


Article

# Generalized Partially Functional Linear Model with Unknown Link Function

Weiwei Xiao <sup>1,\*</sup> , Songxuan Li <sup>1</sup> and Haiyan Liu <sup>2</sup>

<sup>1</sup> School of Science, North China University of Technology, Beijing 100144, China; 2021312050118@mail.ncut.edu.cn

<sup>2</sup> Department of Statistics, University of Leeds, Leeds LS2 9JT, UK; h.liu1@leeds.ac.uk

\* Correspondence: xiaoww@ncut.edu.cn

**Abstract:** In existing models with an unknown link function, the issue of predictors containing both multiple functional data and multiple scalar data has not been studied. To fill this gap, we propose a generalized partially functional linear model, which not only models the relationship between multiple scalar and functional predictors and responses, but also automatically estimates the link function. Specifically, we use the functional principal component analysis method to reduce the dimensionality of functional predictors, estimate the regression coefficients using the maximum likelihood estimation method, estimate the link function using the method of local linear regression, iteratively obtain the final estimator, and establish the asymptotic normality of the estimator. The asymptotic normality is illustrated through simulation experiments. Finally, the proposed model is applied to study the influence of environmental, economic, and medical levels on life expectancy in China. In the study, functional predictors are the daily air quality index, temperature, and humidity of 58 cities in 2020, and scalar predictors are GDP and the number of beds in hospitals. The experimental results indicate that the unknown link function model has a smaller prediction error and better performance than both the model with the known link function and the model without a link function.

**Keywords:** functional data analysis; unknown link function; generalized functional linear model average life expectancy

**MSC:** 00A71



**Citation:** Xiao, W.; Li, S.; Liu, H. Generalized Partially Functional Linear Model with Unknown Link Function. *Axioms* **2023**, *12*, 1089. <https://doi.org/10.3390/axioms12121089>

Academic Editors: Ricardo Abreu-Blaya and Juan Carlos Hernández Gómez

Received: 13 October 2023  
Revised: 16 November 2023  
Accepted: 22 November 2023  
Published: 28 November 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

In 1982, Ramsay [1] first proposed the definition of functional data, laying a foundation for the development of functional data analysis. In 2005, Ramsay and Silverman provided a detailed introduction to the general methods and steps of functional data analysis, including functional principal component analysis and functional linear regression models in their book [2]. In 2012, Horváth and Kokoszka [3] focused on the inferential methods in functional data analysis.

In 2009, Shin [4] proposed a partial functional linear model (PFLM), which explores the relationship between a scalar response variable and mixed-type predictors. In 2012, Shin and Lee [5] derived the asymptotic prediction rate of PFLM and compared it with that of other functional regression models.

In 2002, James [6] proposed generalized linear models with functional predictors and applied them to standard missing data problems. In 2005, Müller and Stadtmüller [7] proposed a generalized functional linear regression model where the response variable is a scalar and the predictor is a random function. They also considered the situation where the link and variance functions were unknown. In 2015, Shang and Cheng [8] proposed a roughness regularization approach in making nonparametric inference for generalized functional linear models with known link functions. In 2019, Wong et al. [9] investigated

a class of partially linear functional additive models that predict a scalar response by both the parametric effects of a multivariate predictor and the non-parametric effects of a multivariate functional predictor.

In a generalized linear model, sometimes the link function may not be known exactly, but can be assumed to be of some general ‘parametric’ form. In 1984, Scallan et al. [10] showed how generalized linear models can be extended to fit models with such link functions.

In 1994, Weisberg and Welsh [11] used kernel smoothing estimation to estimate the link function and estimated regression coefficients through the link function, then alternated between these two steps, which effectively solves the fitting problem when the link function is unknown. However, kernel smoothing estimation may have problems at the boundary, so local polynomial fitting is introduced, which performs better near the boundary.

In 1998, Chiou and Müller [12] considered the condition of the link and the variance functions to be unknown but smooth. Consistency results for the link and the variance function estimators, as well as the sampling distribution of the regression coefficients, were obtained. In 2005, Chiou and Müller [13] introduced a flexible marginal modeling approach for statistical inference for clustered and longitudinal data under minimal assumptions. The predictor was longitudinal data in the model. The estimated estimating equation approach was semi-parametric. The semi-parametric model proposed was fitted by quasi-likelihood regression. The consistency of the estimates of the link and variance functions and the asymptotic limit distribution of regression coefficients were given. In addition, there are other methods to estimate unknown functions. In 2009, Bai et al. [14] focused on single-index models for longitudinal data. They proposed a procedure to estimate the single-index component and the unknown link function based on the combination of the penalized splines and quadratic inference functions. In 2012, Pang and Xue [15] generalized the single-index models to the scenarios with random effects. The link function was estimated by using the local linear smoother. A new set of estimating equations modified for the boundary effects was proposed to estimate the index coefficients. In 2017, Yuan and Diao [16] developed a sieve maximum likelihood estimation for generalized linear models, in which the estimator of the unknown link function was assumed to lie in a sieve space. Various methods of sieves including the B-spline and P-spline-based methods were introduced.

In 2017, Kokoszka and Reimherr [17] wrote a book that introduced the basic concepts, methods, and applications of functional data analysis. The book provided a clear and systematic overview, covering key areas such as representation, smoothing, interpolation, statistical modeling, and inference for functional data. It also included detailed explanations of practical examples and computational methods. In 2023, Rao and Reimherr [18] introduced a novel neural network-based nonlinear model of functional data designed to exploit the structure of functional data and fit it with a derived function gradient optimization algorithm, demonstrating the effectiveness of these methods in dealing with complex functional models and providing new breakthroughs for deep learning applications in the field of functional data analysis.

The relationship between environmental factors and human health has been a topic of significant research interest in recent years. In 2012, Huang et al. [19] explored the relationship between temperature and years of life lost (YLL). The study found that both high and low temperatures lead to an increase in YLL, with high temperatures having a greater impact. In 2020, Yang et al. [20] applied a generalized additive model to assess the associations between daily PM<sub>2.5</sub> exposure and YLL due to respiratory diseases in 96 Chinese cities during 2013–2016. They further estimated the avoidable YLL and potential gains in life expectancy under the assumption that daily PM<sub>2.5</sub> level met World Health Organization standards. In 2021, Deryugina and Molitor [21] explored the factors influencing life expectancy across the United States. The study found that individuals living in areas with severe air pollution, poor water quality, and inadequate healthcare facilities generally had shorter life expectancy and poorer health conditions.

In summary, the existing models with unknown link functions have not addressed the issue of the generalized partially functional regression model, which involves regressing the response variable on multiple functional and scalar predictors. To fill this gap, this study proposes a generalized partially functional linear model with an unknown link function. The proposed model avoids the problem of decreased model accuracy caused by selecting an incorrect link function. The predictors in the proposed model include both multiple functional data and multiple scalar data. It reveals the complex relationships between variables and provides a flexible and effective modeling approach. It can achieve better prediction and explanation.

The paper is organized as follows. All the published works and definitions that are referred to in the process of theorem proving are introduced in Section 2. The abbreviations used in the article are introduced in the Section 3.1. The generalized partial functional linear model with unknown link function is proposed in Section 3.2. The estimation of the regression coefficients and the link function is discussed in Section 3.3. In Section 4, asymptotic normality of estimators are derived. Simulation results are reported in Section 5. The average life expectancy study in 58 cities in China is given in Section 6. In Section 7, a brief summary and limitations of the research are provided. Possible applications and future directions are presented in Section 7.

## 2. Preliminaries

In this section, we provide an overview of the published works and definitions that are relevant to our research. These preliminary concepts and references lay the foundation for a better understanding of the subsequent discussion.

(1) In 1982, Mack and Silverman [22] provided a comprehensive analysis of the weak and strong uniform consistency properties of kernel regression estimates, highlighted their theoretical properties and practical significance in non-parametric regression modeling. In this paper, we directly apply the results of Proposition 4 as Lemma 1 for Theorem 1 in this paper.

(2) In 1995, Masry and Tjøstheim [23] discussed the estimation and identification of nonlinear time series of ARCH type. They provided an estimation method to obtain consistent estimates of the parameters and proved the asymptotic normality. They also explored model identification methods. Their studies are of significant importance for modeling and analyzing financial time series. Theorem 3.3 in their work is used to prove Theorem 1 in this paper.

(3) In 1999, Chiou and Müller [24] focused on the study of non-parametric quasi-likelihood methods. They provided the theoretical derivation process of this method, and explored its applications in statistical inference. Theorem 4.1 in their paper is used to prove Lemma 2 and Lemma 3 for Theorem 2 in this paper.

(4) In 2021, Xiao et al. [25] proposed a generalized partially functional linear regression model where the response variable is 0 or 1 and the predictors were multiple functional and scalar, and the asymptotic property of the estimated coefficients in the model was established. The proof method of Theorem 1 in [25] is used to prove Theorem 2 in this work.

## 3. Model and Estimation

The data we observe for the  $i$ -th subject are  $\{Y_i, X_{i1}(t), X_{i2}(t), \dots, X_{id}(t), Z_i\}$ ,  $i = 1, \dots, n$ . We assume that these data are independent, identically distributed (i.i.d) copies of  $\{Y, X_1(t), \dots, X_d(t), Z\}$ . For  $j = 1, \dots, d$ , the functional predictor  $X_j(t)$  is a random curve.  $X_{ij}(t), i = 1, 2, \dots, n$  are samples of  $X_j(t)$  and  $X_{ij}(t)$  are square integrable on a real bounded interval  $T$ , i.e.,  $X_{ij}(t) \in L^2(T)$ .  $L^2(T)$  refers to the space of square integrable functions defined on  $T$ . And the scalar predictor vector  $Z = (Z_1, Z_2, \dots, Z_q)^T$  is a  $q$  dimensional random vector. The response  $Y$  is a real-valued random variable that may be binary or count.

### 3.1. Abbreviation Introduction

Table 1 is a list of the abbreviations we use in this work along with their corresponding full forms:

**Table 1.** The abbreviations and their corresponding full forms.

Abbreviation	Full Form
FPCA	Functional principal component analysis
KL expansion	Karhunen–Loeve expansion
RMISE	Root Mean Integrated Square Error
SD	Standard Deviation
GCV	Generalized Cross Validation
MAE	Mean Absolute Error
MSE	Mean Squared Error
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative

### 3.2. Model

We establish a model for the relationship between the response variable  $Y_i$  and the predictors  $X_{ij}(t), j = 1, 2, \dots, d$  and  $Z_i$ :

$$Y_i = g\left(\sum_{j=1}^d \int_T X_{ij}(t)\beta_j(t)dt + Z_i^T\gamma\right) + \varepsilon_i, \tag{1}$$

where  $\beta_j(\cdot)$  is the regression coefficient function that needs to be estimated for the functional predictors  $X_{ij}(t)$ ;  $\gamma$  is a  $q$  dimensional vector with the elements to be the regression coefficients for the scalar predictors  $Z_i$  that need to be estimated, i.e.,  $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_q)^T$ . Here  $\varepsilon_i$  is i.i.d copies of  $\varepsilon$ , which is the random error variable and  $\varepsilon = Y - g(\eta)$ ,  $E[\varepsilon|X_j(t), Z] = 0$ , where

$$\eta = \sum_{j=1}^d \int_T X_j(t)\beta_j(t)dt + Z^T\gamma.$$

The relationship between the response variable  $Y$  and  $\eta$  is established through  $g(\cdot)$ , i.e.,  $E[Y|X_j(T), Z] = \mu = g(\eta)$ .  $g(\cdot)$  is the link function that is unknown and needs to be estimated in this paper.

Let  $\sigma^2(\cdot)$  be a variance function that satisfies  $\sigma^2(\cdot) \geq c > 0$  for a constant  $c > 0$ , such that

$$\begin{aligned} \text{Var}[Y|X_j(t), Z] &= \sigma^2(\mu) = \sigma^2(g(\eta)), \\ \text{Var}[\varepsilon] &= E[\varepsilon^2] = \sigma^2(E[Y|X_j(t), Z]). \end{aligned}$$

To reduce the dimensionality of the functional predictors  $X_{ij}(t)$ , we adopt the method of FPCA in this paper. First, we need to standardize the original data by centering them, so that  $E[X_{ij}(t)] = 0, j = 1, \dots, d$ , and  $E[Z_l] = 0, l = 1, \dots, q$ .

By KL expansion and Mercer’s theorem,  $X_{ij}(t)$  can be expanded as

$$X_{ij}(t) = \sum_{k=1}^{\infty} \xi_{ijk}\rho_{jk}(t), \tag{2}$$

where  $\xi_{ijk}$  represents the functional principal component scores, and  $\rho_{jk}(\cdot)$  are called functional principal components, which are the eigenfunctions of the covariance operator

of  $X_{ij}(t)$ . Notice that  $\rho_{jk}(\cdot), k = 1, 2, \dots$  form an orthonormal basis for the function space  $L^2(T)$ . Then regression coefficient function  $\beta_j(t) \in L^2(T)$  can be expanded as

$$\beta_j(t) = \sum_{k=1}^{\infty} \chi_{jk} \rho_{jk}(t). \tag{3}$$

where  $\chi_{jk}$  represents the functional principal component scores.

After plugging the above two expansions into (1), we have

$$Y_i = g \left( \sum_{j=1}^d \sum_{k=1}^{m_j} \xi_{ijk} \chi_{jk} + Z_i^T \gamma \right) + \varepsilon_i. \tag{4}$$

In (4), we truncated the predictors at  $m_j$  (depending on sample size  $n$ ), and  $m_j$  increases asymptotically with  $n \rightarrow \infty$ .

### 3.3. Estimation

Define a parameter vector  $\theta_0$ , where

$$\theta_0 = (\chi_{11}, \chi_{12}, \dots, \chi_{1m_1}, \dots, \chi_{d1}, \chi_{d2}, \dots, \chi_{dm_d}, \gamma_1, \dots, \gamma_q)^T.$$

For the estimation of the parameter vector  $\theta$  and the link function  $g$ , we use an iterative estimation method to obtain the final estimates. Let there exist a constant  $c > 0$ ; with this  $c$  and  $n$ , we can define  $\theta_n = \{\theta : \|\theta - \theta_0\| \leq cn^{-1/2}\}$ . The norm of finite dimensional spaces used in this paper is the Euclidean norm. The overall iterative process is briefly described below:

**Step 1** To obtain the estimate  $\theta^{(0)}$  of  $\theta_0$  by solving Equation (5), it is assumed that the link function  $g(\cdot)$  is known. The link function  $g(\cdot)$  is required to be second-order continuously differentiable to ensure the existence of the Hessian matrix, moreover, for the variance function  $\sigma^2(\cdot)$  is defined on the range of link function and is strictly positive.

$$U(\theta) = \sum_{i=1}^n (Y_i - \mu_i) \frac{g'(\eta_i)}{\sigma^2(\mu_i)} \Delta_i = 0, \tag{5}$$

where  $\eta_i = \sum_{j=1}^d \sum_{k=1}^{m_j} \xi_{ijk} \tilde{\chi}_{jk} + Z_i \tilde{\gamma}$ ,  $\tilde{\chi}$  near  $\chi$ ,  $\tilde{\gamma}$  near  $\gamma$ ,  $\mu_i = g(\eta_i)$  and

$$\Delta_i = (\xi_{i11}, \xi_{i12}, \dots, \xi_{i1m_1}, \dots, \xi_{id1}, \xi_{id2}, \dots, \xi_{idm_d}, z_{i1}, \dots, z_{iq})^T.$$

Here,  $\tilde{\chi}$  and  $\tilde{\gamma}$  represent the corresponding estimated value in step 1 but not the final estimate.

We introduce the following matrix:

$$\begin{aligned} D_0 &= D_{n,q} = (z_{il})_{1 \leq i \leq n, 1 \leq l \leq q}, \\ D_j &= D_{n,m_j} = (\xi_{ijk})_{1 \leq i \leq n, 1 \leq j \leq d, 1 \leq k \leq m_j}, \\ D &= D_{n,q + \sum_{j=1}^d m_j} = (D_0, D_1, \dots, D_d), \\ V &= \text{diag}(\sigma^2(\mu_1), \sigma^2(\mu_2), \dots, \sigma^2(\mu_n)), \\ G &= \text{diag}(g'(\eta_i))_{1 \leq i \leq n}, \\ Y &= (Y_1, Y_2, \dots, Y_n)^T, \\ \mu &= (\mu_1, \mu_2, \dots, \mu_n)^T. \end{aligned}$$

Then, Equation (5) can be expressed in matrix form, i.e.

$$D^T V^{-1} G(Y - \mu) = 0.$$

We can solve it by the weighted least squares method. A Taylor expansion of  $g^{-1}(Y)$ , where

$$\begin{aligned} g^{-1}(Y) &= g^{-1}(\mu) + [g^{-1}(\mu)]'(Y - \mu) \\ &= \eta + G^{-1}(Y - \mu), \end{aligned}$$

and then we can get

$$D^T W(g^{-1}(Y) - \eta) = 0,$$

where  $W = V^{-1}G^2$ . Simplification yields estimates

$$\begin{aligned} \tilde{\chi}_j^{(0)} &= (D_j^T W D_j)^{-1} D_j^T W g^{-1}(Y), \\ \tilde{\gamma}_j^{(0)} &= (D_0^T W D_0)^{-1} D_0^T W g^{-1}(Y), \end{aligned}$$

where  $\tilde{\chi}_j^{(0)} = (\tilde{\chi}_{j1}^{(0)}, \dots, \tilde{\chi}_{jm}^{(0)})^T, j = 1, 2, \dots, d, \tilde{\gamma}^{(0)} = (\tilde{\gamma}_1^{(0)}, \tilde{\gamma}_2^{(0)}, \dots, \tilde{\gamma}_q^{(0)})^T$ .

Let

$$\tilde{\theta}^{(0)} = (\tilde{\chi}_{11}^{(0)}, \tilde{\chi}_{12}^{(0)}, \dots, \tilde{\chi}_{1m_1}^{(0)}, \dots, \tilde{\chi}_{d1}^{(0)}, \tilde{\chi}_{d2}^{(0)}, \dots, \tilde{\chi}_{dm_d}^{(0)}, \tilde{\gamma}_1^{(0)}, \tilde{\gamma}_2^{(0)}, \dots, \tilde{\gamma}_q^{(0)})^T.$$

**Step 2** By local linear regression, the estimates  $g^{(0)}, g'^{(0)}$  of the link functions  $g, g'$  are obtained.

Let the bandwidth  $b = b_n$  of the kernel function  $k(\cdot)$  converge to zero and define  $k_b(\cdot) = b^{-1}k(\cdot/b)$ . Since the convergence rates of  $g(\cdot)$  and  $g'(\cdot)$  are different, their bandwidth choices should also be different. Let  $h_0 = h_{0n}$  denote the bandwidth of  $g(\cdot)$ , and  $h_1 = h_{1n}$  denote the bandwidth of  $g'(\cdot)$ , but in this paper, for simplicity, the bandwidth  $h = h_0 = h_1$  is chosen. Let the distributions of both the functional predictors  $X_j(t)$  and the scalar predictors  $Z$  belong to a compact support set  $U$ , and we have  $\Omega = \{u = \eta_i | X_j(t), Z \in U\}$ . To simplify the expression, we let  $g = g(u; \theta), g' = g'(u; \theta)$ . For a fixed  $\theta$ , apply the method of local linear regression to obtain an initial estimate of  $\tilde{g}^{(0)}$  and  $\tilde{g}'^{(0)}$  for  $g$  and  $g'$ , respectively. We minimize the weighted sum of squares at any point  $u$ , and the formula for calculating the weighted sum of squares is

$$\sum_{i=1}^n [Y_i - g - g'(\eta_i - u)]^2 k_h(\eta_i - u). \tag{6}$$

Through minimizing (6), we can obtain  $\tilde{g}^{(0)}$  and  $\tilde{g}'^{(0)}$ , and they can be represented as  $\tilde{g}^{(0)} = \sum_{i=1}^n \omega_i(u; \theta) Y_i, \tilde{g}'^{(0)} = \sum_{i=1}^n \tilde{\omega}_i(u; \theta) Y_i$ , where

$$\begin{aligned} \omega_i(u; \theta) &= \frac{k_h(\eta_i - u)[\varphi_{n,2}(u; \theta, h) - (\eta_i - u)\varphi_{n,1}(u; \theta, h)]}{\varphi_{n,0}(u; \theta, h)\varphi_{n,2}(u; \theta, h) - \varphi_{n,1}^2(u; \theta, h)}, \\ \tilde{\omega}_i(u; \theta) &= \frac{k_h(\eta_i - u)[(\eta_i - u)\varphi_{n,0}(u; \theta, h) - \varphi_{n,1}(u; \theta, h)]}{\varphi_{n,0}(u; \theta, h)\varphi_{n,2}(u; \theta, h) - \varphi_{n,1}^2(u; \theta, h)}, \\ \varphi_{n,l}(u; \theta, h) &= \frac{1}{n} \sum_{i=1}^n \left(\frac{\eta_i - u}{h}\right)^l k_h(\eta_i - u), l = 0, 1, 2. \end{aligned}$$

**Step 3** Using the method of **Step 1**, the link function is replaced by the estimated link functions  $\tilde{g}^{(\alpha)}$  and  $\tilde{g}'^{(\alpha)}$ , where  $\alpha = 0, 1, 2, \dots$ . To update  $\tilde{\theta}^{(\alpha)}$ , solve the estimation equation (5) for  $\theta$ . From this we can obtain the estimated value of  $\tilde{\theta}^{(\alpha)}$

$$\tilde{\theta}^{(\alpha)} = (\tilde{\chi}_{11}^{(\alpha)}, \tilde{\chi}_{12}^{(\alpha)}, \dots, \tilde{\chi}_{1m_1}^{(\alpha)}, \dots, \tilde{\chi}_{d1}^{(\alpha)}, \tilde{\chi}_{d2}^{(\alpha)}, \dots, \tilde{\chi}_{dm_d}^{(\alpha)}, \tilde{\gamma}_1^{(\alpha)}, \tilde{\gamma}_2^{(\alpha)}, \dots, \tilde{\gamma}_q^{(\alpha)})^T.$$

**Step 4** Using the method in **Step 2**, the parameter vector is replaced by the estimated  $\tilde{\theta}^{(\alpha)} = (\tilde{\chi}_{j1}^{(\alpha)}, \tilde{\chi}_{j2}^{(\alpha)}, \dots, \tilde{\chi}_{jm}^{(\alpha)}, \tilde{\gamma}_1^{(\alpha)}, \tilde{\gamma}_2^{(\alpha)}, \dots, \tilde{\gamma}_q^{(\alpha)})^T$ , where  $\alpha = 1, 2, 3, \dots$ . From this we obtain the estimates  $\tilde{g}^{(\alpha)}$  and  $\tilde{g}'^{(\alpha)}$  for  $g$  and  $g'$ , where  $\alpha = 1, 2, 3, \dots$

**Step 5** Repeat the above steps until  $|\tilde{\theta}^{(\alpha+1)} - \tilde{\theta}^{(\alpha)}|$  converge, and stop the iteration.

**Step 6** The final estimate of the regression coefficient  $\theta$  is obtained as  $\hat{\theta}$ , and the estimate of the link function  $g$  is obtained as  $\hat{g}$ .

#### 4. Asymptotic Properties

To derive the asymptotics of the estimates of the link function  $\hat{g}$  and the regression coefficients  $\hat{\theta}$ , some additional assumptions are required:

- (C1) There exists  $b = \max(4, c)$  for a constant  $c > 0$ , such that  $E[\int_T \|X_j(t)\|^b dt] < \infty$ ,  $j = 1, \dots, d$ ,  $E[\|Z\|^b] < \infty$ ,  $E[\varepsilon] < \infty$ .
- (C2) Let the density function  $f(\cdot)$  of  $\eta_i$  be strictly positive, and  $f(\cdot)$  satisfies the first-order Lipschitz condition when  $\theta \rightarrow \theta_0$ .
- (C3) The kernel function  $k(\cdot)$  satisfies the first-order Lipschitz condition and is a bounded and continuous symmetric probability density function and satisfies  $\int_{-\infty}^{\infty} u^2 k(u) du \neq 0$ ,  $\int_{-\infty}^{\infty} |u|^2 k(u) du < \infty$ .
- (C4)  $nh^4 / \log^2 n \rightarrow \infty$ ,  $nh^5 = O(1)$ . Here,  $h$  is the bandwidth of the kernel function.
- (C5) For  $j = 1, \dots, d$ ,  $m_j n^{-1/4} \rightarrow 0$  as  $n \rightarrow \infty$ .

**Remark 1.** (C1) It is a necessary condition for the asymptotic normality of the estimator. (C2) Ensures that  $\tilde{g}^{(\alpha)}$ ,  $\tilde{g}'^{(\alpha)}$  are far from 0 when  $\tilde{\theta}^{(\alpha)}$  is close enough to  $\theta$ . (C3) The usual assumptions about the kernel function. (C4) The usual assumptions about the bandwidth. (C5) Some controls are applied to  $m$  in order to make the convergence faster.

##### 4.1. Asymptotic Convergence of $g^{(\alpha)}$

**Lemma 1.** Let  $(\zeta_1, \iota_1), \dots, (\zeta_n, \iota_n)$  be independent and identically distributed random vectors. Furthermore, assume that for any  $s > 0$ , there exist  $E|\iota_i|^s < \infty$ ,  $i = 1, \dots, n$  and  $\sup_{\zeta} \int |\iota|^s f(\zeta, \iota) d\iota < \infty$

such that  $f(\cdot, \cdot)$  is the joint density function of  $(\zeta, \iota)$ . Let  $k(\cdot)$  be a bounded and strictly positive kernel function that satisfies the Lipschitz condition, we have

$$\sup_{\zeta} \left| \frac{1}{n} \sum_{i=1}^n [k_h(\zeta_i - \zeta)\iota_i - E[k_h(\zeta_i - \zeta)\iota_i]] \right| = O_p \left[ \left( \frac{\log(1/h)}{nh} \right)^{\frac{1}{2}} \right].$$

**Proof.** See Proposition 4 in Mack and Silverman (1982) [22]. □

**Theorem 1.** If we assume that (C1)–(C5) holds, for  $\sigma^2 > 0$ , then we have

$$\sqrt{nh}[\tilde{g}^{(\alpha)}(u; \theta) - g(u) - I(u)] \xrightarrow{D} N(0, \vartheta^2(u))$$

where  $I(u) = \frac{1}{2}h^2\mu_2g''(u)$ ,  $\vartheta^2(u) = \nu_2\sigma^2 / \sum_{j=1}^d f_j(u)$ , and for the kernel function, let

$$\mu_l = \int u^l k(u) du, \nu_l = \int k^l(u) du, l = 1, 2.$$

**Proof.**

$$\left(\tilde{g}^{(\alpha)}(u; \theta_0), h\tilde{g}'^{(\alpha)}(u; \theta_0)\right)^T = \Gamma_n^{-1}(u; \theta_0)\Phi_n(u; \theta_0), \tag{7}$$

where

$$\begin{aligned} \Gamma_n(u; \theta_0) &= \begin{pmatrix} \varphi_{n,0}(u; \theta_0) & \varphi_{n,1}(u; \theta_0) \\ \varphi_{n,1}(u; \theta_0) & \varphi_{n,2}(u; \theta_0) \end{pmatrix}, \\ \Phi_n(u; \theta_0) &= (\phi_{n,0}(u; \theta_0), \phi_{n,1}(u; \theta_0))^T, \\ \varphi_{n,l}(u; \theta_0) &= \frac{1}{n} \sum_{i=1}^n \left(\frac{\eta_{i0} - u}{h}\right)^l k_h(\eta_{i0} - u), l = 0, 1, 2, \\ \phi_{n,l}(u; \theta_0) &= \frac{1}{n} \sum_{i=1}^n Y_i \left(\frac{\eta_{i0} - u}{h}\right)^l k_h(\eta_{i0} - u), l = 0, 1, \\ \eta_{i0} &= \sum_{j=1}^d \sum_{k=1}^{m_j} \xi_{ijk} \chi_{jk} + Z_i^T \gamma. \end{aligned}$$

By expanding  $\varphi_{n,l}(u; \theta)$ ,  $l = 0, 1, 2, 3$ , we obtain that

$$\begin{aligned} E[\varphi_{n,l}(u; \theta)] &= E\left[\frac{1}{n} \sum_{i=1}^n \left(\frac{\eta_i - u}{h}\right)^l k_h(\eta_i - u)\right] \\ &= \sum_{j=1}^d f_j(u) \mu_l + O(h). \end{aligned} \tag{8}$$

From Lemma 1, it can be proved that for  $l = 0, 1, 2, 3$

$$\varphi_{n,l}(u; \theta) - E[\varphi_{n,l}(u; \theta)] = O_p\left[\left(\frac{\log(1/h)}{nh}\right)^{\frac{1}{2}}\right]. \tag{9}$$

Taking (8) into (9), we can obtain that

$$\varphi_{n,l}(u; \theta) = \sum_{j=1}^d f_j(u) \mu_l + O_p\left[\left(\frac{\log(1/h)}{nh}\right)^{\frac{1}{2}} + h\right].$$

Then

$$\Gamma_n(u; \theta) = \Gamma(u) + O_p\left[\left(\frac{\log(1/h)}{nh}\right)^{\frac{1}{2}} + h\right],$$

where  $\Gamma(u) = \sum_{j=1}^d f_j(u) \otimes \text{diag}(1, \mu_2)$ ,  $\otimes$  indicates the Kronecker product.

Inverting the matrix  $\Gamma_n(u; \theta)$ , we get

$$\Gamma_n^{-1}(u; \theta) = \Gamma^{-1}(u) + O_p\left[\left(\frac{\log(1/h)}{nh}\right)^{\frac{1}{2}} + h\right].$$

Let

$$\phi_{n,l}^*(u; \theta) = \frac{1}{n} \sum_{i=1}^n [Y_i - g(\eta_i)] \left(\frac{\eta_i - u}{h}\right)^l k_h(\eta_i - u), \tag{10}$$

where  $l = 0, 1$ , and

$$\Phi_n^* \equiv \Phi_n^*(u; \theta) = (\phi_{n,0}^*(u; \theta), \phi_{n,1}^*(u; \theta))^T.$$



By expanding  $\phi_{n,l}^*(u; \theta)$ , we obtain when  $l = 0, 1, 2$ ,

$$E[\phi_{n,l}^*(u; \theta)] = E\left[\frac{1}{n} \sum_{i=1}^n [Y_i - g(\eta_i)] \left(\frac{\eta_i - u}{h}\right)^l k_h(\eta_i - u)\right] = O(n^{-\frac{1}{2}}). \tag{11}$$

From Lemma 1, combined with (10) and (11), we can prove that

$$\phi_{n,l}^*(u; \theta) = O_p\left[\left(\frac{\log(1/h)}{nh}\right)^{\frac{1}{2}} + n^{-1/2}\right].$$

The Taylor expansion of  $g(\eta_i)$  at  $u$  is

$$\Phi_n - \Phi_n^* = \Gamma_n \begin{pmatrix} g(u) \\ hg'(u) \end{pmatrix} + \frac{1}{2}h^2 \begin{pmatrix} \varphi_{n,2}g''(u) \\ \varphi_{n,3}g''(u) \end{pmatrix} + o_p(h^2), \tag{12}$$

where  $\Gamma_n = \Gamma_n(u; \theta)$ ,  $\varphi_{n,l} = \varphi_{n,l}(u; \theta)$ ,  $l = 2, 3$ .

Combining (7) and (12), we can obtain

$$\begin{aligned} \begin{pmatrix} \tilde{g}^{(\alpha)} - g \\ h[\tilde{g}'^{(\alpha)} - g'] \end{pmatrix} &= \Gamma_n^{-1}\Phi_n^* + \frac{1}{2}\Gamma_n^{-1}h^2 \begin{pmatrix} \varphi_{n,2}g''(u) \\ \varphi_{n,3}g''(u) \end{pmatrix} + o_p(h^2) \\ &= \Gamma^{-1}(u)\Phi_n^* + \frac{1}{2}h^2 \begin{pmatrix} \mu_2g''(u) \\ \frac{\mu_3}{\mu_2}g''(u) \end{pmatrix} + o_p(h^2 + n^{-1/2}), \end{aligned}$$

where

$$\tilde{g}^{(\alpha)} - g = \left[\sum_{j=1}^d f_j(u)\right]^{-1} \phi_{n,0}^*(u; \theta) + \frac{1}{2}h^2\mu_2g''(u) + o_p(h^2 + n^{-1/2}). \tag{13}$$

Since  $\|\theta - \theta_0\| = O(n^{-1/2})$ , (10) can be transformed into

$$\phi_{n,0}^*(u; \theta) = \frac{1}{n} \sum_{i=1}^n [Y_i - g(\eta_{i0})]k_h(\eta_{i0} - u) + O_p(n^{-1/2}).$$

Taking it into (13) and combining it with Theorem 3.3 in Masry and Tjøstheim (1995) [23], finally, Theorem 1 can be proved.  $\square$

**Corollary 1.** *If we further refine the condition in assumption (C4) such that  $nh^5 \rightarrow 0$ , then it follows that*

$$\sqrt{nh}[\tilde{g}^{(\alpha)}(u; \theta) - g(u)] \xrightarrow{D} N(0, \vartheta^2(u)).$$

#### 4.2. Asymptotic Convergence of $\hat{\theta}$

First, we need to provide some more specific explanations for the estimation iteration process mentioned in ‘‘Estimation’’, which makes some preparation for Theorem 2.

(1) solving  $U(\theta)$  by Equation (5) given the assumption that the link function is known. Assume  $Q_{\theta_0} = -\frac{\partial U(\theta_0)}{\partial \mu} \frac{\partial \mu}{\partial \theta_0^T}$ , then it follows that

$$Q_{\theta_0} = \sum_{i=1}^n \frac{g'(\eta_{i0})^2}{\sigma^2(\mu_{i0})} \Delta_i \Delta_i^T + \sum_{i=1}^n (Y_i - \mu_{i0}) \left[ \frac{g'(\eta_{i0})[\sigma^2(\mu_{i0})]'}{[\sigma^2(\mu_{i0})]^2} - \frac{g''(\eta_{i0})}{\sigma^2(\mu_{i0})} \right] \Delta_i \Delta_i^T.$$

Let  $\bar{\theta} \in \theta_n$ , where

$$\bar{\theta} = (\bar{\chi}_{j1}, \bar{\chi}_{j2}, \dots, \bar{\chi}_{jp}, \bar{\gamma}_1, \dots, \bar{\gamma}_q)^T, j = 1, 2, \dots, d,$$

and satisfies  $\bar{\eta}_i = \sum_{j=1}^d \sum_{k=1}^{m_j} \xi_{ijk} \bar{\chi}_{jk} + Z_i \bar{\gamma}$ ,  $\bar{\mu}_i = g(\bar{\eta}_i)$ . Similarly, we can obtain

$$Q_{\bar{\theta}} = \sum_{i=1}^n \frac{g'(\bar{\eta}_i)^2}{\sigma^2(\bar{\mu}_i)} \Delta_i \Delta_i^T + \sum_{i=1}^n (Y_i - \bar{\mu}_i) \left[ \frac{g'(\bar{\eta}_i) [\sigma^2(\bar{\mu}_i)]'}{[\sigma^2(\bar{\mu}_i)]^2} - \frac{g''(\bar{\eta}_i)}{\sigma^2(\bar{\mu}_i)} \right] \Delta_i \Delta_i^T.$$

(2) Solving  $U^*(\theta)$  given the link function is unknown by

$$U^*(\theta_0) = \sum_{i=1}^n (Y_i - \bar{\mu}_{i0}) \frac{\bar{g}'^{(\alpha)}(\eta_{i0})}{\sigma^2(\bar{\mu}_{i0})} \Delta_i = 0,$$

where  $\bar{\mu}_{i0} = \bar{g}^{(\alpha)}(\eta_{i0})$ . Similarly, we can obtain

$$Q_{\bar{\theta}}^* = \sum_{i=1}^n \frac{\bar{g}'^{(\alpha)}(\bar{\eta}_i)^2}{\sigma^2(\bar{\mu}_i^*)} \Delta_i \Delta_i^T + \sum_{i=1}^n (y_i - \bar{\mu}_i^*) \left[ \frac{\bar{g}'^{(\alpha)}(\bar{\eta}_i) [\sigma^2(\bar{\mu}_i^*)]'}{[\sigma^2(\bar{\mu}_i^*)]^2} - \frac{\bar{g}''^{(\alpha)}(\bar{\eta}_i)}{\sigma^2(\bar{\mu}_i^*)} \right] \Delta_i \Delta_i^T,$$

where  $\bar{\mu}_i^* = g^{(\alpha)}(\bar{\eta}_i)$ .

**Lemma 2.** *If the assumptions (C1)–(C5) hold, we have*

$$\sup_{\bar{\theta} \in \theta_n} \frac{1}{n} |Q_{\bar{\theta}}^* - Q_{\bar{\theta}}| = O_p(1).$$

**Proof.** Let  $M_i = \frac{1}{\sigma^2(\bar{\mu}_i^*)} - \frac{1}{\sigma^2(\bar{\mu}_i)}$ ,  $N_i = \frac{[\sigma^2(\bar{\mu}_i^*)]'}{[\sigma^2(\bar{\mu}_i^*)]^2} - \frac{[\sigma^2(\bar{\mu}_i)]'}{[\sigma^2(\bar{\mu}_i)]^2}$ , By Theorem 4.1 of Chiou and Müller [24], we know that  $\max_{1 \leq i \leq n} |M_i| = o_p(1)$  and  $\max_{1 \leq i \leq n} |N_i| = o_p(1)$ , then

$$\frac{1}{n} |Q_{\bar{\theta}}^* - Q_{\bar{\theta}}| = A + B + C, \tag{14}$$

where A, B, and C can be expressed as

$$\begin{aligned} A &= \frac{1}{n} \sum_{i=1}^n \left[ \frac{\bar{g}'^{(\alpha)}(\bar{\eta}_i)^2}{\sigma^2(\bar{\mu}_i^*)} - \frac{g'(\bar{\eta}_i)^2}{\sigma^2(\bar{\mu}_i)} \right] \Delta_i \Delta_i^T \leq \frac{1}{n} \max_{1 \leq i \leq n} |M_i|, \\ B &= \frac{1}{n} \sum_{i=1}^n (y_i - \bar{\mu}_i) \left[ \frac{\bar{g}'^{(\alpha)}(\bar{\eta}_i) [\sigma^2(\bar{\mu}_i^*)]'}{[\sigma^2(\bar{\mu}_i^*)]^2} - \frac{g'(\bar{\eta}_i) [\sigma^2(\bar{\mu}_i)]'}{[\sigma^2(\bar{\mu}_i)]^2} \right. \\ &\quad \left. - \left( \frac{\bar{g}''^{(\alpha)}(\bar{\eta}_i)}{\sigma^2(\bar{\mu}_i^*)} - \frac{g''(\bar{\eta}_i)}{\sigma^2(\bar{\mu}_i)} \right) \right] \Delta_i \Delta_i^T \\ &\leq \frac{1}{n} \sum_{i=1}^n (y_i - \bar{\mu}_i) \left( \max_{1 \leq i \leq n} |N_i| + \max_{1 \leq i \leq n} |M_i| \right), \\ C &= \frac{1}{n} \sum_{i=1}^n (\bar{\mu}_i^* - \bar{\mu}_i) \left[ \frac{\bar{g}'^{(\alpha)}(\bar{\eta}_i) [\sigma^2(\bar{\mu}_i^*)]'}{[\sigma^2(\bar{\mu}_i^*)]^2} - \frac{\bar{g}''^{(\alpha)}(\bar{\eta}_i)}{\sigma^2(\bar{\mu}_i^*)} \right] \Delta_i \Delta_i^T \\ &= o_p(1). \end{aligned}$$

Then, by (14) we can get

$$\frac{1}{n} |Q_{\bar{\theta}}^* - Q_{\bar{\theta}}| = o_p(1).$$

□

**Lemma 3.** *If the assumptions (C1)–(C5) hold, we have*

$$\left(\frac{1}{\sqrt{n}}\right)|U^*(\theta_0) - U(\theta_0)| = o_p(1).$$

**Proof.** Combining Theorem 1 and  $\max_{1 \leq i \leq n} |M_i| = o_p(1)$  in Lemma 2, we can prove that

$$\begin{aligned} \frac{1}{\sqrt{n}}(U^*(\theta_0) - U(\theta_0)) &= \frac{1}{\sqrt{n}} \left[ (Y_i - \mu_{i0})g'(\eta_{i0}) \left( \frac{1}{\sigma^2(\tilde{\mu}_{i0})} - \frac{1}{\sigma^2(\mu_{i0})} \right) \right. \\ &\quad \left. + \frac{(Y_i - \mu_{i0})}{\sigma^2(\tilde{\mu}_{i0})} (g'(\eta_{i0}) - \tilde{g}'^{(\alpha)}(\eta_{i0})) + \frac{\tilde{g}'^{(\alpha)}(\eta_{i0})}{\sigma^2(\tilde{\mu}_{i0})} (\tilde{\mu}_{i0} - \mu_{i0}) \right] \Delta_i \\ &= o_p(1). \end{aligned}$$

□

**Theorem 2.** *If we assume that (C1)–(C5) hold, we have*

$$\begin{pmatrix} \frac{nd_G^2(\hat{\beta}_1, \beta_1) - m_1}{\sqrt{2m_1}} \\ \vdots \\ \frac{nd_G^2(\hat{\beta}_d, \beta_d) - m_d}{\sqrt{2m_d}} \\ \sqrt{no_1}(\gamma_1 - \hat{\gamma}_1) \\ \vdots \\ \sqrt{no_q}(\gamma_q - \hat{\gamma}_q) \end{pmatrix} \xrightarrow{d} N(0, I),$$

In the case of truncated models for  $m_j$ , let  $\hat{\chi}_j$  be the estimator of  $\chi_j = (\chi_{j1}, \chi_{j2}, \dots, \chi_{jm_j})^T$ ,  $\tilde{\Lambda}_j = (\lambda_{j,k_1k_2})$ ,  $1 \leq k_1, k_2 \leq m_j$ , where  $\lambda_{j,k_1k_2} = E \left[ \frac{g'(\eta)^2}{\sigma^2(\mu)} \xi_{jk_1} \xi_{jk_2} \right]$ . We define  $\tilde{\chi}_j = (\chi_{j(m_j+1)}, \chi_{j(m_j+2)}, \dots)^T$ . Therefore, we have the following expression:

$$d_G^2(\hat{\beta}_j, \beta_j) = (\hat{\chi}_j - \chi_j)^T \tilde{\Lambda}_j (\hat{\chi}_j - \chi_j) + \sum_{k_1, k_2 = m_j}^{\infty} \lambda_{j,k_1k_2} \tilde{\chi}_j^2, j = 1, \dots, d.$$

Furthermore, let  $o_l = E \left[ \frac{g'(\eta)^2}{\sigma^2(\mu)} z_{il}^2 \right]$ , where  $l = 1, \dots, q$ . Here,  $I$  represents a  $(q + \sum_{j=1}^d m_j) \times (q + \sum_{j=1}^d m_j)$  dimensional identity matrix.

**Proof.** By using the Taylor expansion with a suitable mean value  $\bar{\theta}$ , we can obtain

$$U^*(\hat{\theta}) = U^*(\theta_0) - Q_{\bar{\theta}}^*(\hat{\theta} - \theta_0) = 0. \tag{15}$$

Then, by Lemmas 2 and 3, (15) can be deformed as

$$U^*(\hat{\theta}) = U(\theta_0) - Q_{\bar{\theta}}(\hat{\theta} - \theta_0) + o_p(\sqrt{n}) = 0.$$

Then, we can get

$$\hat{\theta} - \theta_0 = Q_{\bar{\theta}}^{-1}U(\theta_0) + o_p\left(\frac{1}{\sqrt{n}}\right).$$

By combining the above equation with

$$U(\tilde{\theta}^{(\alpha)}) = U(\theta_0) - Q_{\tilde{\theta}}(\tilde{\theta}^{(\alpha)} - \theta_0) = 0,$$

we can get

$$\sqrt{n}(\hat{\theta} - \theta_0) = \sqrt{n}(\tilde{\theta}^{(\alpha)} - \theta_0) + o_p(1). \tag{16}$$

By (16), it can be seen that it transforms the relationship between  $\hat{\theta}$  and  $\theta_0$  in the case of unknown link functions into the relationship between  $\tilde{\theta}^{(\alpha)}$  and  $\theta_0$  in the case of known link functions, and then combined with Theorem 1 in [25], the proof of Theorem 2 can be obtained.  $\square$

### 4.3. Asymptotic Convergence of $\hat{g}$

**Theorem 3.** *If we assume that (C1)-(C5) hold, for  $\sigma^2 > 0$ , then we have*

$$\sqrt{nh}[\hat{g}(u; \hat{\theta}) - g(u) - I(u)] \xrightarrow{D} N(0, \vartheta^2(u)).$$

**Proof.**

$$\begin{aligned} \sqrt{nh}[\hat{g}(u; \hat{\theta}) - g(u) - I(u)] &= \sqrt{nh}[\hat{g}(u; \hat{\theta}) - \hat{g}(u; \theta) + \tilde{g}^{(\alpha)}(u; \theta) \\ &\quad - g(u) - I(u)] \\ &\leq \sqrt{nh}[\hat{g}'(u; \theta)|\hat{\theta} - \theta|] \\ &\quad + \sqrt{nh}[\tilde{g}^{(\alpha)}(u; \theta) - g(u) - I(u)] \\ &= \sqrt{nh}[\tilde{g}^{(\alpha)}(u; \theta) - g(u) - I(u)] + o_p(1). \end{aligned}$$

The above expression transforms the relationship between  $\hat{g}$  and  $g$  into the relationship between  $\tilde{g}^{(\alpha)}$  and  $g$  (i.e., Theorem 1). Therefore, by Theorem 1, we can get Theorem 3.  $\square$

**Corollary 2.** *If we further refine the condition in assumption (C4) such that  $nh^5 \rightarrow 0$ , then it follows that*

$$\sqrt{nh}[\hat{g}(u; \hat{\theta}) - g(u)] \xrightarrow{D} N(0, \vartheta^2(u)).$$

**Remark 2.** *Let  $(e_1, \lambda_1), (e_2, \lambda_2), \dots, (e_{m_j}, \lambda_{m_j})$  represent the eigenvalues and eigenvectors of  $\Omega$ , where*

$$e_k = (e_{j1}, \dots, e_{jm_j}), w_k(t) = \sum_{j=1}^d \rho_{jk}(t)e_{jk}, k = 1, 2, \dots, m_j,$$

$$\Omega = \frac{1}{n} E \left( \frac{\hat{g}'(\hat{\eta}_i)^2}{\sigma^2(\hat{\mu}_i)} D_j^T D_j \right), i = 1, \dots, n, j = 1, \dots, d.$$

*Then, the 95% confidence band for the regression coefficient function  $\hat{\beta}_j(t)$  can be expressed as*

$$\hat{\beta}_j(t) \pm \sqrt{r(\alpha) \sum_{k=1}^{m_j} \frac{w_k(t)^2}{e_k}},$$

where  $r(\alpha) = [m_j + \sqrt{2m_j}\Phi(1 - \alpha)], \alpha = 0.05, \Phi(1 - \alpha) = 1.96$ .

## 5. Simulation

We consider a binary response and two functional predictors as well as three scalar predictors. The functional predictors  $X_{i1}(t)$  and  $X_{i2}(t)$  ( $i = 1, \dots, n$ ) are observed at 50 equal distant time points on the interval  $[0, 1]$ .

The sample sizes are  $n = 50, 100, 300$ . Let the score coefficients  $\xi_{ijk}$  for each functional predictor satisfy the following assumptions:

$$\xi_{i1k} \sim N(0, \lambda_{1k}), k = 1, 2, 3, 4,$$

where  $\lambda_{11} = 1, \lambda_{12} = \sqrt{2}/2, \lambda_{13} = 1/2, \lambda_{14} = \sqrt{2}/4$ .

$$\xi_{i2k} \sim N(0, \lambda_{2k}), k = 1, 2, 3,$$

where  $\lambda_{21} = 1, \lambda_{22} = \sqrt{2}/2, \lambda_{23} = 1/2$ .

We define the orthonormal basis functions  $\rho_{1k}(t)$  and  $\rho_{2k}(t), t \in [0, 1]$ , which satisfy

$$\rho_{1k}(t) = \sqrt{2}\sin(2k\pi t), k = 1, 2, 3, 4,$$

$$\rho_{2k}(t) = \sqrt{2}\cos(2k\pi t), k = 1, 2, 3.$$

Then,  $X_{ij}(t)$  can be represented through Karhunen–Loeve expansion as follows:

$$X_{i1}(t) = \sum_{k=1}^4 \xi_{i1k}\rho_{1k}(t),$$

$$X_{i2}(t) = \sum_{k=1}^3 \xi_{i2k}\rho_{2k}(t).$$

Figure 1 shows the 50 trajectories of the two functional predictors  $X_1(t)$  and  $X_2(t)$ .

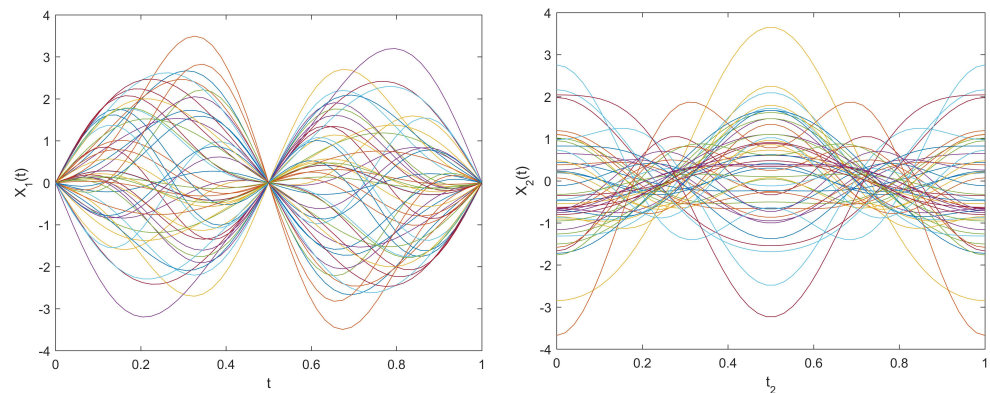


Figure 1. The predictors  $X_1(t)$  and  $X_2(t)$ .

The scalar predictor  $Z = (Z_1, Z_2, Z_3)^T$  satisfies the following assumption

$$Z_1 \sim N(0, 1), Z_2 \sim N(0, \frac{\sqrt{3}}{3}), Z_3 \sim N(0, \frac{\sqrt{5}}{5}).$$

We assume that the regression coefficient functions of the functional predictors satisfy the following assumption

$$\beta_1(t) = \sum_{k=1}^4 \chi_{1k}\rho_{1k}(t),$$

$$\beta_2(t) = \sum_{k=1}^3 \chi_{2k}\rho_{2k}(t),$$

where  $\chi_{1k} = \sqrt{\frac{1}{3k}}, k = 1, 2, 3, 4$  and  $\chi_{2k} = \sqrt{\frac{1}{3k}}, k = 1, 2, 3$ . Moreover, we assume that the regression coefficients  $\gamma = (\gamma_1, \gamma_2, \gamma_3)^T$  of the scalar predictors satisfy  $\gamma_1 = \sqrt{2}/2, \gamma_2 = \sqrt{3}/3, \gamma_3 = 1/2$ .

Define

$$P(X, Z) = g\left(\sum_{j=1}^2 \int_T X_j(t)\beta_j(t) + Z^T\gamma\right).$$

And we select the link function as

$$g(x) = \frac{\exp(x)}{1 + \exp(x)}.$$

We generate binary response

$$Y(X, Z) \sim \text{Binomial}(P(X, Z), 1)$$

as pseudo random sequence.

We obtain a sample

$$(Y_i, X_{i1}(t), X_{i2}(t), Z_i), i = 1, \dots, n,$$

where  $n$  is the sample size. The number of functional principal components that explain 85% of cumulative variation contribution are  $m_1 = 3, 3, 4, m_2 = 2, 3, 3$ , respectively. We run 100 simulations.

Figure 2 shows the asymptotic behavior of the link function under different sample sizes. The black lines in Figure 2 shows the relationship between  $\eta$  and  $\mu$ , where

$$\eta = \sum_{j=1}^2 \int_T X_j(t)\beta_j(t)dt + Z^T\gamma, \mu = g(\eta) = \frac{\exp(\eta)}{1 + \exp(\eta)} \in [0, 1].$$

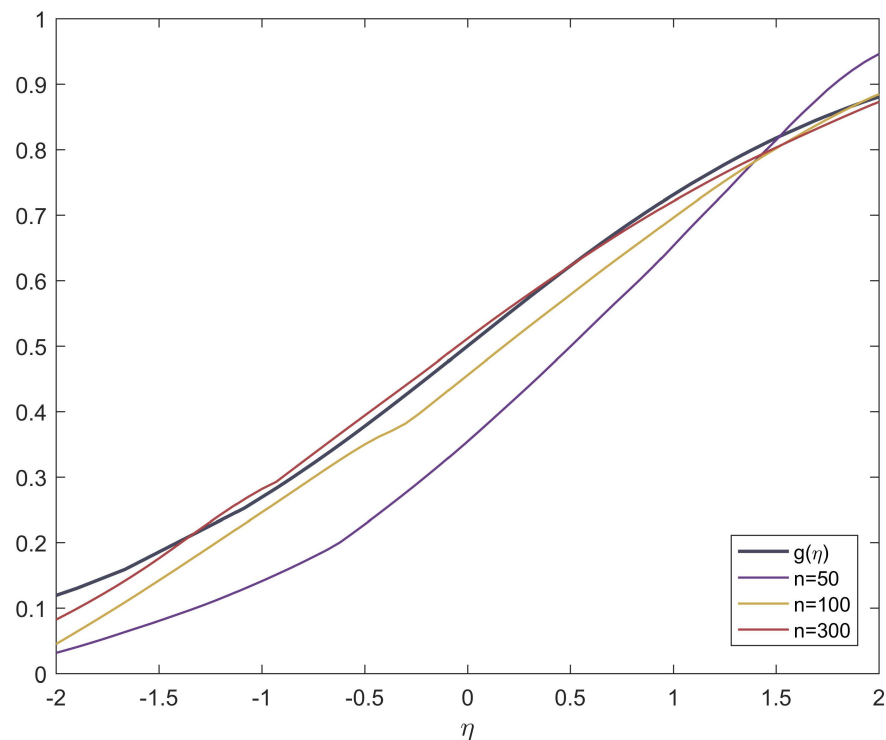
The additional colored lines shown in Figure 2 represent the estimated link function  $\hat{g}$  for different sample sizes. These lines are obtained through iterative processes, starting with an initial value of  $g$  set to  $g(\eta) = \eta$ . The iterative process continues until one of the following conditions is met: 100 iterations have been performed, or the error in the regression coefficients is less than 0.01. The purpose of these lines is to illustrate the relationship between  $\hat{\eta}$  and  $\hat{\mu}$ , where

$$\hat{\eta} = \sum_{j=1}^2 \int_T X_j(t)\hat{\beta}_j(t)dt + Z^T\hat{\gamma}, \hat{\mu} = \hat{g}(\hat{\eta}) \in [0, 1].$$

Since in this case, both  $\hat{\eta}$  and  $\eta$  are in  $[-2, 2]$ , we denote the argument of  $g$  and  $\hat{g}$  by  $\eta$ , and the x-axis in Figure 2 is denoted by  $\eta$  and is shown in the interval  $[0, 1]$ . Table 2 presents the estimates of  $\hat{g}$  evaluated through RMISE under different sample sizes. The RMISE is defined as follows:

$$RMISE = \sqrt{\frac{1}{Q} \int_{-2}^2 (\hat{g}(\eta) - g(\eta))^2 d\eta},$$

where  $Q = 100$  is the number of simulations here. In summary, Figure 2 and Table 2 demonstrate that as the sample size increases, the estimated link function  $\hat{g}$  becomes closer and closer to the true link function  $g$ .



**Figure 2.** Asymptotic properties of the link function  $g$ . The black line in the graph represents the true link function  $g = \exp(\eta) / (1 + \exp(\eta))$ . The purple, yellow, and red lines in the graph represent the estimated link functions  $\hat{g}$  under sample sizes of  $n = 50$ ,  $n = 100$ , and  $n = 300$ , respectively.

**Table 2.** RMISE of  $g$  and  $\hat{g}$  for different sample size  $n$ .

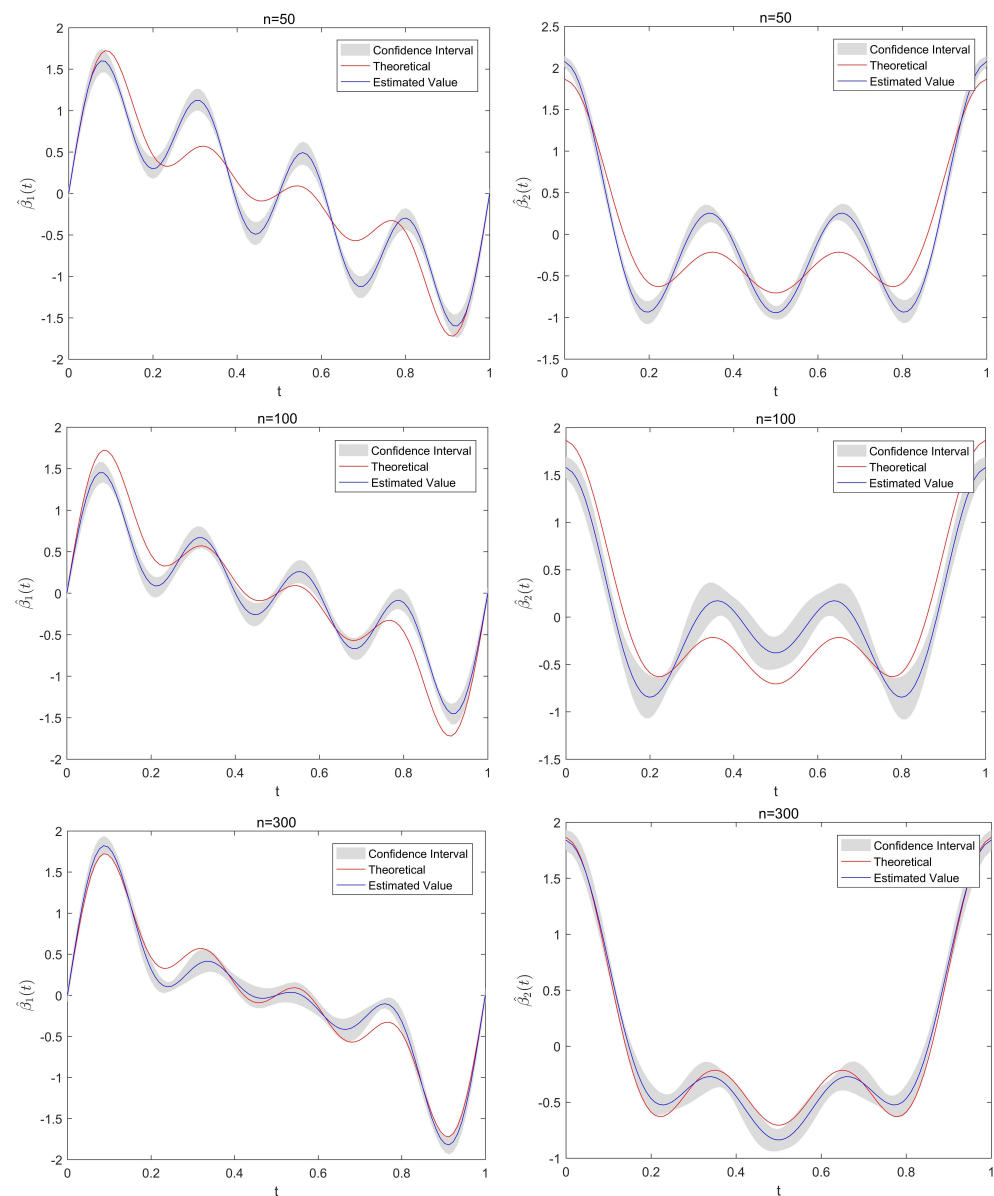
$n$	RMISE
50	0.3540
100	0.2734
300	0.1449

In Table 3, it can be seen that both the SD and RMISE of the estimated regression coefficient functions  $\hat{\beta}_1(t)$  and  $\hat{\beta}_2(t)$  decrease as the sample size  $n$  increases.

**Table 3.** SD and RMISE of the estimated values of  $\hat{\beta}_1(t)$  and  $\hat{\beta}_2(t)$  for different sample sizes  $n$ .

	$n$	SD	RMISE
$\hat{\beta}_1(t)$	50	0.2475	0.3405
	100	0.1344	0.2517
	300	0.0552	0.1204
$\hat{\beta}_2(t)$	50	0.2536	0.3232
	100	0.1261	0.2863
	300	0.0239	0.1033

Figure 3 displays the estimated functional regression coefficients  $\hat{\beta}_1(t)$  and  $\hat{\beta}_2(t)$ , as well as their 95% confidence intervals under different sample sizes. The red curve in the figure represents the theoretical values of  $\beta_1(t)$  and  $\beta_2(t)$ , while the blue curve represents the estimated values  $\hat{\beta}_1(t)$  and  $\hat{\beta}_2(t)$ . The gray shaded area represents the 95% confidence interval of the estimates. It can be seen that as the sample size increases, the estimated values become closer to the true values.



**Figure 3.** Estimated values of regression coefficient function  $\hat{\beta}_1(t)$ ,  $\hat{\beta}_2(t)$  (blue curves) and their 95% confidence intervals (grey area) for difference sample size, where the red curves are the theoretical regression coefficient functions  $\beta_1(t)$ ,  $\beta_2(t)$ .

Table 4 presents the estimated scalar regression coefficient  $\hat{\gamma}$  and corresponding standard deviation under different sample sizes. It can be seen that as the sample size  $n$  increases,  $\hat{\gamma} = (\hat{\gamma}_1, \hat{\gamma}_2, \hat{\gamma}_3)^T$  becomes closer to the true values  $\gamma = (\sqrt{2}/2, \sqrt{3}/3, 1/2)^T$ . Moreover, as the sample size  $n$  increases, the SD becomes smaller, indicating that the estimated values have more certainty.

**Table 4.** Estimated values of scalar regression coefficients  $\hat{\gamma}$  and their SD in brackets for different sample sizes  $n$ .

$n$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$
50	0.7298 (0.191)	0.5928 (0.177)	0.5307 (0.232)
100	0.6892 (0.092)	0.5832 (0.071)	0.4894 (0.096)
300	0.7105 (0.019)	0.5732 (0.018)	0.4988 (0.016)



Table 5 presents the M1 and M2 values for different sample sizes, where  $M1 = \frac{1}{Q} \sum_{i=1}^Q MAE$ ,  $MAE = \frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i|$ ,  $M2 = \frac{1}{Q} \sum_{i=1}^Q MSE$ ,  $MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$ , and  $Y_i$  and  $\hat{Y}_i$  represent the real and the predicted values of the response variable, respectively. We can find that as the sample size increases, the values of M1 and M2 become smaller, indicating that the predictive performance of the model improves.

**Table 5.** The M1 and M2 values for different sample sizes  $n$ .

$n$	M1	M2
50	0.3182	0.1579
100	0.3028	0.1498
300	0.2921	0.1406

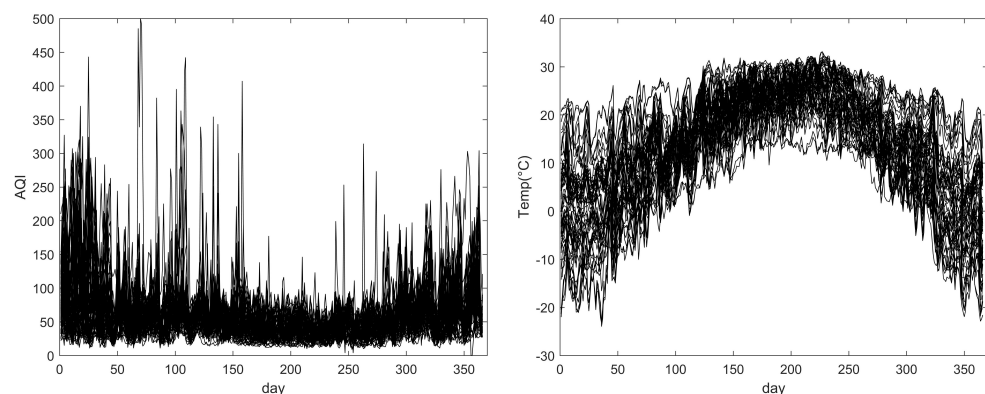
## 6. Application

As is well known, research on average life expectancy is crucial for social development, health policies, and population management. Studies on average life expectancy can help governments, health departments, and social institutions develop relevant policies and plans to improve people's quality of life and health conditions. By understanding people's life expectancy, the efficiency of healthcare systems and the effectiveness of social welfare and public health policies can be evaluated, providing a basis for resource allocation and planning. Additionally, research on average life expectancy can also help people understand population structure and trends, providing references for social-economic development, pension systems, and labor market planning. Therefore, in the application of our proposed model, we investigate factors that influence average life expectancy, including air quality index (AQI), temperature, GDP, and number of beds in hospitals.

### 6.1. Data Description

We collected average daily temperature (Temp) data for 58 cities in China in 2020 from the National Meteorological Science Data Sharing Service Platform, and average daily Air Quality Index (AQI) data from the National Environmental Monitoring Station. We also collected GDP, number of beds in hospitals, and life expectancy data for each city from local statistical bulletins and government documents. Among them, there are two functional predictive variables, which are daily AQI and temperature from 1 January to 31 December 2020, for 366 days in 58 cities. There are also two scalar predictive variables, which are GDP and number of beds in hospitals for the 58 cities in 2020. The response variable is the life expectancy of residents in each city in 2020.

Figure 4 shows the daily AQI and temperature for 58 cities in 2020.



**Figure 4.** Daily AQI (left plot) and daily temperatures (right plot) for 58 cities in 2020; each curve represents one city.

### 6.2. Data Analysis

According to a report released by the National Health Commission, the average life expectancy of Chinese residents in 2020 was 77.9 years. Therefore, we divide the response variable as follows: when the life expectancy of a city is greater than 77.9 years, we represent it as 1; otherwise, when the life expectancy is less than 77.9 years, we represent it as 0. For the functional predictors, we first centralize the data. Second, we conduct FPCA and select the number of functional principal components that explain 75% of the variation. The number of components for AQI and temperature is  $p_{AQI} = 10$  and  $p_{Temp} = 3$ , respectively. We use GCV to demonstrate the predictive accuracy of the estimators. In this application,  $GCV = 0.135$ .

### 6.3. Results Analysis

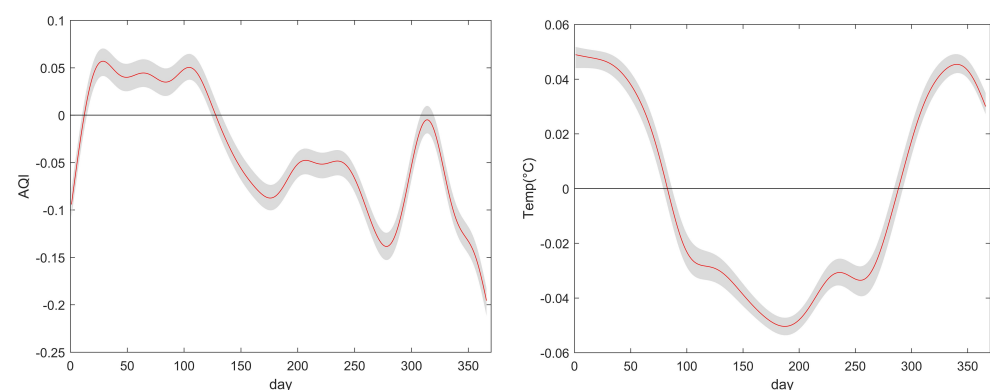
By inputting the data into the generalized partially functional linear model, we obtain the regression coefficient function  $\hat{\beta}(t)$  for the functional predictors and the regression coefficients  $\hat{\gamma}$  for the scalar predictors. The results are shown in Table 6 and Figure 5, respectively.

Table 6 presents the estimated values of the regression coefficient  $\hat{\gamma}$  for scalar predictor variables. We can see that both GDP and number of beds in hospitals have a positive relationship with life expectancy, and are significant at the 5% level. This means that when a region has a higher GDP and more hospital beds, the life expectancy in that region is longer. In other words, the better the economic development and medical resources of a region, the longer the life expectancy.

**Table 6.** Regression coefficients  $\hat{\gamma}$  and their significance levels.

	Estimate	Std.Error	t Value	Pr(> t )
$\hat{\gamma}_{GDP}$	0.6776	0.339	1.9988	0.04639
$\hat{\gamma}_{Beds}$	0.7354	0.367	2.0038	0.04585

In Figure 5, we see the estimated values of the regression coefficient function  $\hat{\beta}(t)$ . For AQI, we can find a negative relationship between AQI and life expectancy in general. The higher the value of AQI, the more serious the air pollution is, and the lower the life expectancy corresponding to it. However, there is a more obvious positive relationship trend in February to April, which may be influenced by some other external factors. For temperature, we can find that the effect of temperature on life expectancy varies with the change of seasons. In spring, summer, and fall (March to October), the effect of temperature on life expectancy is negatively correlated, and in winter (November to February), it is positively correlated, which is consistent with the conclusion in Huang et al. [19].



**Figure 5.** Estimated values of regression coefficient function  $\hat{\beta}(t)$  and their 95% confidence intervals.

To confirm the necessity of considering the unknown link function model, we choose models without a link function and with the logit link function (i.e.,  $g(\eta) = \frac{e^\eta}{1+e^\eta}$ ) and compare them with our proposed models with unknown link functions. In order to evaluate the prediction performance of the three models, we use MAE, MSE, and  $R^2$ . Additionally, we calculate the accuracy using the confusion matrix, where we define TP as the number of samples correctly classified as positive, TN as the number of samples correctly classified as negative, FP as the number of samples incorrectly classified as positive, and FN as the number of samples incorrectly classified as negative (missed detections). We obtain the model's accuracy using the formula

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}.$$

When the values of MAE and MSE are smaller, it indicates that the model has a smaller prediction error and better performance. When  $R^2$  is closer to 1, it indicates that the model has a stronger ability to explain the response variable. The experimental results are shown in Table 7. It can be seen that the model we proposed has the best performance.

**Table 7.** Comparison between Unknown Link Function Model, Logit Link Function Model, and Model without a Link Function.

Link Function	MAE	MSE	$R^2$	Accuracy
Unknown	0.2584	0.1399	0.8916	81.03%
Logit	0.2872	0.2511	0.6673	75.86%
Without	0.4777	0.3146	0.4118	74.14%

## 7. Conclusions

This article proposes a generalized partially functional linear model for scalar response and predictor variables that include both functional and scalar components, without specifying a link function. We use functional principal component analysis to reduce the dimensionality of functional data, estimate the regression coefficients using the maximum likelihood estimation method, estimate the link function using the method of local linear regression, iteratively obtain the final estimator, and establish the asymptotic normality of the estimator. The accuracy of the proposed model is validated through simulation studies.

The article applies the proposed model to the study of average life expectancy. Using daily AQI, temperature, GDP, and number of beds in hospitals for 58 cities in China in 2020, the study explores the impact of environmental, economic, and medical factors on life expectancy. The results indicate that GDP and number of beds in hospitals have a positive correlation with the life expectancy, while the AQI has an overall negative correlation. Temperature has a negative correlation with the average life expectancy in spring, summer, and autumn, and a positive correlation in winter. Overall, the study concludes that the average life expectancy is higher in areas with better environmental, economic, and medical development.

This model can be used in various fields, including economics, bio-medicine, engineering, etc. However, this model still has certain limitations. For example, the relationship between air quality and temperature needs to be further considered. There is a certain correlation between temperature and air quality. Generally, an increase in temperature can lead to the intensified volatilization and diffusion of pollutants in the air, thereby causing a decline in air quality. In the next phase of research, we will consider the interactions between functional predictors to make results more accurate. In addition, the algorithms and optimization methods of the model can be further improved to enhance computational efficiency. Combining this model with other machine learning methods can further improve predictive performance.

**Author Contributions:** W.X.: methodology, software, validation, writing—review, supervision, funding acquisition. S.L.: methodology, software, data curation, writing—original draft. H.L.: writing—review, supervision. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research is supported by the Yujie Talent Project of North China University of Technology (Grant No. 107051360023XN075-04).

**Data Availability Statement:** The original data supporting the results of this study can be obtained from the National Meteorological Science Data Sharing Service Platform, the National Environmental Monitoring Station, and local statistical bulletins.

**Acknowledgments:** The authors would like to thank the referees and the editor for their useful suggestions, which helped us improve the paper.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Ramsay, J.O. When the data are functions. *Psychometrika* **1982**, *47*, 379–396. [[CrossRef](#)]
2. Ramsay, J.O.; Silverman, B.W. *Functional Data Analysis*, 2nd ed.; Springer: New York, NY, USA, 2005.
3. Horváth, L.; Kokoszka, P. *Inference for Functional Data with Application*; Springer: New York, NY, USA, 2012.
4. Shin, H. Partial functional linear regression. *J. Stat. Plan. Inference* **2009**, *139*, 3405–3418. [[CrossRef](#)]
5. Shin, H.; Lee, M.H. On prediction rate in partial functional linear regression. *J. Multivar. Anal.* **2012**, *103*, 93–106. [[CrossRef](#)]
6. James, G.M. Generalized linear models with functional predictors. *J. R. Stat. Soc. Ser. B* **2002**, *64*, 411–432. [[CrossRef](#)]
7. Müller, H.G.; Stadtmüller, U. Generalized functional linear models. *Ann. Stat.* **2005**, *33*, 774–805. [[CrossRef](#)]
8. Shang, Z.F.; Cheng, G. Nonparametric inference in generalized functional linear models. *Ann. Stat.* **2015**, *43*, 1742–1773. [[CrossRef](#)]
9. Wong, R.K.W.; Li, Y.; Zhu, Z.Y. Partially Linear Additive Models for Multivariate Functional Data. *J. Am. Stat. Assoc.* **2019**, *114*, 406–418. [[CrossRef](#)]
10. Scallan, A.; Gilchrist, R.; Green, M. Fitting Parametric Link Functions in Generalized Linear Models. *Comput. Stat. Data Anal.* **1984**, *2*, 37–49. [[CrossRef](#)]
11. Weisberg, S.; Welsh, A.H. Adapting for the missing link. *Ann. Stat.* **1994**, *22*, 1674–1700. [[CrossRef](#)]
12. Chiou, J.M.; Müller, H.G. Quasi-likelihood regression with unknown link and variance functions. *J. Am. Stat. Assoc.* **1998**, *93*, 1376–1387. [[CrossRef](#)]
13. Chiou, J.M.; Müller, H.G. Estimated estimating equations: Semiparametric inference for clustered and longitudinal data. *J. R. Stat. Soc. Ser. B* **2005**, *67*, 531–553. [[CrossRef](#)]
14. Bai, Y.; Fung, W.K.; Zhu, Z.Y. Penalized quadratic inference functions for single-index models with longitudinal data. *J. Multivar. Anal.* **2009**, *100*, 152–161. [[CrossRef](#)]
15. Pang, Z.; Xue, L. Estimation for the single-index models with random effects. *Comput. Stat. Data Anal.* **2012**, *56*, 1837–1853. [[CrossRef](#)]
16. Yuan, M.; Diao, G. Sieve maximum likelihood estimation in generalized linear models with an unknown link function. *Wiley Interdiscip. Rev. Comput. Stat.* **2017**, *10*, e1425. [[CrossRef](#)]
17. Kokoszka, P.; Reimherr, M. *Introduction to Functional Data Analysis*; CRC Press: Boca Raton, FL, USA, 2017. [[CrossRef](#)]
18. Rao, A.R.; Reimherr, M. Nonlinear Functional Modeling Using Neural Networks. *J. Comput. Graph. Stat.* **2023**, *32*, 1248–1257. [[CrossRef](#)]
19. Huang, C.; Barnett, A.G.; Wang, X.; Tong, S. The impact of temperature on years of life lost in Brisbane, Australia. *Nat. Clim. Chang.* **2012**, *2*, 265–270. [[CrossRef](#)]
20. Yang, Y.; Qi, J.L.; Ruan, Z.L.; Yin, P.; Zhang, S.Y.; Liu, J.M.; Liu, Y.N.; Li, R.; Wang, L.J.; Lin, H.L. Changes in Life Expectancy of Respiratory Diseases from Attaining Daily PM2.5 Standard in China: A Nationwide Observational Study. *Innovation* **2020**, *1*, 100064. [[CrossRef](#)] [[PubMed](#)]
21. Deryugina, T.; Molitor, D. The Causal Effects of Place on Health and Longevity. *J. Econ. Perspect.* **2021**, *35*, 147–170. [[CrossRef](#)]
22. Mack, Y.P.; Silverman, B.W. Weak and strong uniform consistency of kernel regression estimates. *Probab. Theory Relat. Fields* **1982**, *63*, 405–415. [[CrossRef](#)]
23. Masry, E.; Tjøstheim, D. Estimation and Identification of Nonlinear ARCH Time Series: Strong Convergence and Asymptotic Normality. *Econom. Theory* **1995**, *11*, 258–289. [[CrossRef](#)]
24. Chiou, J.M.; Müller, H.G. Nonparametric quasi-likelihood. *Ann. Stat.* **1999**, *27*, 36–64. [[CrossRef](#)]
25. Xiao, W.W.; Wang, Y.X.; Liu, H.Y. Generalized partially functional linear model. *Sci. Rep.* **2021**, *11*, 23428. [[CrossRef](#)] [[PubMed](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.