

## Article

# Generalized Partially Functional Linear Model with Interaction between Functional Predictors

Weiwei Xiao <sup>1,\*</sup> , Kejing Mao <sup>1</sup> and Haiyan Liu <sup>2</sup> <sup>1</sup> School of Science, North China University of Technology, Beijing 100144, China; kjmao0312@163.com<sup>2</sup> Department of Statistics, University of Leeds, Leeds LS2 9JT, UK; h.liu1@leeds.ac.uk

\* Correspondence: xiaoww@ncut.edu.cn

**Abstract:** This paper proposes a generalized partially functional linear model with interaction terms. It is suitable for cases where the response variable is scalar, and the predictor variables include a mix of functional and scalar types, while considering the correlations among functional predictor variables. The model uses principal component analysis for dimensionality reduction, employs maximum likelihood estimation to obtain parameter values, proves the asymptotic properties of the estimates, and validates the model's accuracy through data simulation experiments. Finally, the proposed model was applied to investigate the influence of air quality, climate factors, and medical and social indicators, along with their interactions, on cancer incidence, which is a binary response.

**Keywords:** functional data analysis; generalized functional linear model; interaction term; cancer incidence

MSC: 00A71; 62H25



**Citation:** Xiao, W.; Mao, K.; Liu, H. Generalized Partially Functional Linear Model with Interaction between Functional Predictors. *Axioms* **2024**, *13*, 583. <https://doi.org/10.3390/axioms13090583>

Academic Editors: Christos Floros, Christos Kountzakis and Konstantinos Gkillas

Received: 10 July 2024

Revised: 15 August 2024

Accepted: 19 August 2024

Published: 27 August 2024



**Copyright:** © 2024 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

With the advent of the big data era, more and more functional data, providing information about objects varying over a continuum, are collected.

Currently, functional data analysis is being applied in various fields such as medicine, environmental science, and economics and is receiving increasing attention. For details on functional data analysis, see monographs by Ramsay and Silverman [1], Horváth and Kokoszka [2], and Hsing and Eubank [3].

Several variants of functional linear regression models have been proposed to investigate the influence of functional and/or scalar predictors on functional or scalar response and, therefore, to make predictions. Cardot [4], Tony [5], and others have utilized spline methods for estimation and prediction in functional linear regression models. In 2007, Cardot et al. [6] extended the population least squares method to functional linear models, proposing smooth spline estimates for model function coefficients and providing asymptotic results for this estimation. In 2012, Delaigle and Hall [7] utilized partial least squares to demonstrate the consistency and convergence of functional linear models. Tony and Ming [8] studied the estimation and prediction issues of functional linear regression models within the framework of reproducing kernel Hilbert spaces. Nevertheless, these models cannot deal with general responses such as binary and Poisson.

An important tool for functional data analysis is the functional linear regression model, while the generalized functional regression model is an extension of the functional linear regression model. As research progressed, the generalized functional regression model was introduced to handle more complex response variables. This model was first introduced by Nelder and Wedderburn [9] in 1972, and it investigates the relationship between continuous and discrete response variables and the predictor variables through a link function. In 2002, James [10] proposed generalized linear models with functional predictors and applied it to

standard missing data problems. In 2005, Müller and Stadtmüller [11] proposed a generalized functional linear regression model where the response variable is a discrete scalar and the predictor is functional. In 2011, Goldsmith et al. [12] developed fast fitting methods for generalized functional linear models that can be applied to various functional data designs, including functions measured with and without error and sparsely or densely sampled. In 2021, Xiao et al. [13] proposed a generalized partially functional linear regression model where the response variable is general and the predictors are scalar and functional. However, none of these models incorporate the interaction of functional predictors.

To better address complex data that include both functional predictors and scalar predictors, scholars have improved the functional linear model and proposed a functional regression model with mixed predictors. In 2016, Kong et al. [14] explored the estimation and variable selection problems in cases where the parametric part is high-dimensional and the functional predictors are multidimensional. Yao [15] and Ma et al. [16] further built upon the work of Kong [14], conducting more in-depth research and proving the large sample properties of the estimators. In 2020, Xu et al. [17] studied the estimation and hypothesis testing issues for models with multiple functional predictors and demonstrated the corresponding large sample properties.

In many practical applications, we need to consider the interactions between variables, and failure to consider the interaction term may lead to the problem of missing variables in the model, thus introducing inaccurate predictions and inappropriate interpretations. By introducing interaction terms, the inaccuracy can be reduced and the model can be made more reliable, thereby improving the prediction by the model and providing more reliable decision support. Indeed, functional linear regressions models with interaction between functional predictors have been proposed recently; several examples follow. In 2016, Usset et al. [18] proposed a functional regression model with a scalar response and multiple functional predictors with two-way interactions in addition to their main effects. In 2019, Luo and Qin [19] proposed function-on-function regression models with interaction and quadratic effects, together with an efficient estimation method that has a minimum prediction error. In 2013, Yang et al. [20] introduced a class of nonlinear multivariate time-frequency functional models that can identify important features in each signal as well as the interaction of signals. Some models considered the interaction of two different time points in the functional data. In 2020, Matsui [21] proposed a functional quadratic model which took the interaction between two different time points of the functional data into consideration. In 2020, Sun and Wang [22] also considered a quadratic regression model where the predictor and the response are both functional; it estimated predictions for the coefficient functions, and unknown responses and asymptotics were demonstrated. Nonetheless, these models cannot be applied to general scalar responses. As far as we know, only Fuchs et al. [23] in 2015 considered general scalar response with functional predictors to include linear functional interaction terms. However, one drawback of the method of Fuchs et al. [23] is that scalar predictors are not included, and a second drawback is that the asymptotic properties of estimated regression coefficients were not established.

A practical motivation of this paper is the investigation of the influence of air qualities, climate factors, medical and social indicators, and their interactions on cancer incidence, which is a binary response. Cancer is one of the leading causes of death in humans; therefore, it is crucial to analyze the factors related to cancer incidence. Studying cancer incidence can help improve public health and quality of life, reduce social medical costs, and promote human health and socio-economic development. In 2022, Qiu et al. [24] pointed out that cancer incidence in China is much higher than those in the United States and the United Kingdom due to the fact that China faces problems such as a large population, uneven development in various regions, and a relative lag in cancer control strategies. In 2014, Qin et al. [25] indicated that long-term exposure to air pollutants or short-term exposure to some high concentrations of air pollutants such as PM<sub>2.5</sub> may be associated with some increased incidence rates of overall cancer, especially prostate cancer and female breast cancer. In 2022, Wu et al. [26] found that areas with high green coverage have a lower

risk of cancer. In 2023, Cao et al. [27] analyzed the relationship between per capita GDP and cancer incidence in 55 regions of China, showing that regions with high GDP have high cancer incidence. In 2017, Xu et al. [28] conducted a statistical analysis of the current situation of PM2.5 in Changzhou in China and considered an interaction between PM2.5 and relative humidity during the same period, indicating a certain degree of interaction between the two. In 2022, Yang et al. [29] used the generalized linear model to study the effects of PM2.5 and relative humidity on visibility and found a significant interaction between PM2.5 and relative humidity.

Therefore, we collected data on average daily PM2.5 concentration (from 1 January 2015 to 31 December 2020), average daily humidity (from 1 January 2015 to 31 December 2020), per capita GDP, green coverage rate in built-up areas, the proportion of medical personnel (PMP) (which is the ratio of the number of licensed (assistant) doctors to the population in the locality), and the binary cancer incidence in 49 cities in China from <http://www.cnemc.cn/>, <http://www.stats.gov.cn/sj/ndsjs/> and <http://www.chinancpcn.org.cn/home>. Our aim was to investigate the influence of PM2.5 concentration, air humidity, per capita GDP, green coverage, and PMP on cancer incidence, with the focus not only on the main effects but also on the interaction between PM2.5 concentration and air humidity to, therefore, make predictions.

Existing models with interaction terms between functional predictors and general scalar responses cannot deal with multiple functional and scalar predictors, which is the case in our motivated datasets. Moreover, the asymptotic properties of estimators have not been addressed in existing models. Therefore, in Section 2, we fully consider the combined influence of functional predictors, scalar predictors, and interactions between functional predictors on general scalar response by proposing a generalized partially functional linear model with interaction terms. In Section 3, the asymptotic properties of our proposed estimators are established. Extensive simulation studies are given in Section 4. Section 5 is reserved for the real data analysis.

## 2. Model and Estimation

### 2.1. Model Introduction

Suppose we have  $n$  subjects, and the data we observe for the  $i$ -th subject are  $\{(X_{i1}(t_1), t_1 \in T_1), (X_{i2}(t_2), t_2 \in T_2), Z_i, Y_i\}, i = 1, \dots, n$ . For  $j = 1, 2$ , the functional predictor  $X_{ij}(t_j)$  is a random curve, which is observed for subject  $i$  and  $X_{ij}(t_j) \in L^2(T_j)$ , where  $T_j$  is a bounded interval of  $\mathbb{R}$ . Notice that, for the sake of simplicity in notations, we only consider the case with two functional predictors, and the case with multiple functional predictors can be easily similarly established. The scalar predictor vector  $Z = (Z_1, Z_2, \dots, Z_q)^T$  is a  $q$ -dimensional random vector. The response  $Y_i$  is a real-valued random variable that may be continuous or discrete (e.g., binary, count, etc.).

We assume that there is a known link function  $g(\cdot)$ , which is a monotone and twice continuously differentiable function with bounded derivatives that is, thus, invertible.

We introduce the following generalized partially functional linear model with interaction between the functional predictors:

$$Y_i = g\left(\alpha + \int_{T_1} X_{i1}(t_1)\beta_1(t_1)dt_1 + \int_{T_2} X_{i2}(t_2)\beta_2(t_2)dt_2 + \iint_{T_1 \times T_2} X_{i1}(t_1)X_{i2}(t_2)\beta(t_1, t_2)dt_1dt_2 + Z_i^T \gamma\right) + \varepsilon_i, \tag{1}$$

where  $\alpha \in \mathbb{R}$  is the intercept,  $\beta_1(t_1), \beta_2(t_2)$ , and  $\beta(t_1, t_2)$  are the regression coefficient functions corresponding to the two functional predictors and the interaction term, respectively,  $Z_i$  represents the  $i$ -th random variable in  $Z$ , and  $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_q)^T$  is the regression coefficient corresponding to the multiple scalar predictors  $Z$ . It is assumed that  $\varepsilon_i$  has mean 0 and variance  $\sigma^2$  and that  $\varepsilon_i$  is independent with  $\varepsilon_j$  if  $i \neq j$ .

Define the linear operator  $\ell$ :

$$\ell = \alpha + \int_{T_1} X_1(t_1)\beta_1(t_1)dt_1 + \int_{T_2} X_2(t_2)\beta_2(t_2)dt_2 + \iint_{T_1 \times T_2} X_1(t_1)X_2(t_2)\beta(t_1, t_2)dt_1dt_2 + Z^T \gamma.$$

We specify

$$E(Y|X_1(\cdot), X_2(\cdot), Z) = \eta = g(\ell), \text{Var}(Y|X_1(\cdot), X_2(\cdot), Z) = \sigma^2(\eta).$$

For simplicity, we assume that the predictors  $X_j(t_j)$  and  $Z$  are both centralized, i.e.,  $E(X_j(t_j)) = 0, j = 1, 2$  and  $E(Z_l) = 0, l = 1, \dots, q$ . The Karhunen–Loève expansion is a mathematical technique for representing stochastic processes in terms of an orthonormal basis derived from the process’s covariance function. It decomposes a random process into a series of orthogonal functions, each weighted by uncorrelated random coefficients. The basis functions are the eigenfunctions of the covariance function, and the expansion efficiently captures the process’s variability, often using only a few terms. For a detailed introduction to the Karhunen–Loève expansion, please refer to Equation (2.8) in [2]. Based on Karhunen–Loève expansion,  $X_{ij}(t_j)$  can be expanded as

$$X_{i1}(t_1) = \sum_{k=1}^{\infty} \chi_{i1k} \varphi_{1k}(t_1),$$

$$X_{i2}(t_2) = \sum_{l=1}^{\infty} \chi_{i2l} \varphi_{2l}(t_2),$$

where  $\chi_{i1k}, \chi_{i2l}$  are the functional principal component scores,  $\varphi_{1k}(t_1), \varphi_{2l}(t_2)$  are the functional principal component bases, and

$$\int_{T_1} \varphi_{1k}^2(t_1)dt_1 = 1, \int_{T_2} \varphi_{2l}^2(t_2)dt_2 = 1.$$

Using the functional principal component bases, the regression coefficient functions  $\beta_j(t_j), \beta(t_1, t_2)$  are expanded as

$$\beta_1(t_1) = \sum_{k=1}^{\infty} b_{1k} \varphi_{1k}(t_1),$$

$$\beta_2(t_2) = \sum_{l=1}^{\infty} b_{2l} \varphi_{2l}(t_2),$$

$$\beta(t_1, t_2) = \sum_{k=1}^{\infty} \sum_{l=1}^{\infty} u_{kl} \varphi_{1k}(t_1) \varphi_{2l}(t_2).$$

Plugging the above expansions into Model (1) and truncating the predictors at  $p_j$ , which increases asymptotically with  $n \rightarrow \infty$ , we can obtain truncated Model (2):

$$Y_i = g\left(\alpha + \sum_{k=1}^{p_1} \chi_{i1k} b_{1k} + \sum_{l=1}^{p_2} \chi_{i2l} b_{2l} + \sum_{k=1}^{p_1} \sum_{l=1}^{p_2} \rho_{ikl} u_{kl} + Z_i^T \gamma\right) + \varepsilon_i, \tag{2}$$

where  $\rho_{ikl} = \chi_{i1k} \cdot \chi_{i2l}$ .

### 2.2. Parameter Estimation

Define the parameter vector

$$\Omega = (b_{11}, \dots, b_{1p_1}, b_{21}, \dots, b_{2p_2}, u_{11}, \dots, u_{1p_2}, u_{21}, \dots, u_{2p_2}, \dots, u_{p_1 1}, \dots, u_{p_1 p_2}, \gamma_0, \gamma_1, \dots, \gamma_q)^T,$$

and define

$$\ell_i = \alpha + \sum_{k=1}^{p_1} \chi_{i1k} b_{1k} + \sum_{l=1}^{p_2} \chi_{i2l} b_{2l} + \rho_i^T u + Z_i^T \gamma,$$

$$\eta_i = g(\ell_i),$$

$$\omega_i = (\chi_{i11}, \dots, \chi_{i1p_1}, \chi_{i21}, \dots, \chi_{i2p_2}, \rho_{i11}, \dots, \rho_{i1p_2}, \rho_{i21}, \dots, \rho_{i2p_2}, \dots, \rho_{ip_11}, \dots, \rho_{ip_1p_2}, z_{i0}, z_{i1}, \dots, z_{iq})^T,$$

where

$$b_j = (b_{j1}, \dots, b_{jp_j})^T, j = 1, 2,$$

$$u = (u_{11}, \dots, u_{1p_2}, u_{21}, \dots, u_{2p_2}, \dots, u_{p_11}, \dots, u_{p_1p_2})^T,$$

$$\gamma = (\gamma_0, \gamma_1, \dots, \gamma_q)^T,$$

$$\rho_i = (\rho_{i11}, \dots, \rho_{i1p_2}, \rho_{i21}, \dots, \rho_{i2p_2}, \dots, \rho_{ip_11}, \dots, \rho_{ip_1p_2})^T,$$

$z_{i0} = 1$ , and  $\gamma_0 = \alpha$ .

The maximum likelihood estimate  $\hat{\Omega}$  of  $\Omega$  can be obtained by solving Equation (3):

$$U(\Omega) = \sum_{i=1}^n \frac{(Y_i - g(\ell_i))g'(\ell_i)}{\sigma^2(\eta_i)} \omega_i = 0, \tag{3}$$

$$\hat{\Omega} = (\hat{b}_{11}, \dots, \hat{b}_{1p_1}, \hat{b}_{21}, \dots, \hat{b}_{2p_2}, \hat{u}_{11}, \dots, \hat{u}_{1p_2}, \hat{u}_{21}, \dots, \hat{u}_{2p_2}, \dots, \hat{u}_{p_11}, \dots, \hat{u}_{p_1p_2}, \hat{\gamma}_0, \hat{\gamma}_1, \dots, \hat{\gamma}_q)^T,$$

where

$$\hat{b}_j = (\hat{b}_{j1}, \dots, \hat{b}_{jp_j})^T, j = 1, 2,$$

$$\hat{u} = (\hat{u}_{11}, \dots, \hat{u}_{1p_2}, \hat{u}_{21}, \dots, \hat{u}_{2p_2}, \dots, \hat{u}_{p_11}, \dots, \hat{u}_{p_1p_2})^T,$$

$\hat{\alpha} = \hat{\gamma}_0$ , and  $\hat{\gamma} = (\hat{\gamma}_0, \hat{\gamma}_1, \dots, \hat{\gamma}_q)^T$  are the estimates of  $b_j, u, \alpha, \gamma$ , respectively.

Introducing the following matrices:

$$V = \text{diag}(\sigma^2(\eta_1), \dots, \sigma^2(\eta_n)),$$

$$W = \text{diag}(g'(\ell_1), g'(\ell_2), \dots, g'(\ell_n)),$$

$$A_0 = A_{n,q+1} = \left( \frac{g'(\ell_i) z_{im}}{\sigma(\eta_i)} \right)_{1 \leq i \leq n, 0 \leq m \leq q},$$

$$A_j = A_{n,p_j} = \left( \frac{g'(\ell_i) \chi_{ijr}}{\sigma(\eta_i)} \right)_{1 \leq i \leq n, 0 \leq r \leq p_j}, j \in \{1, 2\},$$

$$A_{12} = A_{n,t} = \left( \frac{g'(\ell_i) \rho_{it}}{\sigma(\eta_i)} \right)_{1 \leq i \leq n, 1 \leq t \leq p_1 p_2},$$

$$A = A_{n,q+1+p_1+p_2+p_1 p_2} = \text{diag}(A_1, A_2, A_{12}, A_0),$$

and the vectors  $Y = (Y_1, \dots, Y_n)^T, \eta = (\eta_1, \dots, \eta_n)^T$ , then Equation (3) can be written as

$$A^T V^{-\frac{1}{2}} (Y - \eta) = 0.$$

The estimation of  $\Omega$  is usually solved iteratively using a weighted least squares method. By Taylor expansion, we have

$$\begin{aligned} g^{-1}(Y) &= g^{-1}(\eta) + [g^{-1}(\eta)]'(Y - \eta) \\ &= \ell + W^{-1}(Y - \eta); \end{aligned}$$

thus, there is

$$A^T H(g^{-1}(Y) - \ell) = 0,$$

where  $H = V^{-\frac{1}{2}}W$ .

Simplify to obtain estimates of  $b_j, \gamma, u$ :

$$\tilde{b}_j = (A_j^T A_j)^{-1} A_j^T H g^{-1}(Y),$$

$$\tilde{\gamma} = (A_0^T A_0)^{-1} A_0^T H g^{-1}(Y),$$

$$\tilde{u} = (A_{12}^T A_{12})^{-1} A_{12}^T H g^{-1}(Y).$$

Repeat the above process until convergence; then, the estimate of  $\Omega$  is obtained:

$$\hat{\Omega} = (\hat{b}_{11}, \dots, \hat{b}_{1p_1}, \hat{b}_{21}, \dots, \hat{b}_{2p_2}, \hat{u}_{11}, \dots, \hat{u}_{1p_2}, \hat{u}_{21}, \dots, \hat{u}_{2p_2}, \dots, \hat{u}_{p_1 1}, \dots, \hat{u}_{p_1 p_2}, \hat{\gamma}_0, \hat{\gamma}_1, \dots, \hat{\gamma}_q)^T.$$

### 3. Asymptotic Properties

Considering the truncated Model (2), we have the metric

$$d_G^2(\hat{\beta}_j, \beta_j) = (\hat{b}_j - b_j)^T \tilde{\Gamma}_j (\hat{b}_j - b_j) + \sum_{k_1, k_2 = p_j + 1}^{\infty} \lambda_{j, k_1 k_2} \bar{b}_j^2, j = 1, 2,$$

where

$$b_j = (b_{j1}, b_{j2}, \dots, b_{jp_j}),$$

$$\tilde{\Gamma}_j = (\lambda_{j, k_1 k_2})_{1 \leq k_1, k_2 \leq p_j}$$

is a symmetric positive definite matrix and

$$\lambda_{j, k_1 k_2} = E \left[ \frac{g'(\ell)^2}{\sigma^2(\eta)} \chi_{jk_1} \chi_{jk_2} \right]_{1 \leq k_1, k_2 \leq p_j}$$

is an eigenvalue of the generalized self-covariance operator  $A_{G_j}$  with kernel

$$G_j(s, t) = E \left[ \frac{g'(\ell)^2}{\sigma^2(\eta)} X_j(s) X_j(t) \right],$$

and we have

$$\tilde{\Gamma}_j^{-1} = (\xi_{j, k_1 k_2})_{1 \leq k_1, k_2 \leq p_j}$$

and

$$\bar{b}_j = (b_{j(p_j+1)}, b_{j(p_j+2)}, \dots)^T.$$

Combined with Corollary 4.1 in Müller (2005) [11], we have

$$\sum_{k_1, k_2 = p_j + 1}^{\infty} \lambda_{j, k_1 k_2} \bar{b}_j^2 = o\left(\frac{\sqrt{p_j}}{n}\right).$$

We specify

$$\|f\|^2 = \int_S f(s)^2 ds, f \in L^2(S),$$

$$\|g\|^2 = \int_S \int_T g(s, t)^2 ds dt, g \in L^2(S \times T)$$

and

$$(f \otimes g)(x, y) = f(x)g(y), x \in X, y \in Y,$$

where  $X, Y$  are the domains of  $f, g$ , respectively.

Define  $C_{X_j}$  as the covariance function of a random function  $X_j$ , for  $j = 1, 2$ . By Mercer’s theorem,

$$C_{X_1}(t_{11}, t_{12}) = \sum_{k \geq 1} \lambda_k \varphi_{1k}(t_{11}) \varphi_{1k}(t_{12}),$$

$$C_{X_2}(t_{21}, t_{22}) = \sum_{l \geq 1} \sigma_l \varphi_{2l}(t_{21}) \varphi_{2l}(t_{22}),$$

where  $t_{11}, t_{12} \in T_1; t_{21}, t_{22} \in T_2; \lambda_k$  and  $\varphi_{1k}, k = 1, 2, \dots$ , are the non-negative eigenvalues and the corresponding eigenfunctions of the covariance function  $C_{X_1}(t_{11}, t_{12})$ ; and  $\sigma_l$  and  $\varphi_{2l}, l = 1, 2, \dots$ , are the non-negative eigenvalues and the corresponding eigenfunctions of the covariance function  $C_{X_2}(t_{21}, t_{22})$ .

In order to derive the asymptotic nature of the regression coefficients, we have made the following assumptions in addition to the basic conditions in Section 2:

- (i) The connected function  $g(\cdot)$  is monotonically invertible and has bounded second-order derivatives, the derivative of the variance function  $\sigma^2(\cdot)$  is continuously bounded, and there exists an  $\sigma(\cdot) > \Delta > 0$ ;
- (ii) The scalar predictor variable  $Z$  and the functional predictor variable  $X_j(t_j)$  are independent of each other;
- (iii) When  $n \rightarrow \infty$ ,  $p_j$  satisfies  $p_j \rightarrow \infty$  and  $p_j n^{-\frac{1}{4}} \rightarrow 0$ ;
- (iv)  $E[\int_{T_j} \{X_j(t_j)\}^4 dt_j] < \infty$ ;
- (v) Define  $\mu_{X_1,k} = \min_{1 \leq k \leq p_1} (\lambda_{p_1} - \lambda_{k+1}), \mu_{X_2,l} = \min_{1 \leq l \leq p_2} (\sigma_l - \sigma_{l+1})$ , and  $\mu_{X_1,k} > 0, \mu_{X_2,l} > 0$ ;
- (vi) Define  $d_n = \|\hat{C}_X - C_X\|, \tilde{K}_n = \min\{k \geq 1 : \lambda_k \leq 2d_n\} - 1, \tilde{L}_n = \min\{l \geq 1 : \sigma_l \leq 2d_n\} - 1; d_n \rightarrow 0, \tilde{K}_n \rightarrow \infty$ , and  $\tilde{L}_n \rightarrow \infty$  when  $n \rightarrow \infty$ .

**Lemma 1.** *If the above basic conditions and assumptions hold, while  $p_1 \leq \tilde{K}_n, p_2 \leq \tilde{L}_n$  and*

$$\sum_{k=1}^{p_1} \sum_{l=1}^{p_2} \frac{1}{n} \left( \frac{1}{\mu_{X_1,k}^2} + \frac{1}{\mu_{X_2,l}^2} \right) \rightarrow 0,$$

we have  $\|\hat{\beta} - \beta\|^2 = O_p\left(\sum_{t=1}^{p_1 p_2} (u_t - \hat{u}_t)^2\right)$ .

**Proof.**

$$\begin{aligned} \|\beta - \hat{\beta}\|^2 &= \left\| \sum_{k=1}^{\infty} \sum_{l=1}^{\infty} u_{kl} \varphi_{1k} \otimes \varphi_{2l} - \sum_{k=1}^{p_1} \sum_{l=1}^{p_2} \hat{u}_{kl} \hat{\varphi}_{1k} \otimes \hat{\varphi}_{2l} \right\|^2 \\ &= \left\| \sum_{k=1}^{p_1} \sum_{l=1}^{p_2} u_{kl} (\varphi_{1k} \otimes \varphi_{2l} - \hat{\varphi}_{1k} \otimes \hat{\varphi}_{2l}) + \sum_{k=1}^{p_1} \sum_{l=1}^{p_2} (u_{kl} - \hat{u}_{kl}) \hat{\varphi}_{1k} \otimes \hat{\varphi}_{2l} \right. \\ &\quad \left. + \sum_{k=1}^{p_1} \sum_{l > p_2} u_{kl} \varphi_{1k} \otimes \varphi_{2l} + \sum_{k > p_1} \sum_{l=1}^{\infty} u_{kl} \varphi_{1k} \otimes \varphi_{2l} \right\|^2 \\ &\leq 4 \left\| \sum_{k=1}^{p_1} \sum_{l=1}^{p_2} u_{kl} (\varphi_{1k} \otimes \varphi_{2l} - \hat{\varphi}_{1k} \otimes \hat{\varphi}_{2l}) \right\|^2 + 4 \sum_{k=1}^{p_1} \sum_{l=1}^{p_2} (u_{kl} - \hat{u}_{kl})^2 \\ &\quad + 4 \sum_{k=1}^{p_1} \sum_{l > p_2} u_{kl}^2 + 4 \sum_{k > p_1} \sum_{l=1}^{\infty} u_{kl}^2 \\ &= 4I_1 + 4 \sum_{t=1}^{p_1 p_2} (u_t - \hat{u}_t)^2 + 4R_{\beta}(p_1, p_2)^2, \end{aligned}$$

where  $R_\beta(p_1, p_2) = \left( \sum_{k=1}^{p_1} \sum_{l>p_2} u_{kl}^2 + \sum_{k>p_1} \sum_{l=1}^{\infty} u_{kl}^2 \right)^{\frac{1}{2}} \rightarrow 0, \|\beta\|^2 < \infty (p_1, p_2 \rightarrow \infty)$ .

From the Cauchy–Schwarz’s inequality and Yifan Sun (2020) [22] Lemmas 1 and 2, we obtain

$$\begin{aligned} I_1 &= \left\| \sum_{k=1}^{p_1} \sum_{l=1}^{p_2} u_{kl} (\varphi_{1k} \otimes \varphi_{2l} - \hat{\varphi}_{1k} \otimes \hat{\varphi}_{2l}) \right\|^2 \\ &= \int_{T_1} \int_{T_2} \left[ \sum_{k=1}^{p_1} \sum_{l=1}^{p_2} u_{kl} (\varphi_{1k} \otimes \varphi_{2l} - \hat{\varphi}_{1k} \otimes \hat{\varphi}_{2l}) \right]^2 dsdt \\ &\leq \int_{T_1} \int_{T_2} \left( \sum_{k=1}^{p_1} \sum_{l=1}^{p_2} u_{kl}^2 \right) \left[ \sum_{k=1}^{p_1} \sum_{l=1}^{p_2} (\varphi_{1k} \otimes \varphi_{2l} - \hat{\varphi}_{1k} \otimes \hat{\varphi}_{2l})^2 \right] dsdt \\ &\leq \|\beta\|^2 \sum_{k=1}^{p_1} \sum_{l=1}^{p_2} \|\varphi_{1k} \otimes \varphi_{2l} - \hat{\varphi}_{1k} \otimes \hat{\varphi}_{2l}\|^2 \\ &= \|\beta\|^2 \sum_{k=1}^{p_1} \sum_{l=1}^{p_2} \frac{C}{n} \left( \frac{1}{\mu_{X_1,k}^2} + \frac{1}{\mu_{X_2,l}^2} \right). \end{aligned}$$

Therefore, we have

$$I_1 = O_p \left( \sum_{k=1}^{p_1} \sum_{l=1}^{p_2} \frac{1}{n} \left( \frac{1}{\mu_{X_1,k}^2} + \frac{1}{\mu_{X_2,l}^2} \right) \right) \rightarrow 0.$$

Thus, we conclude that

$$\|\beta - \hat{\beta}\| = O_p \left( \sum_{t=1}^{p_1 p_2} (u_t - \hat{u}_t)^2 \right).$$

Therefore, Lemma 1 is proven. □

**Theorem 1.** *If the above conditions and assumptions hold, then we have*

$$\begin{pmatrix} \frac{nd_G^2(\hat{\beta}_1, \beta_1) - p_1}{\sqrt{2p_1}} \\ \frac{nd_G^2(\hat{\beta}_2, \beta_2) - p_2}{\sqrt{2p_2}} \\ \frac{nd^2(\hat{\beta}, \hat{\beta}) - p_1 p_2 \tau}{\sqrt{2p_1 p_2 \tau}} \\ \sqrt{n\Theta_0}(\gamma_0 - \hat{\gamma}_0) \\ \sqrt{n\Theta_1}(\gamma_1 - \hat{\gamma}_1) \\ \vdots \\ \sqrt{n\Theta_q}(\gamma_q - \hat{\gamma}_q) \end{pmatrix} \rightarrow N(0, I),$$

where  $\Theta_m = E \left[ \frac{g'(\ell_i)^2}{\sigma^2(\eta_i)} z_{im}^2 \right], \tau = E \left[ \frac{g'(\ell_i)^2}{\sigma^2(\eta_i)} \rho_{it}^2 \right]$ , and  $I$  is a unit matrix of  $(q + 1 + p_1 + p_2) \times (q + 1 + p_1 + p_2)$ .

**Proof.** A Taylor expansion-based approach is used to prove the asymptotic normality of the estimates. The Hessian of the proposed likelihood is  $J_\Omega = \Delta_\Omega U(\Omega)$  and

$$A^T A = \sum_{i=1}^n \frac{g'(\ell_i)^2}{\sigma^2(\eta_i)} \omega_i \omega_i^T.$$

Thus, we have



$$\begin{aligned}
 J_{\Omega} &= \frac{\partial U(\Omega)}{\partial \Omega} = \frac{\partial U(\Omega)}{\partial \ell_i} \frac{\partial \ell_i}{\partial \Omega} \\
 &= - \sum_{i=1}^n \frac{g'^2(\ell_i) \omega_i \omega'_i}{\sigma^2(g(\ell_i))} + \sum_{i=1}^n \left( \frac{g''(\ell_i)}{\sigma^2(\eta_i)} - \frac{g'^2(\ell_i) \sigma^{2'}(\eta_i)}{\sigma^4(\eta_i)} \right) (Y_i - g(\ell_i)) \omega_i \omega'_i \\
 &= -A^T A + R.
 \end{aligned}$$

The remainder term R can be ignored and, using Taylor expansions, we obtain a  $\tilde{\Omega}$  that lies between  $\Omega$  and  $\hat{\Omega}$ . We have

$$\frac{U(\Omega) - U(\hat{\Omega})}{\Omega - \hat{\Omega}} = J_{\tilde{\Omega}};$$

therefore,

$$\sqrt{n}(\Omega - \hat{\Omega}) = [I + M + N]^{-1} \left( \frac{A^T A}{n} \right)^{-1} \frac{U(\Omega)}{\sqrt{n}},$$

where  $M = \left( \frac{A^T A}{n} \right)^{-1} \frac{J_{\tilde{\Omega}} - J_{\Omega}}{n}$ ,  $N = \left( \frac{A^T A}{n} \right)^{-1} \frac{J_{\Omega} - A^T A}{n}$ .

From Lemma 7.1 in Müller (2005) [11], it follows that

$$\sqrt{n}(\Omega - \hat{\Omega}) \sim \left( \frac{A^T A}{n} \right)^{-1} \frac{U(\Omega)}{\sqrt{n}}.$$

Asymptotic convergence of the lower proof  $\left( \frac{A^T A}{n} \right)^{-1} \frac{U(\Omega)}{\sqrt{n}}$  gives

$$\left( \frac{A^T A}{n} \right)^{-1} \frac{U(\Omega)}{\sqrt{n}} = \left( \frac{A^T A}{n} \right)^{-1} \frac{A^T V^{-\frac{1}{2}}(Y - \eta)}{\sqrt{n}} = \left( \frac{A^T A}{n} \right)^{-1} \frac{A^T \bar{\varepsilon}}{\sqrt{n}},$$

where  $\bar{\varepsilon} = \frac{\varepsilon}{\sigma(\eta)}$  and follows a standard normal distribution.

Thus, we have

$$\sqrt{n}(\Omega - \hat{\Omega}) \sim \left( \frac{A^T A}{n} \right)^{-1} \frac{A^T \bar{\varepsilon}}{\sqrt{n}}.$$

Since  $\beta_j(t_j)$ ,  $\beta(t_1, t_2)$ , and  $\gamma$  are of different data types,  $\sqrt{n}(\Omega - \hat{\Omega})$  is divided into three terms, i.e.,

$$\sqrt{n}(b_j - \hat{b}_j) \sim \left( \frac{A_j^T A_j}{n} \right)^{-1} \frac{A_j^T \bar{\varepsilon}}{\sqrt{n}}, j = 1, 2,$$

$$\sqrt{n}(u - \hat{u}) \sim \left( \frac{A_{12}^T A_{12}}{n} \right)^{-1} \frac{A_{12}^T \bar{\varepsilon}}{\sqrt{n}},$$

$$\sqrt{n}(\gamma - \hat{\gamma}) \sim \left( \frac{A_0^T A_0}{n} \right)^{-1} \frac{A_0^T \bar{\varepsilon}}{\sqrt{n}},$$

where  $A_j^T A_j$  are symmetric matrices and  $A_0^T A_0$  are diagonal matrices.

First, we prove

$$\sqrt{n}(b_j - \hat{b}_j) \sim \left( \frac{A_j^T A_j}{n} \right)^{-1} \frac{A_j^T \bar{\varepsilon}}{\sqrt{n}}, j = 1, 2.$$

Let

$$\mathcal{X}_{nj} = \frac{\tilde{\Lambda}_j^{-\frac{1}{2}} A_j^T \bar{\varepsilon}}{\sqrt{n}}, \mathcal{Z}_{nj} = \left( \frac{A_j^T A_j}{n} \right)^{-1} \frac{A_j^T \bar{\varepsilon}}{\sqrt{n}},$$

$$\Psi_{nj} = \tilde{\Gamma}_j^{\frac{1}{2}} \left( \frac{A_j^T A_j}{n} \right)^{-1} \tilde{\Gamma}_j^{\frac{1}{2}}.$$

Thus, we have

$$\begin{aligned} nd_G^2(\beta, \hat{\beta}) &= Z_{nj}^T \tilde{\Gamma}_j Z_{nj} = \mathcal{X}_{nj}^T \Psi_{nj}^2 \mathcal{X}_{nj} \\ &= \mathcal{X}_{nj}^T \mathcal{X}_{nj} + 2\mathcal{X}_{nj}^T (\Psi_{nj} - I_{nj}) \mathcal{X}_{nj} + \mathcal{X}_{nj}^T (\Psi_{nj} - I_{nj})(\Psi_{nj} - I_{nj}) \mathcal{X}_{nj}, \end{aligned}$$

From Lemma 7.2 in Müller (2005) [11], it follows that

$$nd_G^2(\hat{\beta}_j, \beta_j) = \mathcal{X}_{nj}^T \mathcal{X}_{nj}.$$

Then,

$$\begin{aligned} \mathcal{X}_{nj}^T \mathcal{X}_{nj} &= \frac{1}{n} \sum_{k_1=1}^{p_j} \left( \sum_{k_2=1}^{p_j} \zeta_{j,k_1 k_2}^{\frac{1}{2}} \sum_{i=1}^n \frac{g'(\eta_i) \chi_{ijk_2}}{\sigma(\mu_i)} \bar{\varepsilon}_i \right)^2 \\ &= E + F, \end{aligned}$$

where

$$\begin{aligned} E &= \frac{1}{n} \sum_{i=1}^n \bar{\varepsilon}_i^2 \sum_{k_2, k_2'=1}^{p_j} \frac{g'(\ell_i)^2}{\sigma^2(\eta_i)} \chi_{ijk_2} \chi_{ijk_2'} \sum_{k_1=1}^{p_j} \zeta_{j,k_1 k_2}^{\frac{1}{2}} \zeta_{j,k_1 k_2'}^{\frac{1}{2}}, \\ F &= \frac{1}{n} \sum_{i_1 \neq i_2=1}^n \bar{\varepsilon}_{i_1} \bar{\varepsilon}_{i_2} \frac{g'(\ell_{i_1}) g'(\ell_{i_2})}{\sigma(\eta_{i_1}) \sigma(\eta_{i_2})} \sum_{k_2, k_2'=1}^{p_j} \chi_{i_1 j k_2} \chi_{i_2 j k_2'} \sum_{k_1=1}^{p_j} \zeta_{j,k_1 k_2}^{\frac{1}{2}} \zeta_{j,k_1 k_2'}^{\frac{1}{2}}. \end{aligned}$$

Since  $\bar{\varepsilon}$  follows a standard normal distribution, we have

$$E[\mathcal{X}_{nj}^T \mathcal{X}_{nj}] = p_j, \text{Var}[\mathcal{X}_{nj}^T \mathcal{X}_{nj}] = 2p_j.$$

Therefore,

$$\frac{nd_G^2(\hat{\beta}_j, \beta_j) - p_j}{\sqrt{2p_j}} \rightarrow N(0, 1) \quad j = 1, 2.$$

The following proves that

$$\sqrt{n}(u - \hat{u}) \sim \left( \frac{A_{12}^T A_{12}}{n} \right)^{-1} \frac{A_{12}^T \bar{\varepsilon}}{\sqrt{n}}.$$

For the coefficient function of the interaction term, we have the metric  $d^2(\beta, \hat{\beta}) = \|\beta - \hat{\beta}\|^2$ , so, according to Lemma 1, we have

$$d^2(\beta, \hat{\beta}) = \sum_{t=1}^{p_1 p_2} (u_t - \hat{u}_t)^2 = (u - \hat{u})^T (u - \hat{u}).$$

Let

$$\begin{aligned} Q_{nt} &= \left( \frac{A_{12}^T A_{12}}{n} \right)^{-1} \frac{A_{12}^T \bar{\varepsilon}}{\sqrt{n}}, \\ \mathcal{A}_{nt} &= \frac{A_{12}^T \bar{\varepsilon}}{\sqrt{n}}, \mathcal{F}_{nt} = \left( \frac{A_{12}^T A_{12}}{n} \right)^{-1}. \end{aligned}$$

Thus, we have

$$\begin{aligned} nd^2(\beta, \hat{\beta}) &= Q_{nt}^T Q_{nt} = \mathcal{A}_{nt}^T \mathcal{F}_{nt}^2 \mathcal{A}_{nt} \\ &= \mathcal{A}_{nt}^T \mathcal{A}_{nt} + 2\mathcal{A}_{nt}^T (\mathcal{F}_{nt} - I_{nt}) \mathcal{A}_{nt} + \mathcal{A}_{nt}^T (\mathcal{F}_{nt} - I_{nt})(\mathcal{F}_{nt} - I_{nt}) \mathcal{A}_{nt}, \end{aligned}$$

From Lemma 7.2 in Müller (2005) [11], it follows that

$$nd^2(\beta, \hat{\beta}) = \mathcal{A}_{nt}^T \mathcal{A}_{nt}.$$

Then,

$$\begin{aligned} \mathcal{A}_{nt}^T \mathcal{A}_{nt} &= \frac{1}{n} \sum_{t=1}^{p_1 p_2} \left( \sum_{i=1}^n \frac{g'(\ell_i) \rho_{it}}{\sigma(\eta_i)} \bar{\varepsilon}_i \right)^2 \\ &= \frac{1}{n} \sum_{i=1}^n \bar{\varepsilon}_i^2 \sum_{t=1}^{p_1 p_2} \frac{g'(\ell_i)^2}{\sigma^2(\eta_i)} \rho_{it}^2 \\ &\quad + \frac{1}{n} \sum_{i_1 \neq i_2=1}^n \bar{\varepsilon}_{i_1} \bar{\varepsilon}_{i_2} \frac{g'(\ell_{i_1})}{\sigma(\eta_{i_1})} \frac{g'(\ell_{i_2})}{\sigma(\eta_{i_2})} \sum_{t=1}^{p_1 p_2} \rho_{i_1 t} \rho_{i_2 t}. \end{aligned}$$

We have

$$E[\mathcal{A}_{nk}^T \mathcal{A}_{nk}] = p_1 p_2 \tau, \text{Var}[\mathcal{A}_{nk}^T \mathcal{A}_{nk}] = 2p_1 p_2 \tau.$$

Thus, there is

$$\frac{nd^2(\beta, \hat{\beta}) - p_1 p_2 \tau}{\sqrt{2p_1 p_2 \tau}} \rightarrow N(0, 1).$$

Next, prove that

$$\sqrt{n}(\gamma - \hat{\gamma}) \sim \left( \frac{A_0^T A_0}{n} \right)^{-1} \frac{A_0^T \bar{\varepsilon}}{\sqrt{n}}.$$

Let

$$\mathcal{Z}_0 = \left( \frac{A_0^T A_0}{n} \right)^{-1} \frac{A_0^T \bar{\varepsilon}}{\sqrt{n}}.$$

Then, its matrix form is

$$\mathcal{Z}_0 = \sqrt{n} \left( \sum_{i=1}^n \frac{g'(\ell_i)^2}{\sigma^2(\eta_i)} z_{im}^2 \right)^{-1} \sum_{i=1}^n \frac{g'(\ell_i) z_{im}}{\sigma(\eta_i)} \bar{\varepsilon}_i.$$

Therefore, we have

$$\sqrt{n}(\gamma_m - \hat{\gamma}_m) \sim \sqrt{n} \left( \sum_{i=1}^n \frac{g'(\ell_i)^2}{\sigma^2(\eta_i)} z_{im}^2 \right)^{-1} \sum_{i=1}^n \frac{g'(\ell_i) z_{im}}{\sigma(\eta_i)} \bar{\varepsilon}_i.$$

Since

$$E[\sqrt{n}(\gamma_m - \hat{\gamma}_m)] = 0,$$

$$\text{Var}[\sqrt{n}(\gamma_m - \hat{\gamma}_m)] = \left( E \left[ \frac{g'(\ell_i)^2}{\sigma^2(\eta_i)} z_{im}^2 \right] \right)^{-1} = \Theta_m^{-1}.$$

There is

$$\sqrt{n} \Theta_m (\gamma_m - \hat{\gamma}_m) \rightarrow N(0, 1).$$

Therefore, Theorem 1 is proven.  $\square$

#### 4. Simulation

In this simulation, we consider the case that has two functional predictors, three scalar predictors, an interaction term between the two functional predictors, and a binary response. In order to include the case in which the functional predictors do not have the same domain, we define the functional predictors  $X_{i1}(t_1), t_1 \in [0, 1]$  and  $X_{i2}(t_2), t_2 \in [-1, 1], i = 1, \dots, n$ , where  $n$  can be any positive integer. In the latter sample size,  $n$  takes the values of 50, 100,

and 500, and, for each  $n$ , we run 100 simulations. First, we define two standard orthogonal bases  $\varphi_{1k}(t_1), t_1 \in [0, 1]$  and  $\varphi_{2l}(t_2), t_2 \in [-1, 1]$ , satisfying

$$\varphi_{1k}(t_1) = \sqrt{2} \cos(2k\pi t_1), k = 1, \dots, 4,$$

$$\varphi_{2l}(t_2) = \sqrt{2} \sin(2l\pi t_2), l = 1, \dots, 5.$$

Under the Gaussian assumption, we define the two randomly generated functional principal component scores  $\chi_{i1k}, \chi_{i2l}$  that satisfy

$$\chi_{i1k} \sim N(0, \lambda_{1k}), k = 1, \dots, 4,$$

$$\chi_{i2l} \sim N(0, \lambda_{2l}), l = 1, \dots, 5,$$

where  $\lambda_{11} = 8, \lambda_{12} = 6, \lambda_{13} = 4, \lambda_{14} = 2, \lambda_{21} = 4, \lambda_{22} = 2, \lambda_{23} = 1, \lambda_{24} = \frac{1}{2}, \lambda_{25} = \frac{1}{4}$ . Notice that the first three functional principal components explain up to 90% of the variation in the two predictors. So, we have

$$X_{i1}(t_1) = \sum_{k=1}^4 \chi_{i1k} \varphi_{1k}(t_1),$$

$$X_{i2}(t_2) = \sum_{l=1}^5 \chi_{i2l} \varphi_{2l}(t_2).$$

Fifty images of  $X_1(t_1)$  and  $X_2(t_2)$  are shown in Figure 1.

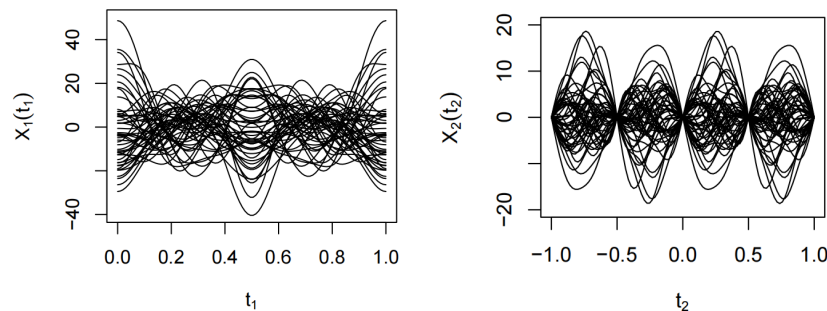


Figure 1. Functional predictors  $X_1(t_1)$  and  $X_2(t_2)$ .

For scalar predictors, we assume  $Z_1 \sim N(0, 2), Z_2 \sim N(0, 5)$ , and  $Z_3 \sim N(0, 6)$ . We assume that the theoretical values of the regression coefficients are

$$\gamma = (4, 6, 8)^T,$$

$$\beta_1(t_1) = \sum_{k=1}^4 b_{1k} \varphi_{1k}(t_1),$$

$$\beta_2(t_2) = \sum_{l=1}^5 b_{2l} \varphi_{2l}(t_2),$$

where  $b_{1k} = \frac{k}{5}, b_{2l} = \frac{l^2}{25}$ .

For the interaction term, its principal component score is denoted by  $\rho_{ikl}$  and satisfies

$$\rho_{ikl} = \chi_{i1k} \chi_{i2l},$$

$$\psi_{kl}(t_1, t_2) = \varphi_{1k}(t_1) \varphi_{2l}(t_2),$$

$$\beta(t_1, t_2) = \sum_{k=1}^4 \sum_{l=1}^5 u_{kl} \varphi_{1k}(t_1) \varphi_{2l}(t_2),$$

where  $u_{kl} = \left(\frac{k}{10}\right)^2$ .

The corresponding response variable is generated by

$$p(X_i, Z_i) = g\left(\int_{T_1} X_{i1}(t_1)\beta_1(t_1)dt_1 + \int_{T_2} X_{i2}(t_2)\beta_2(t_2)dt_2 + \iint_{T_1 \times T_2} X_{i1}(t_1)X_{i2}(t_2)\beta(t_1, t_2)dt_1dt_2 + Z_i^T \gamma\right),$$

where the link function  $g(x) = \frac{\exp(x)}{1+\exp(x)}$  and  $Y(X, Z) \sim \text{Bernoulli}(p(X, Z))$  is a sequence of pseudo-random numbers.

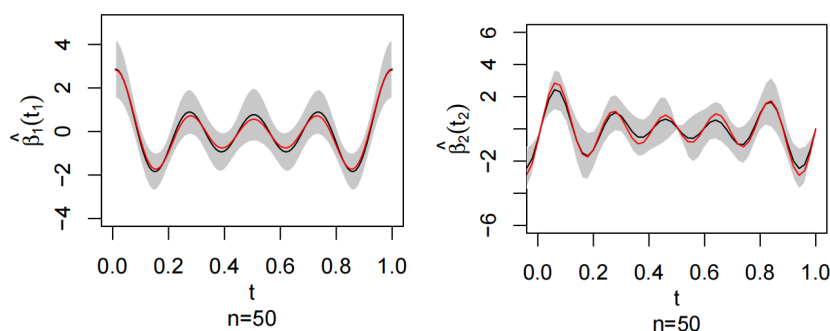
The principal component analysis was performed for  $n = 50, 100, 500$ , and the running results showed that the principal component scores of  $X_1$  with 90% cumulative contribution were 3, 3, 3 for each sample size, respectively and the principal component scores of  $X_2$  with 90% cumulative contribution were 2, 2, 2.

Table 1 shows how the standardized prediction error (SPE) varies with different sample sizes, and the results show that the model’s predictions become more and more accurate as the sample size increases. Here, SPE is defined by  $\sum_i |\hat{Y}_i - Y_i| / \sum_i |Y_i|$ .

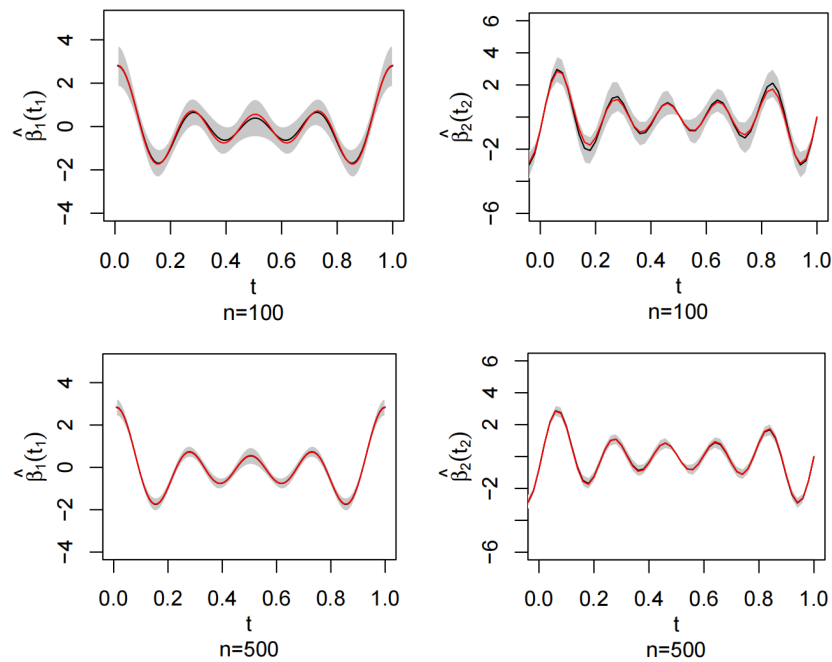
**Table 1.** Standardized prediction error for different sample sizes.

$n$	SPE
50	0.0156
100	0.0130
500	0.0106

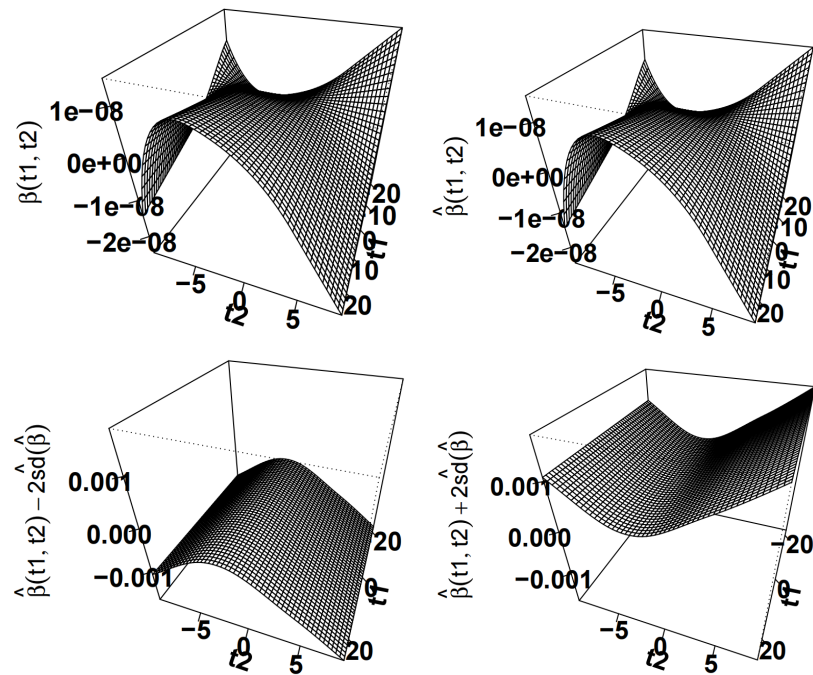
Figure 2 shows  $\hat{\beta}_1(t_1), \hat{\beta}_2(t_2)$  and the corresponding 95% confidence interval bands for different sample sizes, where the red curves are the theoretical values of  $\beta_1(t_1)$  and  $\beta_2(t_2)$  and the black curves are the corresponding estimates  $\hat{\beta}_1(t_1)$  and  $\hat{\beta}_2(t_2)$ . From Figure 2, it can be seen that, as the sample size increases, the estimated value becomes closer to the theoretical value. Figure 3  $\hat{\beta}(t_1, t_2)$  shows the visualized 3D plot with  $\hat{\beta}(t_1, t_2)$  in the middle panel and the 95% confidence intervals for  $\hat{\beta}(t_1, t_2)$  in the left and right panels.



**Figure 2.** Cont.



**Figure 2.** Estimated regression coefficient functions  $\hat{\beta}_1(t_1)$ ,  $\hat{\beta}_2(t_2)$  (black curves) and their 95% confidence bands (grey area) for different sample sizes, where the red curves are the theoretical regression coefficient functions  $\beta_1(t_1)$ ,  $\beta_2(t_2)$ .



**Figure 3.**  $\beta(t_1, t_2)$  and  $\hat{\beta}(t_1, t_2)$  are visualized in 3D.

Table 2 shows the estimated values of  $\hat{\gamma}$  and their corresponding standard deviations for different sample sizes. It can be seen that, as  $n$  increases, the standard deviation becomes smaller and the estimated value of  $\gamma$  becomes closer to the theoretical value, where the theoretical values of  $\gamma$  are 4, 6, and 8, respectively. Table 3 shows the standard deviation and root mean square error for  $\hat{\beta}_1$ ,  $\hat{\beta}_2$  and  $\hat{\beta}(t_1, t_2)$  for different sample sizes. Here, we use the coefficients of the basis expansion of the regression coefficient function to calculate the root mean square error. For example, the root mean square error of  $\hat{\beta}_1$  is  $\sqrt{\sum_{k=1}^4 (\hat{b}_{1k} - b_{1k})^2}$ .

The results show that, as  $n$  increases, both the standard deviation and the RMS error become smaller, indicating that, as sample size increases, the prediction becomes more accurate.

**Table 2.** Estimates of the regression coefficients and their standard deviations.

$n$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$
50	3.912 (0.193)	5.902 (0.068)	8.053 (0.045)
100	4.013 (0.062)	5.958 (0.026)	8.015 (0.018)
500	3.998 (0.022)	5.993 (0.006)	7.996 (0.007)

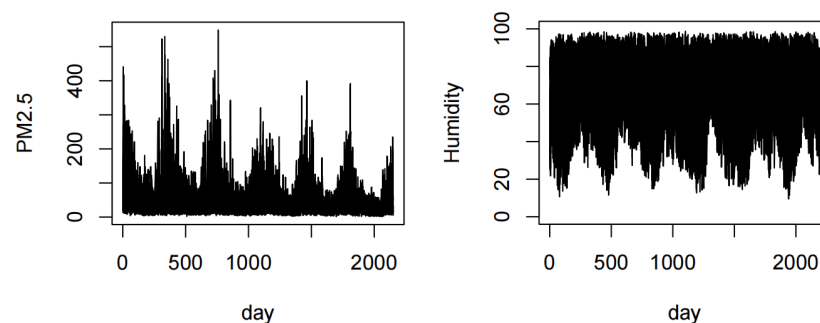
**Table 3.** Standard deviation and root mean square error of the estimated values of the regression coefficient function.

	$n$	Sd	RMSE
$\hat{\beta}_1$	50	0.034	0.026
	100	0.018	0.006
	500	0.008	0.001
$\hat{\beta}_2$	50	0.325	0.204
	100	0.126	0.044
	500	0.043	0.004
$\hat{\beta}(t_1, t_2)$	50	0.137	0.094
	100	0.054	0.024
	500	0.020	0.004

### 5. Application

To investigate the influence of the influence of air qualities, climate factors, medical and social indicators, and their interactions on cancer incidence using the proposed model, we collected data on average daily PM2.5 concentration, average daily humidity, per capita GDP, green coverage rate in built-up areas, the proportion of medical personnel (PMP), and the incidence of cancer in 49 cities in China from <http://www.cnemc.cn/>, <http://www.stats.gov.cn/sj/ndsjs/>, and <http://www.chinancpcn.org.cn/home>.

There are two functional predictors (average daily PM2.5 concentration and average daily humidity from 1 January 2015 to 31 December 2020), three scalar predictors (per capita GDP, greenery coverage, and PMP in 2020), and the response is the cancer incidence in 2020. The ratio of the number of new cancer cases to the total number of people in China in 2020 is 0.3156%. The data of the cancer incidence can only contain 0 and 1, indicating high or low cancer incidence rate. When the cancer incidence of a city was less than 0.3156%, the city was considered to have a low cancer incidence rate, denoted by 0; otherwise, the cancer incidence was high, denoted by 1. Figure 4 shows average daily PM2.5 concentration and daily relative humidity in 21 cities selected from the 49 cities.



**Figure 4.** PM2.5 concentrations and average daily humidity in 21 selected cities in 2020.

We chose  $g(x) = \frac{\exp(x)}{1+\exp(x)}$  as the link function. The model was first subjected to principal component analysis and then the number of principal components was determined based on the cumulative contribution to obtain the number of functional principal components for PM2.5 concentrations and relative humidity, which were chosen as  $p_{PM2.5} = 7$ ,  $p_{Humidity} = 14$  in order to explain 75% of the variation.

The prediction accuracy is shown by the Generalized Cross Validation (GCV) with a value of 0.0038.

The results of the regression coefficients for the scalar predictor variable  $\hat{\gamma}$  are shown in Table 4, where we can see that the per capita GDP is positively correlated with the incidence of cancer, i.e., the higher the GDP per capita, the higher the incidence of cancer in that city, which is consistent with the findings of Cao et al. [27]. The reason for this situation is that the promotion of cancer screening, early diagnosis, and treatment in the more economically developed regions has, to some extent, facilitated the detection of the disease. The greenery coverage is negatively correlated with the cancer incidence, i.e., the higher the greenery coverage, the lower the cancer incidence, which is also consistent with the findings of Wu et al. [26]. A high green coverage rate implies better air quality, which in turn reduces the risk of cancer. Additionally, a high green coverage rate may provide more outdoor recreational spaces, promoting physical activity and exercise, contributing to maintaining good physical health, and, thus, reducing the risk of cancer. The PMP is positively correlated with the incidence of cancer. As we all know, cancer incidence is age-related, and older people are more susceptible to cancer. The higher PMP, the better the medical conditions, the longer the average life expectancy of the people, and, therefore, the higher the cancer incidence.

**Table 4.** Estimates of regression coefficients and their levels of significance.

	Estimate	Std. Error	t Value	Pr (>  t )
GDP	$1.165 \times 10^{-6}$	$2.632 \times 10^{-7}$	4.424	$6.45 \times 10^{-5}$ ***
Greenery coverage	$-2.062 \times 10^{-1}$	$1.920 \times 10^{-1}$	-2.654	0.0378 *
PMP	$3.676 \times 10^{-4}$	$1.642 \times 10^{-4}$	2.239	0.0491 *

PMP—Proportion of medical personnel. \* indicates that the result is significant at the 5% significance level. \*\*\* indicates that the result is significant at the 0.1% significance level.

The regression coefficient functions  $\hat{\beta}_1(t_1)$  and  $\hat{\beta}_2(t_2)$  for the functional predictors are shown in Figure 5. From Figure 5, we can see that the effect of PM2.5 concentration on cancer incidence is generally positively correlated, i.e., the higher the PM2.5 concentration, the higher the cancer incidence. This result is consistent with Qin et al. [25] from 2014. Regarding the effect of humidity on cancer incidence, there is a more significant positive correlation between humidity and cancer incidence, i.e., the higher the humidity, the higher the cancer incidence. In high-humidity environments, there may be a higher presence of mold and fungi, and the spores and harmful substances released by these microorganisms may have negative effects on human health, increasing the risk of cancer. In high-humidity environments, pollutants in the air are more likely to adhere to suspended particles, making them more easily inhalable by humans. These pollutants include PM2.5, organic compounds, and heavy metals, which are believed to be associated with the occurrence of cancer. High humidity increases the survival time of bacteria and viruses in the air, increasing the chances of people becoming infected with diseases. Certain viruses such as hepatitis B virus and human papillomavirus (HPV) are believed to be associated with the occurrence of cancer.



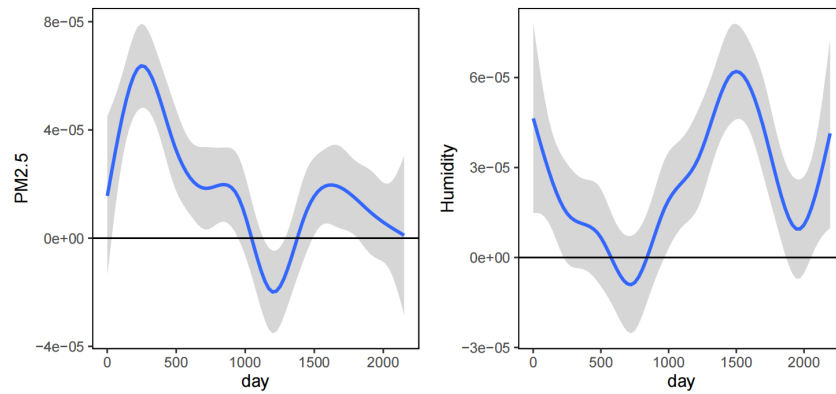


Figure 5. Regression coefficient functions  $\hat{\beta}_1(t_1), \hat{\beta}_2(t_2)$  and their 95% confidence bands

The interaction surface estimate  $\hat{\beta}(t_1, t_2)$  (middle)  $\pm$  two times the estimated standard errors (left and right) are given in Figure 6. Figure 7 shows the contour map of  $\hat{\beta}(t_1, t_2)$ , from which it can be seen that  $\hat{\beta}(t_1, t_2)$  decreases and then increases with  $t_1$  when  $t_2 \in [0, 1100]$  and increases and then decreases with  $t_1$  when  $t_2 \in [1100, 2192]$ . In the conditions of higher humidity, PM2.5 particles may be more prone to settling, reducing the suspended harmful particles in the air, potentially lowering the incidence of cancer. Conversely, in lower-humidity conditions, PM2.5 may be more likely to remain suspended in the air, increasing the risk of respiratory system exposure, thereby raising the incidence of cancer. Additionally, the concentrations of PM2.5 and humidity may not fluctuate synchronously throughout the day. By introducing interaction terms, the model can capture the temporal complexities, making the estimation results more in line with real-world conditions.

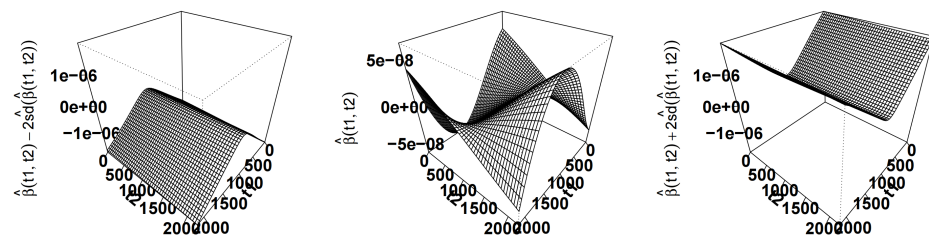


Figure 6. Visualization of  $\hat{\beta}(t_1, t_2)$  in 3D.

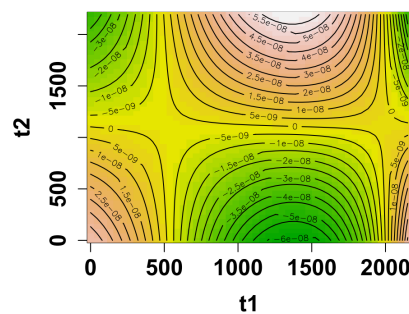


Figure 7. Contour map of  $\hat{\beta}(t_1, t_2)$ .

To verify the necessity of considering the interaction term, i.e., to demonstrate the effectiveness of our proposed method, we compare mod1 proposed in this paper with mod2, which does not include the interaction term, i.e.,

$$\begin{aligned}
 \text{mod1} : Y_i = & g\left(\alpha + \int_{T_1} X_{i1}(t_1)\beta_1(t_1)dt_1 + \int_{T_2} X_{i2}(t_2)\beta_2(t_2)dt_2 \right. \\
 & \left. + \iint_{T_1 \times T_2} X_{i1}(t_1)X_{i2}(t_2)\beta(t_1, t_2)dt_1dt_2 + Z_i^T \gamma\right) + \varepsilon_i.
 \end{aligned}$$

$$\text{mod2} : Y_i = g\left(\alpha + \int_{T_1} X_{i1}(t_1)\beta_1(t_1)dt_1 + \int_{T_2} X_{i2}(t_2)\beta_2(t_2)dt_2 + Z_i^T \gamma\right) + \varepsilon_i.$$

The general standards for evaluating model performance are AIC (Akaike Information Criterion), residual, R-squared, RMSE (root mean square error), and MAE (mean absolute error). The smaller values of AIC, residuals, RMSE, and MAE indicate that the model’s fitting effect and generalization ability are better. The R-squared takes a value between 0 and 1, and, the bigger the value, the better the model’s fitting effect. According to Table 5, we can see that the AIC, residuals, RMSE, and MAE values of mod1 are smaller and that R-squared is much closer to 1 compared to that of mod2, which indicates that mod1 has a better performance. Thus, including the interaction term between PM2.5 concentration and relative humidity makes the research results more meaningful.

**Table 5.** Results of model comparison.

	AIC	R-Squared	Residual	RMSE	MAE
mod1	8.281	0.9287	0.7816	0.1263	0.1036
mod2	35.592	0.6465	3.2158	0.2562	0.1989

### 6. Discussion

This paper proposes a generalized partially functional linear model with interaction terms. We first use principal component analysis to reduce the dimensionality of the functional data, followed by maximum likelihood estimation to obtain the estimates of the unknown parameters, then prove the asymptotic property of the estimators, and finally perform data simulations and apply our model to a real data example.

As the incidence and mortality of cancer in China are increasing year by year, it is necessary to study the influencing factors and formulate corresponding measures. The effect of PM2.5 concentration, average daily humidity, per capita GDP, the greenery coverage of built-up areas, and PMP on cancer incidence in 49 cities in China was investigated, which showed that the effect of PM2.5 concentration and relative humidity on cancer incidence was generally positively correlated. The effect of greenery coverage in built-up areas on cancer incidence is negatively correlated, while the effect of per capita GDP and the proportion of medical personnel on cancer incidence is positively correlated. The higher the economic level and the more developed the medical conditions, the longer the average life expectancy of people and, therefore, the higher the cancer incidence. Comparing this model with the model without the interaction term shows that considering the role of the interaction term leads to more accurate and meaningful predictions.

Our research lays a foundation for further study on the generalized partially functional linear model with interaction terms and of unknown link function or variance function.

**Author Contributions:** W.X.: methodology, validation, writing—review, supervision, funding acquisition. K.M.: 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. 107051360024XN153).

**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 to improve the paper.

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

## References

1. Ramsay, J.O.; Silverman, B.W. *Functional Data Analysis*, 2nd ed.; Springer: New York, NY, USA, 2005.
2. Horváth, L.; Kokoszka, P. *Inference for Functional Data with Application*; Springer: New York, NY, USA, 2012.
3. Hsing, T.; Eubank, R. *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*; Springer: New York, NY, USA, 2015.
4. Cardot, H.; Ferraty, F.; Sarda, P. Functional linear model. *Stat. Probab. Lett.* **1999**, *45*, 11–22. [[CrossRef](#)]
5. Tony, C.; Peter, H. Prediction in functional linear regression. *Ann. Stat.* **2006**, *34*, 2159–2179. [[CrossRef](#)]
6. Cardot, H.; Crambes, C.; Kneip, A.; Sarda, P. Smoothing Splines Estimators in Functional Linear Regression with Errors in Variables. *Comput. Stat. Data Anal.* **2007**, *51*, 4832–4848. [[CrossRef](#)]
7. Delaigle, A.; Hall, P. Methodology and theory for partial least squares applied to functional data. *Ann. Stat.* **2012**, *40*, 322–352. [[CrossRef](#)]
8. Cai, T.; Yuan, M. Minimax and adaptive prediction for functional linear regression. *J. Am. Stat. Assoc.* **2012**, *107*, 1201–1216. [[CrossRef](#)]
9. Nelder, J.A.; Wedderburn, R.W.M. Generalized Linear Models. *J. R. Stat. Soc.* **1972**, *135*, 370–384. [[CrossRef](#)]
10. James, G.M. Generalized linear models with functional predictors. *J. R. Stat. Soc. Ser. B* **2002**, *64*, 411–432. [[CrossRef](#)]
11. Müller, H.G.; Stadtmüller, U. Generalized functional linear models. *Ann. Stat.* **2005**, *33*, 774–805. [[CrossRef](#)]
12. Goldsmith, J.; Bobb, J.; Crainiceanu, C.M.; Caffo, B.; Reich, D. Penalized Functional Regression. *J. Comput. Graph. Stat.* **2011**, *20*, 830–851. [[CrossRef](#)]
13. Xiao, W.W.; Wang, Y.X.; Liu, H.Y. Generalized partially functional linear model. *Sci. Rep.* **2021**, *11*, 23428. [[CrossRef](#)]
14. Kong, D.; Xue, K.; Yao, F.; Zhang, H.H. Partially functional linear regression in high dimensions. *Biometrika* **2016**, *103*, 147–159. [[CrossRef](#)]
15. Yao, F.; Sue, C.S.; Wang, F. Regularized partially functional quantile regression. *J. Multivar. Anal.* **2017**, *156*, 39–56. [[CrossRef](#)]
16. Ma, H.; Li, T.; Zhu, H.; Zhu, Z. Quantile regression for functional partially linear model in ultra-high dimensions. *Comput. Stat. Data Anal.* **2019**, *129*, 135–147. [[CrossRef](#)]
17. Xu, W.; Ding, H.; Zhang, R.; Liang, H. Estimation and inference in partially functional linear regression with multiple functional covariates. *J. Stat. Plan. Inference* **2020**, *209*, 44–61. [[CrossRef](#)]
18. Usset, J.; Staicu, A.M.; Maity, A. Interaction models for functional regression. *Comput. Stat. Data Anal.* **2016**, *94*, 317–329. [[CrossRef](#)]
19. Luo, R.; Qi, X. Interaction Model and Model Selection for Function-on-Function Regression. *J. Comput. Graph. Stat.* **2019**, *28*, 309–322. [[CrossRef](#)]
20. Yang, W.H.; Winkle, C.K.; Holan, S.H.; Wildhaber, M.L. Ecological Prediction with Nonlinear Multivariate Time-Frequency Functional Data Models. *J. Agric. Biol. Environ. Stat.* **2013**, *18*, 450–474. [[CrossRef](#)]
21. Matsui, H. Quadratic regression for functional response models. *Econom. Stat.* **2020**, *13*, 125–136. [[CrossRef](#)]
22. Sun, Y.; Wang, Q. Function-on-function quadratic regression models. *Comput. Stat. Data Anal.* **2020**, *142*, 106814. [[CrossRef](#)]
23. Fuchs, K.; Scheipl, F.; Greven, S. Penalized scalar-on-functions regression with interaction term. *Comput. Stat. Data Anal.* **2015**, *81*, 38–51. [[CrossRef](#)]
24. Qiu, H.; Cao, S.; Xu, R. Cancer incidence, mortality, and burden in China: A time-trend analysis and comparison with the United States and United Kingdom based on the global epidemiological data released in 2020. *Cancer Commun.* **2021**, *41*, 1037–1048. [[CrossRef](#)]
25. Qin, X.; Wan, F.; Zhang, H.; Dai, B.; Shi, G.; Zhu, Y.; Ye, D. Relationship between air pollution PM2.5 concentration and cancer. In Proceedings of the 8th Chinese Oncology Academic Conference and the 13th Cross-Strait Oncology Academic Conference, Jinan, China, 12–13 September 2014; p. 76.
26. Wu, X.; Feng, Y.; Chang, K.; Jia, X.; Xue, F. Analysis of the causal relationship between green coverage and the incidence of cancer. *J. Shandong Univ. (Health Sci.)* **2022**, *60*, 115–119.
27. Cao, W.; Li, F.; Liang, Y.; Yu, D. Analysis of the relationship between the level of economic development and cancer incidence and mortality in selected regions of China. *Chin. J. Dis. Control Prev.* **2023**, *27*, 209–215. [[CrossRef](#)]

28. Xu, J.; Kuang, H.Y.; Wang, G.Q.; Chen, M.; Lu, L. Analysis of the relationship between PM2.5 and air relative humidity. *Agric. Technol.* **2017**, *37*, 148–149.
29. Yang, Z.; Wang, Y.K.; Xu, X.H.; Yang, J.; Ou, C.Q. Quantifying and characterizing the impacts of PM2.5 and humidity on atmospheric visibility in 182 Chinese cities: A nationwide time-series study. *J. Clean. Prod.* **2022**, *368*, 133182. [[CrossRef](#)]

**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.