1002/sim.4780110717.
4. Wood, S. Generalized Additive Models: An Introduction with R, Second ed.; CRC Press: Boca Raton, FL, USA, 2017; doi:10.1201/9781315370279.
5. Dunn, P. K.; Smyth, G.K. Generalized Linear Models with Examples in R; Springer Texts in Statistics; Springer: New York, NY, USA, 2018.
6. Lai, P.; Tian, Y.; Lian, H. Estimation and variable selection for generalised partially linear single-index models. *J. Nonparametr. Stat.* 2014, 26, 171–185.
7. Ferraty, F.; Vieu, P. *Nonparametric Functional Data Analysis: Theory and Practice;* Springer Series in Statistics; Springer: New York, NY, USA, 2006.
8. Caroll, R.J.; Fan, J.; Gijbels, I.; Wand, M.P. Generalized partially linear single-index models. *J. Am. Stat. Assoc.* 1997, 92, 477–489.
9. Li, C.S., Lu, M. A lack-of-fit test for generalized linear models via single-index techniques. *Comput. Stat.* 2018, 33, 731–756, https://doi.org/10.1007/s00180-018-0802-2.
10. Li, W.; Cao, G. Efficient estimation for generalized partially linear single-index models. *Bernoulli* 2018, 24, 1101–1127.
11. Aneiros-Perez, G.; Vieu, P. Semi functional partial linear regression. *Stat. Probab. Lett.* 2006, 76, 1102–1110.
12. Horváth, L.; Kokoszka, P. Inference for Functional Data with Applications. *Comput. Sci.* 2012, doi:10.1007/978-1-4614-3655-3.
13. Aneiros-Perez, G.; Vieu, P. Partial linear modelling with multi-functional covariates. *Comput. Stat.* 2015, 30, 647–671.
14. Ramsay, J.O.; Silverman, B.W. *Functional Data Analysis;* Springer: New York, NY, USA, 2005; http://dx.doi.org/10.1002/0470013192.bsa239.
15. Yu, P.; Du, J.; Zhang, Z. Single-index partially functional linear regression model. *Stat. Pap.* 2020, 61, 1107–1123.
16. Cao, R., Du, J., Zhou, J. and Xie, T. FPCA-based estimation for generalized functional partially linear models. *Stat Pap.* 2020, 61, 2715–2735, https://doi.org/10.1007/s00362-018-01066-8.
17. Rachdi, M.; Alahiane, M.; Ouassou, I.; Vieu, P. Generalized Functional Partially Linear Single-index Models. In *IWFOS 2020: Functional and High-Dimensional Statistics and Related Fields;* Springer: Cham, Switzerland, 2020; pp. 221–228, doi:10.1007/978-3-030-47756-1.
18. De Boor, C. *A Practical Guide to Splines;* Revised Edition of Applied Mathematical Sciences; Springer: Berlin, Germany, 2001; Volume 27.
19. Yu, Y.; Ruppert, D. Penalized spline estimation for partially linear single-index models. *J. Am. Stat. Assoc.* 2002, 16, 1042–1054.
20. Peng, Q.; Zhou, J.; Tang, N. Varying coefficient partially functional linear regression models. *Stat. Pap.* 2015, 57, 827–841.
21. Pollard, D. Asymptotics for least absolute deviation regression estimators. *Econom. Theory* 1991, 7, 186–199.
22. Stone, C.J. The dimensionality reduction principle for generalized additive models. *Ann. Stat.* 1986, 14, 590–606.
23. Van der Vaart, A.W.; Wellner, J.A. *Weak Convergence and Empirical Processes with Applications To Statistics;* Springer: New York, NY, USA, 1996.
24. Huang, J. Efficient estimation of the partly linear additive Cox model. *Ann. Stat.* 1999, 27, 1536–1563.
25. Li, W.; Yang, L. Spline estimation of single-index models. *Stat. Sin.* 2009, 19, 765–783.
26. Xue, L.; Yang, L. Additive coefficient modelling via polynomial spline. *Stat. Sin.* 2006, 16, 1423–1446.