

# 带有交互项且连接函数未知的广义函数型线性模型研究

刘依林<sup>1</sup>, 张博文<sup>2</sup>, 肖维维<sup>1\*</sup>

<sup>1</sup>北方工业大学理学院, 北京

<sup>2</sup>南京理工大学数学与统计学院, 江苏 南京

收稿日期: 2026年2月23日; 录用日期: 2026年3月18日; 发布日期: 2026年3月25日

## 摘要

本文提出一种广义部分函数型线性模型。该模型不仅能够同时处理标量型与函数型预测变量, 还通过引入未知的连接函数并考虑函数型预测变量间的交互作用, 以更贴合实际数据的复杂结构。具体而言, 本文首先采用函数型主成分分析对高维函数型数据进行降维。在参数估计上, 我们提出一种迭代算法: 在假设连接函数已知的条件下, 利用极大似然估计法估计回归系数; 随后, 在回归系数固定的条件下, 采用局部线性核回归方法估计连接函数。二者交替迭代, 直至收敛, 从而同步获得回归系数与连接函数的最终估计。为验证模型的有效性, 我们将其应用于肉类蛋白质含量高低的分类研究。实证结果表明, 该模型能有效揭示预测变量与响应变量之间的复杂关系。模型的五折交叉验证准确率达到97.03%, 性能显著优于连接函数已知的带交互项模型(准确率94.35%)及不带交互项模型(准确率90.45%), 展现了其优越的预测能力。

## 关键词

广义部分函数型线性模型, 函数型主成分分析, 肉类蛋白质含量

# Research on the Generalized Functional Linear Model with Interaction Terms and an Unknown Link Function

Yilin Liu<sup>1</sup>, Bowen Zhang<sup>2</sup>, Weiwei Xiao<sup>1\*</sup>

<sup>1</sup>Department of Computer Science, North China University of Technology, Beijing

<sup>2</sup>School of Mathematics and Statistics, Nanjing University of Science and Technology, Nanjing Jiangsu

Received: February 23, 2026; accepted: March 18, 2026; published: March 25, 2026

\*通讯作者。

文章引用: 刘依林, 张博文, 肖维维. 带有交互项且连接函数未知的广义函数型线性模型研究[J]. 应用数学进展, 2026, 15(3): 529-540. DOI: 10.12677/aam.2026.153124

## Abstract

In this paper, a generalized partial functional linear model is proposed, which not only accommodates composite predictors including scalar and functional variables, but also incorporates an unknown link function and accounts for the interaction effects among functional predictors, thus achieving a better fit to the actual characteristics of the data. Specifically, functional principal component analysis is first adopted for dimensionality reduction. Under the assumption of a known link function, the maximum likelihood estimation method is employed to estimate the regression coefficients; subsequently, with the estimated regression coefficients fixed, the local linear kernel regression is applied to estimate the unknown link function. This iterative alternating estimation procedure is repeated until convergence, and the final estimates of the regression coefficients and the link function are thus obtained. To verify the good validity and performance of the proposed model, it is applied to the research on the protein content of meat products. The empirical results demonstrate that the model can effectively characterize the relationship between the predictors and the response variable, with an accuracy of 97.03% achieved by five-fold cross-validation. In comparison with the model with a known link function and interaction terms (with an accuracy of 94.35%) and the model with a known link function but without interaction terms (with an accuracy of 90.45%), the proposed model exhibits superior performance and a higher prediction accuracy.

## Keywords

Generalized Partial Functional Linear Model, Functional Principal Component Analysis, Protein Content of Meat Products

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

近年来,在统计学与数据科学迅速发展的背景下,函数型数据分析成为应对高维、连续时间或空间观测数据挑战的重要工具。在实际应用中,研究者经常面临以函数或曲线形式呈现的变量,这类数据具有复杂的结构特征,传统的多元回归模型难以对其进行有效的建模与解释。

为了解决上述问题,1972年由Nelder和Wedderburn [1]首次引入函数型线性模型,分别研究了连续型和离散型响应变量与预测变量之间的关系。但是传统的广义线性模型在处理复杂数据时,模型的预设假设可能难以精确地契合数据的真实分布。针对这一问题,Scalla等人(1984) [2]提出了一种参数化连接函数的方法。该方法事先设定连接函数的形式,并通过数据来估计这些参数,从而调整模型的拟合效果,提高预测精度。然而,在实际应用层面,通过引入参数来优化连接函数的方式存在其固有的局限性。因为我们通常很难事先确定连接函数的具体形式。另外,由于数据的复杂性和独特性,标准的连接函数可能无法准确地拟合数据,从而限制模型的预测能力。而Weisberg和Welsh (1994) [3]结合了Wedderburn (1974) [4]提出的拟似然函数,在连接函数未知的广义线性模型拟合中,使用核平滑估计连接函数,再通过估计出的连接函数估计回归系数,利用交替迭代的方式,有效地解决了连接函数未知时的拟合难题。Chiou和Müller (1998) [5]对未知的方差函数和连接函数进行非参数估计,并将估计代入拟似然函数中,运用局部多项式拟合得到了连接函数估计和方差函数估计的一致性结果,以及回归系数的渐近极限分布。Cardot和Crambes等人[6]将总体最小二乘法推广到函数型线性模型中,提出了模型函数系数的光滑样条

估计, 并给出了该估计的渐近结果。Goldsmith 等人[7]提出了函数型数据广义线性模型的快速拟合方法, 通过对大量光滑特征向量进行投影, 并利用惩罚样条回归估计系数函数。Fuchs 等人[8]将具有函数型预测变量的标量响应的广义模型扩展到包含线性泛函交互项, 系数函数的估计使用基展开式和最大似然估计法。

基于以上研究背景, 本文在前人研究的基础上, 构建了连接函数未知且含有函数型预测变量交互项的广义函数型线性模型, 该模型不仅引入了未知的连接函数, 还考虑到了函数型预测变量间的交互作用。另外, 该模型还具有同时处理函数型预测变量和标量型预测变量的能力, 能够更准确的阐述响应变量与预测变量之间的关系, 具有重要的理论价值和现实意义。

## 2. 广义部分函数型线性模型

### 2.1. 模型介绍

假设有标量型响应变量  $Y_i$ ,  $i=1, \dots, n$ ,  $X_{ij}(t_j)$ ,  $j=1, 2$  是影响  $Y_i$  的函数型预测变量, 其中  $X_{ij}(t_j) \in L^2(T_j)$ ,  $T_j$  是  $R$  上的一个有界区间,  $q$  维向量  $Z = (Z_1, Z_2, \dots, Z_q)$  是影响  $Y_i$  的标量型预测变量, 响应变量  $Y_i$  和预测变量  $X_{ij}(t_j)$ 、 $Z$  有以下关系:

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

其中,  $\alpha \in R$  是截距,  $\beta_1(t_1)$ 、 $\beta_2(t_2)$  和  $\beta(t_1, t_2)$  是两个函数型预测变量和交互项对应的未知回归系数函数,  $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_q)^T$  是标量型预测变量  $Z$  对应的未知回归系数,  $g(\cdot)$  是未知的连接函数, 假设  $\varepsilon_i$  是独立同分布的, 且符合正态分布, 均值为 0, 方差为  $\sigma^2$ 。

对于函数型预测变量, 我们首先利用函数型主成分分析的方法对其进行降维。假设函数型预测变量  $X_{ij}(t_j)$  和标量型预测变量  $Z$  均已中心化。通过 Karhunen-Loeve 展开和 Mercer 定理, 将函数型预测变量  $X_{ij}(t)$  展开为

$$X_{i1}(t) = \sum_{k=1}^{\infty} \chi_{ik} \varphi_{1k}(t) \quad (2)$$

$$X_{i2}(t) = \sum_{l=1}^{\infty} \chi_{i2l} \varphi_{2l}(t) \quad (3)$$

其中,  $\chi_{ik}$ 、 $\chi_{i2l}$  为函数型主成分得分,  $\varphi_{1k}(t)$ 、 $\varphi_{2l}(t)$  为函数型主成分基, 且为  $L^2(T)$  空间中的标准正交基。

同理, 将回归系数函数  $\beta_j(t)$ 、 $\beta(t_1, t_2)$  展开为

$$\beta_1(t) = \sum_{k=1}^{\infty} b_{1k} \varphi_{1k}(t) \quad (4)$$

$$\beta_2(t) = \sum_{l=1}^{\infty} b_{2l} \varphi_{2l}(t) \quad (5)$$

$$\beta(t_1, t_2) = \sum_{k=1}^{\infty} \sum_{l=1}^{\infty} u_{kl} \varphi_{1k}(t_1) \varphi_{2l}(t_2) \quad (6)$$

其中,  $b_{1k}$ 、 $b_{2l}$  是回归系数函数的系数,  $u_{kl}$  是交互项的回归系数函数的系数。

将上述展开式代入模型(1)中, 并将函数型预测变量在  $m_j$  处截断, 其中  $m_j$  取决于样本容量  $n$ , 且随着

$n \rightarrow \infty$  而不断增加, 得到

$$Y_i = g \left( \alpha + \sum_{k=1}^{m_1} \chi_{i1k} b_{1k} + \sum_{l=1}^{m_2} \chi_{i2l} b_{2l} + \sum_{k=1}^{m_1} \sum_{l=1}^{m_2} \chi_{i1k} \chi_{i2l} u_{kl} + Z^T \gamma \right) + \varepsilon_i \quad (7)$$

## 2.2. 回归系数与连接函数的估计

定义参数向量

$$\Omega = (b_{11}, \dots, b_{1m_1}, b_{21}, \dots, b_{2m_2}, u_{11}, \dots, u_{1m_2}, u_{21}, \dots, u_{2m_2}, \dots, u_{m_11}, \dots, u_{m_1m_2}, \gamma_0, \gamma_1, \dots, \gamma_q)^T$$

$$\eta = \alpha + \sum_{k=1}^{m_1} \chi_{i1k} b_{1k} + \sum_{l=1}^{m_2} \chi_{i2l} b_{2l} + \sum_{k=1}^{m_1} \sum_{l=1}^{m_2} \chi_{i1k} \chi_{i2l} u_{kl} + Z^T \gamma \quad (8)$$

$$\mu_i = g(\eta_i) \quad (9)$$

$$\rho_{ikl} = \chi_{i1k} \chi_{i2l} \quad (10)$$

$$\omega_i = (\chi_{i11}, \dots, \chi_{i1m_1}, \chi_{i21}, \dots, \chi_{i2m_2}, \rho_{i11}, \dots, \rho_{i1m_2}, \dots, \rho_{im_11}, \dots, \rho_{im_1m_2}, z_{i0}, z_{i1}, \dots, z_{iq})^T$$

对于未知的回归系数向量  $\Omega$  和连接函数  $g$ , 我们采取一种迭代交替更新的方法来进行估计。下面简单叙述迭代过程。

步骤一: 假设连接函数为恒等映射, 利用极大似然估计来估计参数向量  $\Omega$ 。样本的似然函数为

$$L(\Omega) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\varepsilon_i^2}{2\sigma^2}} = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(Y_i - \mu_i)^2}{2\sigma^2}} \quad (11)$$

对似然函数取对数并对  $\Omega$  求偏导, 再令偏导为 0, 即求解下列方程, 可求出参数向量的极大似然估计  $\Omega^{(0)}$ 。

$$\sum_{i=1}^n \frac{(Y_i - g(\eta_i))}{\sigma^2} \cdot g'(\eta_i) \cdot \omega_i = 0 \quad (12)$$

步骤二: 在估计出的参数向量  $\Omega^{(0)}$  的基础上, 利用局部线性核回归去估计连接函数  $g$  及其导数  $g'$ 。设核函数为  $K(\cdot)$ , 带宽为  $h$ , 定义与带宽有关的核函数为  $K_h(\cdot)$ , 本文选取的核函数为高斯核函数。设函数型预测变量  $X_{ij}(t_j)$  和标量型预测变量  $Z$  都属于一个紧集  $U$ , 则  $P = \{x = \eta_i | X_{ij}(t_j), Z \in U\}$ , 为了估计  $g$ 、 $g'$ , 构造任意  $x$  处的加权平方和为

$$\arg \min_{(g, g')} \sum_{i=1}^n K_h(\eta_i - x) [Y_i - g - g'(\eta_i - x)]^2 \quad (13)$$

化为矩阵形式, 即

$$\arg \min_{(g, g')} \sum_{i=1}^n K_h(\eta_i - x) [Y_i - g - g'(\eta_i - x)]^2 = (Y - X\zeta)^T W (Y - X\zeta) \quad (14)$$

其中,

$$W = \text{diag}(K_h(\eta_1 - x), \dots, K_h(\eta_n - x)), \quad Y = (Y_1, Y_2, \dots, Y_n)^T, \quad \zeta = (g, g')^T, \quad X = (1, \dots, \eta_i - x)$$

对  $\zeta$  求导并令其为 0, 则

$$\zeta = (X^T W X)^{-1} X^T W Y \quad (15)$$

即解出  $g$  与  $g'$ 。

步骤三: 运用步骤一的方法, 将连接函数替换为估计出的连接函数  $g$  和  $g'$ , 通过求解含有  $\Omega$  的方程

来更新  $\Omega^{(\alpha)}$ ，其中  $\alpha=1,2,\dots$  表示迭代次数，得到  $\Omega$  新的估计值。

步骤四：运用步骤二的方法，将参数向量替换为估计后的  $\Omega^{(\alpha)}$ ，其中  $\alpha=1,2,\dots$ ，由此得到连接函数  $g$  和  $g'$  新的估计值  $g^{(\alpha)}$  和  $g'^{(\alpha)}$ 。

步骤五：重复上述步骤，直到  $|\Omega^{(\alpha+1)} - \Omega^{(\alpha)}|$  收敛，停止迭代更新。

步骤六：得到最终的参数向量的估计值  $\hat{\Omega}$ ，连接函数  $g$  的估计值  $\hat{g}$ 。

### 3. 数据模拟

我们考虑两个函数型预测变量、三个标量型预测变量以及二元 0-1 型响应变量的情况。对于函数型预测变量  $X_{i1}(t_1)$ ， $t_1 \in [0,1]$ ， $X_{i2}(t_2)$ ， $t_2 \in [-1,1]$ ， $i=1,\dots,n$ ， $n$  可以为任意正整数，这里我们取  $n$  为 500。

定义两个标准正交基为  $\varphi_{1k}(t_1)$ ， $t_1 \in [0,1]$  和  $\varphi_{2l}(t_2)$ ， $t_2 \in [-1,1]$ 。

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

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

生成两个函数型主成分得分

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

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

其中  $\lambda_{11} = \sqrt{2}$ ， $\lambda_{12} = 1$ ， $\lambda_{13} = \frac{\sqrt{2}}{2}$ ， $\lambda_{14} = \frac{1}{2}$ ， $\lambda_{21} = 2$ ， $\lambda_{22} = \sqrt{2}$ ， $\lambda_{23} = 1$ ， $\lambda_{24} = \frac{\sqrt{2}}{2}$ ， $\lambda_{25} = \frac{1}{2}$ 。

所以有

$$X_{i1}(t_1) = \sum_{k=1}^4 \chi_{i1k} \varphi_{1k}(t_1), \quad (16)$$

$$X_{i2}(t_2) = \sum_{l=1}^5 \chi_{i2l} \varphi_{2l}(t_2), \quad (17)$$

函数型预测变量  $X_1(t_1)$  和  $X_2(t_2)$  的图像如图 1 和图 2 所示：

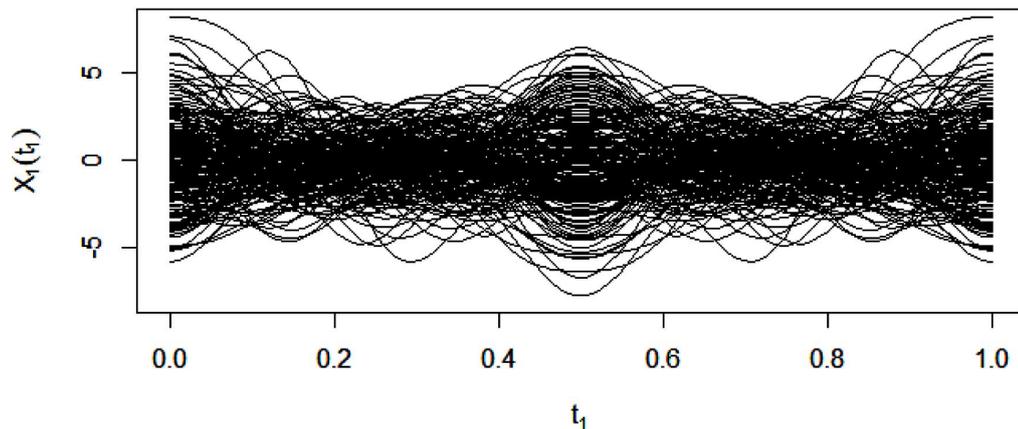


Figure 1. Graph of the functional predictor  $X_1(t_1)$

图 1. 函数型预测变量  $X_1(t_1)$  的图像

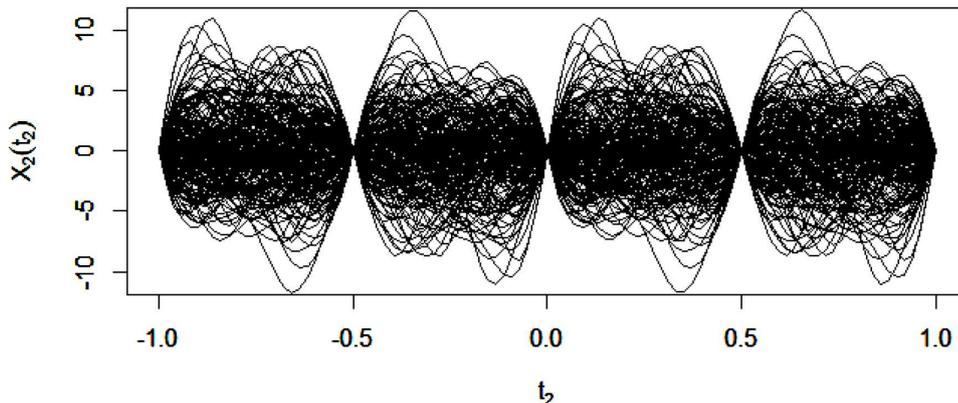


Figure 2. Graph of the functional predictor  $X_2(t_2)$

图 2. 函数型预测变量  $X_2(t_2)$  的图像

对于标量型预测变量  $Z$ , 满足  $Z_1 \sim N(0,1)$ ,  $Z_2 \sim N\left(0, \frac{\sqrt{3}}{3}\right)$ ,  $Z_3 \sim N\left(0, \frac{\sqrt{5}}{5}\right)$ 。假设标量型预测变量的回归系数理论值为  $\gamma_1 = \frac{\sqrt{2}}{2}$ ,  $\gamma_2 = \frac{\sqrt{3}}{3}$ ,  $\gamma_3 = \frac{\sqrt{5}}{5}$ 。

假设函数型预测变量的回归系数函数为

$$\beta_1(t_1) = \sum_{k=1}^4 b_{1k} \varphi_{1k}(t_1), \tag{18}$$

$$\beta_2(t_2) = \sum_{l=1}^5 b_{2l} \varphi_{2l}(t_2), \tag{19}$$

其中  $b_{11} = \sqrt{\frac{1}{3}}$ ,  $b_{12} = \sqrt{\frac{1}{6}}$ ,  $b_{13} = \frac{1}{3}$ ,  $b_{14} = \sqrt{\frac{1}{12}}$ ,  $b_{21} = \frac{1}{3}$ ,  $b_{22} = \frac{1}{6}$ ,  $b_{23} = \frac{1}{9}$ ,  $b_{24} = \frac{1}{12}$ ,  $b_{25} = \frac{1}{15}$ 。

对于交互项  $\beta(t_1, t_2)$ , 回归系数函数的系数记为  $u_{kl}$ , 且有

$$u_{kl} = \left(\frac{k}{10}\right)^2, k = 1, 2, \dots, 20$$

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

交互项的主成分基为  $\phi_{kl}(t_1, t_2)$ , 满足

$$\phi_{kl}(t_1, t_2) = \varphi_{1k}(t_1) \varphi_{2l}(t_2), \tag{20}$$

因此交互项可以表示为

$$\beta(t_1, t_2) = \sum_{k=1}^4 \sum_{l=1}^5 u_{kl} \varphi_{1k}(t_1) \varphi_{2l}(t_2) \tag{21}$$

选择真实的连接函数为  $g(x) = \frac{\exp(x)}{1 + \exp(x)}$ , 生成与函数型预测变量和标量型预测变量有关的二元响应变量  $Y(X, Z)$ , 且有

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

其中  $Y(X, Z)$  服从伯努利分布。

我们选取样本量  $n$  为 500。图 3 和图 4 展示了样本量为 500 情况下，估计出的  $\hat{\beta}_1(t_1)$ ， $\hat{\beta}_2(t_2)$  和其相应的 95% 置信区间带情况，可以看出样本量为 500 时，估计值已经很接近真实值。

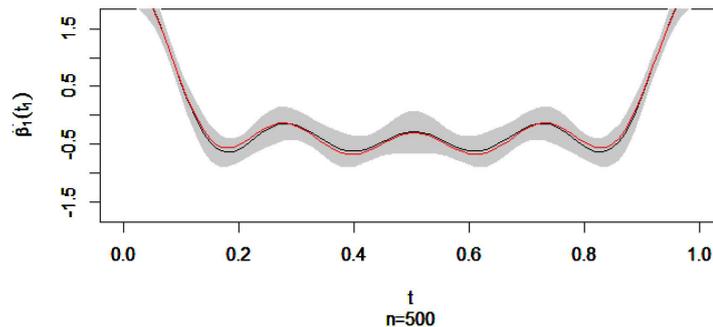


Figure 3. Graph of the regression coefficient function  $\beta_1(t_1)$

图 3. 回归系数函数  $\beta_1(t_1)$  的图像

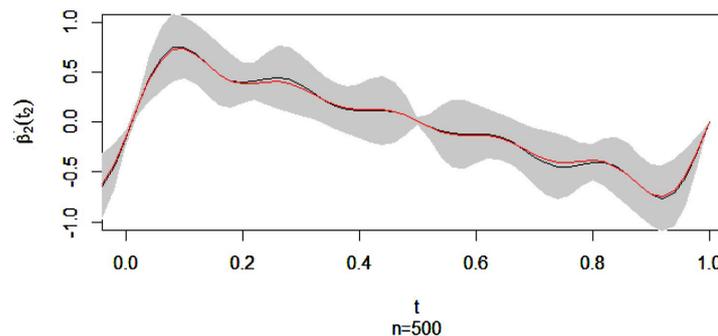


Figure 4. Graph of the regression coefficient function  $\beta_2(t_2)$

图 4. 函数型预测变量  $\beta_2(t_2)$  的图像

黑色曲线是回归系数函数估计  $\hat{\beta}_1(t_1)$ ， $\hat{\beta}_2(t_2)$ ，灰色区域是他们的 95% 置信区间带，红色曲线是  $\hat{\beta}_1(t_1)$ ， $\hat{\beta}_2(t_2)$  的真实值。

图 5 展示了  $n$  为 500 时估计出的连接函数  $\hat{g}$ ，其中蓝色代表真实的连接函数，红色代表估计出的连接函数，可以看出，估计出的连接函数已经很接近真实值。

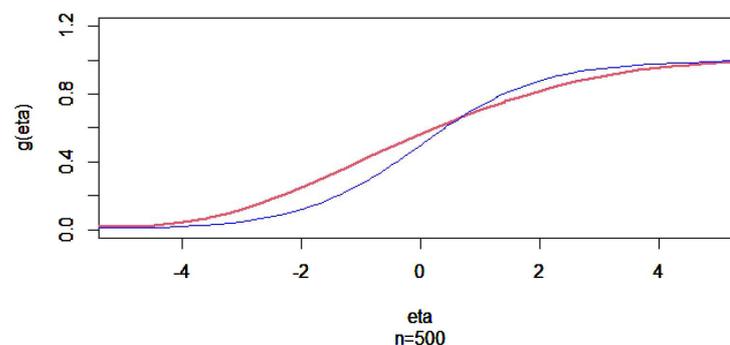


Figure 5. Plot of the true link function and the estimated link function

图 5. 真实连接函数与估计出的连接函数图像

图 6 展示了  $n$  为 500 时交互项回归系数函数  $\beta(t_1, t_2)$  的估计值与真实值作差的等高线图, 可以看出其差值大部分在  $[-0.5, 0.5]$  之间。

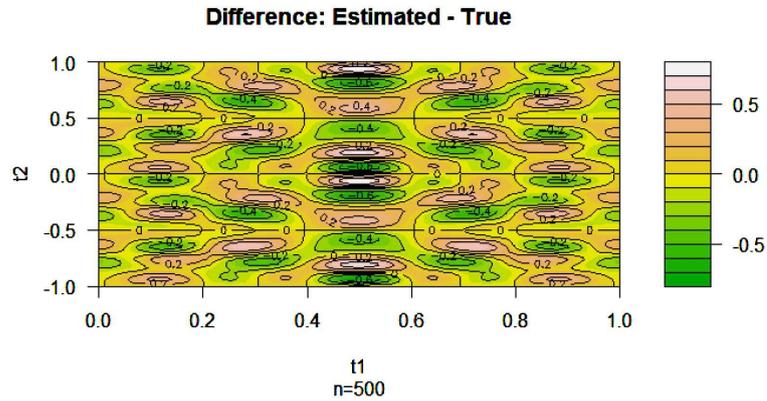


Figure 6. Difference between the estimated surface and the true surface  
图 6. 估计曲面与真实曲面的差值

表 1 显示了  $n$  为 500 时  $\gamma$  和其估计值  $\hat{\gamma}$  的值。通过表中数据可以看出,  $n$  为 500 时估计值接近真实值。

Table 1. Comparison between estimated and true values of scalar predictor

表 1. 标量型预测变量的估计值与真实值对比

$n$	$\hat{\gamma}_1(\gamma_1)$	$\hat{\gamma}_2(\gamma_2)$	$\hat{\gamma}_3(\gamma_3)$
500	0.7055 (0.7071)	0.5671 (0.5774)	0.4898 (0.4472)

#### 4. 实例研究

本节所使用的数据源自 R 语言中的 Tecator 数据集。选取其中的 198 个肉类样本, 其中蛋白质含量高于 18% 记为 1, 低于 18% 记为 0, 以此构成二元响应变量  $Y$ 。将 198 个肉类样本在 850~1050 波长范围内的光谱曲线数据作为第一个函数型预测变量  $X_1$ , 将其导数作为第二个函数型预测变量  $X_2$ ; 将每个样本光谱曲线对应的 MEA (均值)、SKE (偏度)、KUR (峰度) 作为三个标量型预测变量, 分别记为  $Z_1$ 、 $Z_2$ 、 $Z_3$ 。

图 7 和图 8 分别展示了两个函数型预测变量的图像。

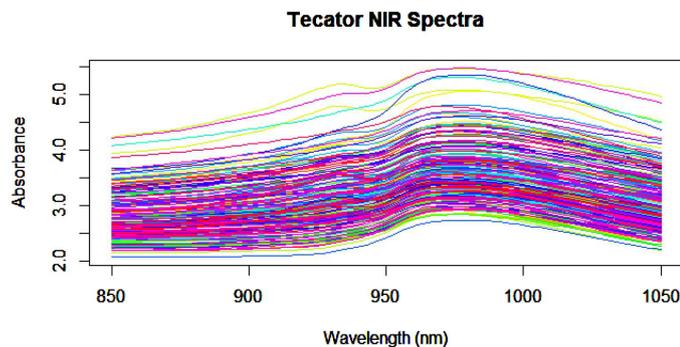
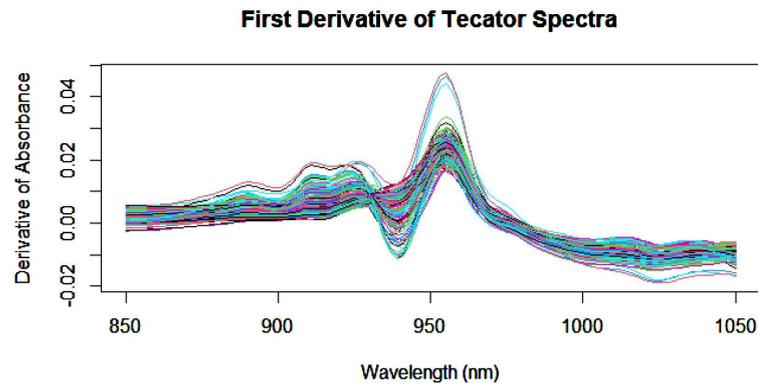


Figure 7. Near-infrared spectral curves of meat samples  
图 7. 肉类近红外光谱曲线



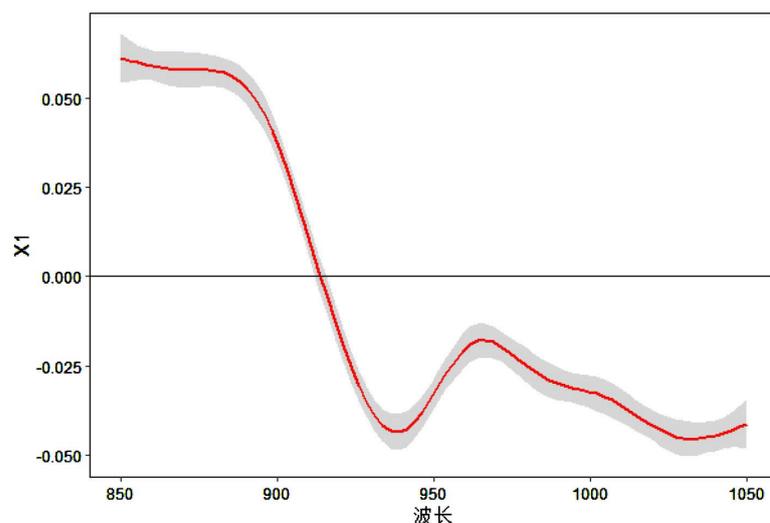
**Figure 8.** First-order derivative curves of near-infrared spectra of meat samples  
**图 8.** 肉类近红外光谱一阶导数曲线

将数据代入模型(1)中, 选取累计贡献率 85%的函数型主成分作截断, 则两个函数型预测变量的主成分基截断个数分别为  $m_{\text{Abs}} = 2$ ,  $m_{\text{Der}} = 2$ 。标量型预测变量回归系数的结果如表 2 所示。实验结果说明 MEA (均值)、SKE (偏度)、KUR (峰度)均与肉类蛋白质含量呈负相关, 即这三个值越大, 则对应的肉类蛋白质含量越低。

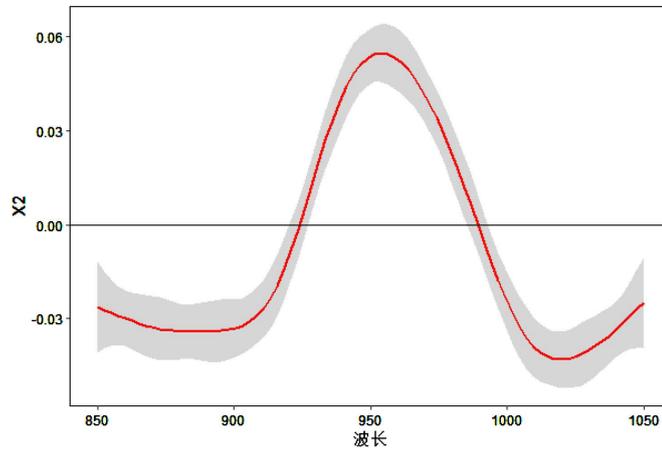
**Table 2.** Estimated values of scalar predictors  
**表 2.** 标量型预测变量的估计值

$\hat{\gamma}$	Estimate	Std. Error	t value	Pr (> t )
$\hat{\gamma}_{\text{MEA}}$	-0.2528	0.1673	-1.5111	0.13241
$\hat{\gamma}_{\text{SKE}}$	-1.0363	0.3350	-3.093	0.00228**
$\hat{\gamma}_{\text{KUR}}$	-1.9254	0.6303	-3.055	0.00257**

图 9 和图 10 分别是两个函数型预测变量回归系数函数的图像。



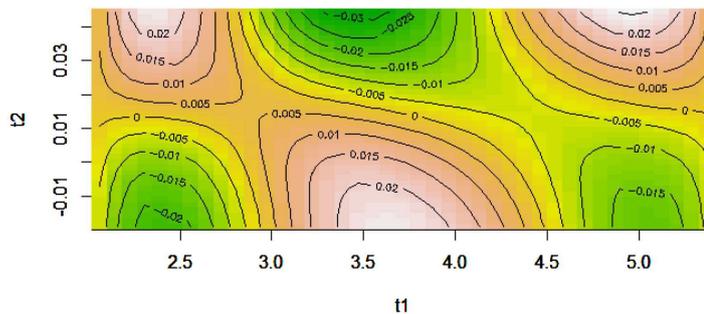
**Figure 9.** Regression coefficient of the functional predictor  $X_1$   
**图 9.** 函数型预测变量  $X_1$  的回归系数



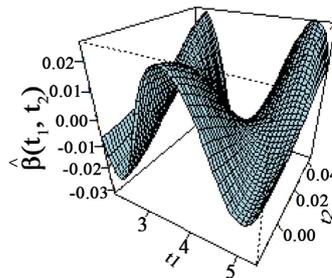
**Figure 10.** Regression coefficient of the functional predictor  $X_2$   
**图 10.** 函数型预测变量  $X_2$  的回归系数

结论表明肉类的近红外光谱曲线与其蛋白质含量有明显的相关性。由图 9 可以看出肉类在波长为约 850~920 nm 时的近红外光吸收与其蛋白质含量呈正相关，在波长为 920~1050 nm 处的近红外光吸收与其蛋白质含量呈负相关。由图 10 可以看出在 920~990 nm 左右的波长段，近红外光谱数据的一阶导数与样本肉类蛋白质含量呈正相关，吸收斜率变化越大，对应肉类样本中的蛋白质含量越高，在其余波长段则呈负相关。因此肉类加工企业、品质检测机构等可利用此模型快速估测肉类样本蛋白质含量的高低，有效降低检测成本。

图 11 和图 12 分别展示了近红外光谱曲线及其导数交互项的回归系数的等高线图和三维曲面图。从图中可看出二者之间明显存在交互效应，对肉类蛋白质含量高低判断产生影响。



**Figure 11.** Contour plot of the interaction term  
**图 11.** 交互项的等高线图



**Figure 12.** 3D surface of the interaction term  
**图 12.** 交互项的三维曲面

## 5. 模型对比

为验证本文所提模型的有效性,我们设计了一项对比实验,具体包含两个目标:(1) 检验在连接函数未知时模型的性能;(2) 评估在模型中考虑函数型预测变量交互项的必要性。

为此,我们将所提模型与以下两个基准模型进行比较。

模型 1 [9]: 连接函数已知,且包含交互项:

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

其中  $g$  已知:

模型 2 [10]: 连接函数已知,但不包含交互项。

$$Y = g\left(\alpha + \int_{T_1} X_{i1}(t_1)\beta_1(t_1)dt_1 + \int_{T_2} X_{i2}(t_2)\beta_2(t_2)dt_2\right) + \varepsilon_i$$

模型评估采用以下两类指标:首先,在拟合优度方面,选用赤池信息准则(AIC)、残差与决定系数(R-squared)三个指标;其次,在泛化能力方面,通过五折交叉验证计算各模型的预测性能指标。其中 AIC 的值和残差值越小说明模型的拟合效果和泛化能力越好,而 R-squared 的值越大说明模型拟合效果越好。如表 3 所示,本文提出的模型在多项指标上均优于对比模型:其 AIC 值与残差和更小, R-squared 值更高,且预测准确率也更具优势。这些结果一致表明,本文模型具有更优的拟合性能。

**Table 3.** Model comparison results

**表 3.** 模型对比结果

模型	AIC	R-squared	残差	准确率
本文模型	271.5259	0.9258	1.2942	97.03%
模型 1	279.8883	0.9014	1.8448	94.35%
模型 2	292.0326	0.6731	2.2093	90.45%

## 6. 结论

本研究旨在建立一类能够刻画函数型预测变量交互作用、且连接函数形式未知的广义函数型线性模型,并应用于肉类蛋白质含量的分析。实证结果表明:肉类样本在 850~1050 nm 波长区间内的近红外光谱曲线及其一阶导数,均与蛋白质含量显著相关;此外,光谱曲线的均值、偏度与峰度等统计量亦与蛋白质含量呈负相关。因此,在评估肉类蛋白质含量时,可综合依据上述光谱及衍生特征进行判定。

## 参考文献

- [1] Nelder, J.A. and Wedderburn, R.W.M. (1972) Generalized Linear Models. *Journal of the Royal Statistical Society. Series A (General)*, **135**, 370-384. <https://doi.org/10.2307/2344614>
- [2] Scallan, A., Gilchrist, R. and Green, M. (1984) Fitting Parametric Link Functions in Generalised Linear Models. *Computational Statistics & Data Analysis*, **2**, 37-49. [https://doi.org/10.1016/0167-9473\(84\)90031-8](https://doi.org/10.1016/0167-9473(84)90031-8)
- [3] Weisberg, S. and Welsh, A.H. (1994) Adapting for the Missing Link. *The Annals of Statistics*, **22**, 1674-1700. <https://doi.org/10.1214/aos/1176325749>
- [4] Wedderburn, R.W.M. (1974) Quasi-Likelihood Functions, Generalized Linear Models, and the Gauss—Newton Method. *Biometrika*, **61**, 439-447. <https://doi.org/10.1093/biomet/61.3.439>
- [5] Chiou, J. and Müller, H. (1998) Quasi-Likelihood Regression with Unknown Link and Variance Functions. *Journal of*

- the American Statistical Association*, **93**, 1376-1387. <https://doi.org/10.1080/01621459.1998.10473799>
- [6] Cardot, H., Crambes, C., Kneip, A. and Sarda, P. (2007) Smoothing Splines Estimators in Functional Linear Regression with Errors-In-Variables. *Computational Statistics & Data Analysis*, **51**, 4832-4848. <https://doi.org/10.1016/j.csda.2006.07.029>
- [7] Goldsmith, J., Bobb, J., Crainiceanu, C.M., Caffo, B. and Reich, D. (2011) Penalized Functional Regression. *Journal of Computational and Graphical Statistics*, **20**, 830-851. <https://doi.org/10.1198/jcgs.2010.10007>
- [8] Fuchs, K., Scheipl, F. and Greven, S. (2015) Penalized Scalar-On-Functions Regression with Interaction Term. *Computational Statistics & Data Analysis*, **81**, 38-51. <https://doi.org/10.1016/j.csda.2014.07.001>
- [9] 王艺璇. 广义部分函数型数据分析及其应用[D]: [硕士学位论文]. 北京: 北方工业大学, 2022.
- [10] 毛可敬. 带有交互项的广义部分函数型线性模型研究及应用[D]: [硕士学位论文]. 北京: 北方工业大学, 2024.