Journal of Statistical Modelling: Theory and Applications
Vol. 3, No. 1, 2022, pp. 73-81
Yazd University Press 2022

JSMTA

Research Paper

Some asymptotic properties of functional linear regression model with points of impact

Alireza Shirvani [1], Omid Khademnoe[2], S. Mohammad Ebrahim Hosseini-Nasab*[3]
[1,3]Department of Statistics, Faculty of Mathematical Sciences, Shahid Beheshti University, Tehran, Iran
[2]Department of Statistics, Faculty of Sciences, University of Zanjan, Zanjan, Iran

Received: November 15, 2022/ Revised: February 20, 2023/ Accepted: February 25, 2023

Abstract: The functional linear regression model with points of impact is a recent augmentation of the classical functional linear model with many practically important applications. It is assumed that there exists an unknown number of impact points, that is discrete observation times where the corresponding functional values possess significant influences on the response variable. In this paper, we obtain some asymptotic properties of the model that can be used for further statistical inferences about the response variable. Specifically, rates of convergence for eigenfunctions estimates of the predictor covariance operator evaluated at the impact points estimates are derived. These are important results, because we do not have true eigenfunctions and impact points in applications and we have to use their estimates instead.

Keywords: Functional Linear regression models; Point of impact; Principal component analysis.
Mathematics Subject Classification (2010): 46S50.

1 Introduction

Functional data analysis (FDA) is a branch of statistics that analyzes data providing information about curves, surfaces or anything else varying over a continuum. In its most general form, under an FDA framework, each sample element of functional data is considered to be a random function. The physical continuum over which these

*Corresponding author: m_hosseininasab@sbu.ac.ir

functions are defined is often time, but may also be spatial location, wavelength, probability, etc. Intrinsically, functional data are infinite dimensional. The high intrinsic dimensionality of these data brings challenges for theory as well as computation, where these challenges vary with how the functional data were sampled. However, the high or infinite dimensional structure of the data is a rich source of information and there are many interesting challenges for research and data analysis. Advances in technology facilitate collecting and storing the data that are essentially in the form of curves. Because in practice, one may collect the values of the curves at a finite set of points, the multivariate methods can be applied to these kinds of data. However, the number of points observed per function may be much larger than the total number of functions. On the other hand, the existence of strong correlations between the values of a function at consecutive points can make the multivariate methods inefficient. Therefore, some theoretical justification is needed to provide the required definitions and concepts regarding the essential nature of the data. Ramsay and Silverman (2005) gave the theoretical and methodological development of FDA.

An important part of FDA is functional linear regression. This model has been studied in depth in theoretical and applied statistical literature. The most frequently used approach for estimating the slope function then is based on functional principal components analysis (FPCA). See, e.g., Frank and Friedman (1993); Bosq (2000); Muller and Stadtmuller (2005) in the context of generalized linear models. Alternative approaches and further theoretical results can, for example, be found in Crambes et al. (2009), Cardot and Johannes (2010), Comte and Johannes (2012), Delaigle and Hall (2012), Khademnoe and Hosseini-Nasab (2016), and Shi and Cao (2022). Hall and Horowitz (2007) showed that the approach based on FPCA to estimating the slope function, achieved optimal convergence rates. James et al. (2009) proposed an approach to obtain interpretable, flexible and accurate estimate of the slope function. Li and Hsing (2010) considered a functional linear model, where only a finite number of the predictor projections affects the response variable. They focused on the FPCA to determine the number of effective projections.

To our knowledge, Mckeague and Sen (2010) are the first to explicitly study identifiability and estimation of a “point of impact”, that is, discrete observation time where the corresponding functional value possess significant influence on the response variable, in a functional regression model. They show that consistent estimators are obtained by least squares, and calculated the convergence rates of the impact point and the slope function estimators. Kneip et al. (2016) considered functional linear regression, where scalar responses are modeled in dependence of i.i.d. predictor random functions. They studied a generalization of the classical functional linear regression model and assumed that there exists an unknown number of points of impact. In addition to estimating a functional slope parameter, they determined the number and locations of points of impact as well as corresponding regression coefficients. They showed that this number can be estimated consistently. Furthermore, rates of convergence for location of impact points estimates, regression coefficients and the slope parameter were derived. Based on the work of Kneip et al. (2016), Poss et al. (2020) investigated a nonparametric functional regression with points of impact when the response variable is scalar. Liebl et al. (2020) used the spline approach introduced by Crambes et al. (2009) to estimate the slope function and coefficients in the functional linear regression model with points of impact, and showed the efficiency of this approach by a simulation study.

Functional linear regression models with points of impact have many applications in some fields including engineering, psychology, medical sciences and etc. For example, in genome-wide expression studies, we may identify the location of genes that their activity is associated with clinical outcomes, or psychologists may be interested in understanding the moments of an affective documentary video that can recall a particular emotional state for the video viewers.

In this paper, we investigate some asymptotic properties of the estimators provided by Kneip et al. (2016), which are required in further statistical inferences. The paper is organized as follows. In Section 2, we introduce functional linear regression model with points of impact. Section 3 contains main results of our work, where we derive some asymptotic properties of the estimators in the functional linear regression model with points of impact. The results of analyzing Google AdWords data are found in Section 4. Finally Section 5 contains our conclusions.

2 Model and notations

2.1 Functional linear regression model with points of impact

We consider linear regression involving a scalar response variable Y and a functional predictor variable $X \in L^2(\mathcal{I})$, $\mathcal{I} = [a, b]$, where $[a, b]$ is a bounded interval of $\mathbb{R}$. It is assumed that data consist of an i.i.d. sample (X_i, Y_i), $i = 1, \ldots, n$, from (X, Y). The functional variable X is such that $E[\int_{\mathcal{I}} X^2(t)dt] < \infty$ and for simplicity the variables are supposed to be centered in the following: $E(Y) = 0$ and $E(X(t)) = 0$ for $t \in \mathcal{I}$, a.e. In this paper, we study the following functional linear regression model with points of impact

$$Y_i = \int_{\mathcal{I}} \beta\left(t\right) X_i\left(t\right) dt + \sum_{r=1}^{S} \beta_r X_i\left(\tau_r\right) + \varepsilon_i, \quad i = 1, \ldots, n, \tag{1}$$

where ε_i, $i = 1, \ldots, n$ are i.i.d. centered real random variables with $E(\varepsilon_i^2) = \sigma^2 < \infty$, which are independent of $X_i(t)$ for all t, $\beta \in L^2(\mathcal{I})$ is an unknown, bounded slope function and $\int_{\mathcal{I}} \beta\left(t\right) X_i\left(t\right) dt$ describes a common effect of the whole trajectory $X_i(\cdot)$ on Y_i. In addition, the model incorporates an unknown number $S \in \mathbb{N}$ of "points of impact", that is, specific time points $\tau_1, \ldots, \tau_S$ with the property that the corresponding functional values $X_i(\tau_1), \ldots, X_i(\tau_S)$ possess some significant influence on the response variable Y_i. Throughout the paper, we will assume that all points of impact are in the interior of the interval, $\tau_r \in (a, b)$, $r = 1, \ldots, S$.

2.2 Parameter estimates

In model (1), the function $\beta(t)$, the number $S \geq 0$, as well as τ_r and β_r, $r = 1$, ..., S, are unknown and have to be estimated from the data. Let $\lambda_1 \geq \lambda_2 \geq \cdots > 0$ denote the nonzero eigenvalues of the covariance operator Γ of X, while $\psi_1, \psi_2, \ldots$ denote the corresponding system of orthonormal eigenfunctions. We use the Kneip et al. (2016) method to estimate the number and locations of impact points. In situations where it can be assumed that $\int_{\mathcal{I}} \beta(t) X_i(t)\, dt = 0$ a.e., we have $Y_i = \sum\limits_{r=1}^{S} \beta_r X_i(\tau_r) + \varepsilon_i,\ i = 1, \ldots, n$ and the regression coefficients may be obtained by least squares method when replacing S by $\hat{S}$ and the unknown points of impact τ_r by their estimates $\hat{\tau_r}$. More precisely, in this case an estimator $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \ldots, \hat{\beta}_{\hat{S}})^T$ of $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_S)^T$ is determined by minimizing $\sum_{i=1}^{n}(Y_i - \sum_{s=1}^{\hat{S}} b_s X_i(\hat{\tau}_s))^2$ over all possible values $b_1, \ldots, b_{\hat{S}}$. In the general case with $\beta(t) \neq 0$ for some t, let $\hat{\lambda}_1, \hat{\lambda}_2, \ldots$ and $\hat{\psi}_1, \hat{\psi}_2, \ldots$ denote eigenvalues and eigenfunctions of the empirical covariance operator of $X_1, \ldots, X_n$. This is easily verified that $\int_a^b \beta(t) X_i(t) dt = \sum_{r=1}^{\infty} \alpha_r \langle X_i, \psi_r \rangle$, where $\alpha_r = \langle \beta, \psi_r \rangle$. Given estimates $\hat{\tau}_1, \ldots, \hat{\tau}_{\hat{S}}$ and a suitable cut-off parameter k, estimates $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \ldots, \hat{\beta}_{\hat{S}})^T$ of $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_S)^T$ and $\hat{\alpha}_1, \ldots, \hat{\alpha}_k$ of α_1, ..., α_k are determined by minimizing $\sum_{i=1}^{n}(Y_i - \sum_{r=1}^{k} a_r \langle X_i, \hat{\psi}_r \rangle - \sum_{s=1}^{\hat{S}} b_s X_i(\hat{\tau}_s))^2$ over all $a_r,\ b_s,\ r = 1, \ldots, k,\ s = 1, \ldots, \hat{S}$. Based on the estimated coefficients $\hat{\alpha}_1, \ldots, \hat{\alpha}_k$, an estimator of the slope function β is then given by $\hat{\beta}(t) = \sum_{r=1}^{k} \hat{\alpha}_r \hat{\psi}_r(t)$.

3 Main results

The theoretical properties related to this work are presented in Theorems 3.1-3.4, below.

Theorem 3.1. *If the conditions of Theorem 5 of Kneip et al. (2016) hold, then for large enough positive integer j, we have*

$$\sum_{l \neq j}^{\infty} \frac{\lambda_l^{\frac{1}{2}} \lambda_j^{\frac{1}{2}}}{|\lambda_l - \lambda_j|} \leq const. j \log j.$$

Proof. We are first going to decompose the sum into four terms as follows

$$\sum_{l \neq j}^{\infty} \frac{\lambda_l^{\frac{1}{2}} \lambda_j^{\frac{1}{2}}}{|\lambda_l - \lambda_j|} = \mathcal{T}_1 + \mathcal{T}_2 + \mathcal{T}_3 + \mathcal{T}_4,$$

where

$$\mathcal{T}_1 = \sum_{l=1}^{[\frac{j}{2}]} \frac{\lambda_l^{\frac{1}{2}} \lambda_j^{\frac{1}{2}}}{|\lambda_l - \lambda_j|}, \mathcal{T}_2 = \sum_{l=[\frac{j}{2}]+1}^{j-1} \frac{\lambda_l^{\frac{1}{2}} \lambda_j^{\frac{1}{2}}}{|\lambda_l - \lambda_j|}, \mathcal{T}_3 = \sum_{l=j+1}^{2j} \frac{\lambda_l^{\frac{1}{2}} \lambda_j^{\frac{1}{2}}}{|\lambda_l - \lambda_j|}, \mathcal{T}_4 = \sum_{l=2j+1}^{\infty} \frac{\lambda_l^{\frac{1}{2}} \lambda_j^{\frac{1}{2}}}{|\lambda_l - \lambda_j|}.$$

Now, look at the terms of the above relation. Hall and Hosseini-Nasab (2006) showed that if the Assumptions 3 and 4 of Kneip et al. (2016) hold, then

$$|\lambda_j - \lambda_l|^{-1} \leq const. \frac{\max\{l, j\}}{|l - j| \max\{\lambda_l, \lambda_j\}}.$$

Thus, based on conditions of Theorem 5 of Kneip et al. (2016) it is concluded that

$$\mathcal{T}_1 = \sum_{l=1}^{[\frac{j}{2}]} \frac{\lambda_l^{\frac{1}{2}} \lambda_j^{\frac{1}{2}}}{|\lambda_l - \lambda_j|} \quad \leq \quad const. \sum_{l=1}^{[\frac{j}{2}]} \frac{\lambda_l^{\frac{1}{2}} \lambda_j^{\frac{1}{2}} j}{(j - l)\lambda_l}$$

$$
\begin{aligned}
&\leq const. \sum_{l=1}^{[\frac{j}{2}]} \frac{\lambda_l^{\frac{1}{2}} \lambda_l^{\frac{1}{2}} j}{(j-l)\lambda_l} \\
&= const. \sum_{l=1}^{[\frac{j}{2}]} \frac{j}{j-l} \\
&\leq const.j \sum_{l=1}^{[\frac{j}{2}]} \frac{j}{j-[\frac{j}{2}]} \\
&\leq const.j,
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{T}_2 = \sum_{l=[\frac{j}{2}]+1}^{j-1} \frac{\lambda_l^{\frac{1}{2}} \lambda_j^{\frac{1}{2}}}{|\lambda_l - \lambda_j|} &\leq const. \sum_{l=[\frac{j}{2}]+1}^{j-1} \frac{\lambda_l^{\frac{1}{2}} \lambda_j^{\frac{1}{2}} j}{(j-l)\lambda_l} \\
&\leq const. \sum_{l=[\frac{j}{2}]+1}^{j-1} \frac{\lambda_l^{\frac{1}{2}} \lambda_l^{\frac{1}{2}} j}{(j-l)\lambda_l} \\
&= const.j \sum_{l=[\frac{j}{2}]+1}^{j-1} \frac{1}{j-l} \\
&\leq const.j(1 + \int_{[\frac{j}{2}]+1}^{j-1} \frac{1}{j-x} dx) \\
&\leq const.j \log j,
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{T}_3 = \sum_{l=j+1}^{2j} \frac{\lambda_l^{\frac{1}{2}} \lambda_j^{\frac{1}{2}}}{|\lambda_l - \lambda_j|} &\leq const. \sum_{l=j+1}^{2j} \frac{\lambda_l^{\frac{1}{2}} \lambda_j^{\frac{1}{2}} l}{(l-j)\lambda_j} \\
&\leq const. \sum_{l=j+1}^{2j} \frac{l}{l-j} \\
&\leq const.j \sum_{l=j+1}^{2j} \frac{1}{l-j} \\
&\leq const.j \log j,
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{T}_4 = \sum_{l=2j+1}^{\infty} \frac{\lambda_l^{\frac{1}{2}} \lambda_j^{\frac{1}{2}}}{|\lambda_l - \lambda_j|} &= const. \sum_{l=2j+1}^{\infty} \frac{\lambda_l^{\frac{1}{2}} \lambda_j^{\frac{1}{2}}}{\lambda_j - \lambda_l} \\
&\leq const. \sum_{l=2j+1}^{\infty} \frac{\lambda_j}{\lambda_j - \lambda_l} \\
&\leq cons.j \log j.
\end{aligned}
$$

See Khademnoe and Hosseini-Nasab (2016) to prove the last inequality. Based on the above results, it is concluded that

$$\sum_{l\neq j}^{\infty} \frac{\lambda_l^{\frac{1}{2}}\lambda_j^{\frac{1}{2}}}{|\lambda_l - \lambda_j|} \leq const.j \log j.$$

Hence the result follows and the Theorem 3.1 is proved. □

Note that the result given in Theorem 3.1 is a general result in functional linear regression. Theorem 3.1 is important for obtaining the asymptotic results related to the estimators in functional linear regression, and then for finding the asymptotic distribution of prediction in the functional linear regression model with points of impact and will be widely used in our next works.

Definition 3.2. *$Y_n = O_p(n^{\alpha})$ if and only if for each $\epsilon > 0$, there exist $M > 0$ such that $P(|n^{-\alpha}Y_n| > M) < \epsilon$, for all n.*

Theorem 3.3. *Suppose that there exist some $\mu > 1$ and some $\sigma^2 < C < \infty$ such that $C^{-1}j^{-\mu} \leq \lambda_j \leq Cj^{-\mu}$ and $\lambda_j - \lambda_{j+1} \geq C^{-1}j^{-\mu-1}$ for all $j \geq 1$. Under the conditions of Theorem 5 of Kneip et al. (2016), for $r = 1, \ldots, S$ and $j \geq 1$ we have*

$$\psi_j(\hat{\tau}_r) = \psi_j(\tau_r) + O_p(j^{\frac{\mu}{2}}.n^{\frac{-1}{2}}).$$

Proof. We have

$$\begin{aligned} \psi_j(t) &= \lambda_j^{-1}\Gamma\big(\psi_j(t)\big), \\ \Gamma\big(\psi_j(t)\big) &= E\big[\langle X, \psi_j\rangle X(t)\big], \end{aligned}$$

for $j = 1, 2, \ldots$. The Cauchy-Schwarz inequality, Assumption 3 and Equation (4.6) of Kneip et al. (2016), together imply that

$$\begin{aligned} |\psi_j(\hat{\tau}_r) - \psi_j(\tau_r)| &= \lambda_j^{-1}\Big|\Gamma\big(\psi_j(\hat{\tau}_r)\big) - \Gamma\big(\psi_j(\tau_r)\big)\Big| \\ &= \lambda_j^{-1}\Big|E\Big[\langle X, \psi_j\rangle\big(X(\hat{\tau}_r) - X(\tau_r)\big)\Big]\Big| \\ &\leq \lambda_j^{-1}\sqrt{E[\langle X, \psi_j\rangle^2]E[(X(\hat{\tau}_r) - X(\tau_r))^2]} \\ &= \lambda_j^{-\frac{1}{2}}\sqrt{E[(X(\hat{\tau}_r) - X(\tau_r))^2]} \\ &= \lambda_j^{-\frac{1}{2}}O_p(n^{\frac{-1}{2}}) = O_p(j^{\frac{\mu}{2}}.n^{\frac{-1}{2}}). \end{aligned}$$

□

Theorem 3.4. *Under the conditions of Theorem 3.3, for $r = 1, ..., S$ and $j \geq 1$ we have*

$$\hat{\psi}_j(\hat{\tau}_r) = \hat{\psi}_j(\tau_r) + O_p(j^{\frac{\mu}{2}}.n^{\frac{-1}{2}}).$$

Proof. Equation (A.11) of Kneip et al. (2016) implies that

$$\hat{\lambda}_j = \lambda_j + O_p(n^{\frac{-1}{2}}),$$

while (A.19) from Kneip et al. (2016) yields

$$(\hat{\psi}_j(\hat{\tau}_r) - \hat{\psi}_j(\tau_r))^2 \leq \hat{\lambda}_j^{-1} \frac{1}{n} \sum_{i=1}^{n} (X_i(\hat{\tau}_r) - X_i(\tau_r))^2,$$

thus

$$(\hat{\psi}_j(\hat{\tau}_r) - \hat{\psi}_j(\tau_r))^2 \leq (\lambda_j + O_p(n^{\frac{-1}{2}}))^{-1} \frac{1}{n} \sum_{i=1}^{n} (X_i(\hat{\tau}_r) - X_i(\tau_r))^2.$$

By Taylor expansions we have

$$(\lambda_j + O_p(n^{\frac{-1}{2}}))^{-1} = \lambda_j^{-1} + O_p(n^{\frac{-1}{2}}),$$

and from Assumption 3 of Kneip et al. (2016) we obtain

$$\lambda_j^{-1} \leq const.j^{\mu}.$$

These results and (4.6) of Kneip et al. (2016) together imply that

$$(\hat{\psi}_j(\hat{\tau}_r) - \hat{\psi}_j(\tau_r))^2 \leq \big(O(j^{\mu}) + O_p(n^{\frac{-1}{2}})\big).O_p(n^{-1}) = O_p(j^{\mu}.n^{-1}),$$

and the assertion of the theorem is an immediate consequence. □

4 Application to real data

The methodology is applied to a real data set called Google AdWords. The data set is important and one of the most popular online advertising platform. The main pricing mechanism at Google AdWords is the so-called Pay-Per-Click (PPC) mechanism. Here, advertisers (e.g., an online outdoor shop in our application) can bid for a sponsored "impression" to be presented along with Google's search results when a user search a query related to a specific keyword. The basic building block of an online ad campaign is a text collection of keywords related to the promulgated products. The limited number of sponsored impressions is allocated by an auction. When impression comes out on the display, their advertisers are selected according to their ad-rank, which is basically their original bid, that is, the maximum "costs-per-click" an advertiser is willing to pay times the quality score. The relevance of an advertiser's impression is measured at a discrete metric from 1, the lowest, to 10, the best. Google AdWords auctions are time continuous and an advertiser only pays if a user clicks on the displayed impression.

To model the relationship between the yearly clicks and the yearly trajectories of daily impressions from April 1st, 2012 to March 31th, 2013, Liebl et al. (2020) used a functional linear regression model with points of impact. More specifically, they used the logarithm of the yearly sum of clicks as the response variable, that is $Y_i = log(\sum_{t=1}^{365} clicks_{it})$, where the index i denotes the ith keyword of the given ad campaign. Moreover, $X_i(t) = \log(impressions_{it})$, $i = 1, \ldots, 903$, and $t = 1, \ldots, 365$.

Using Kneip et al. (2016) estimation procedure, we have recognized five points of impact. Locations of these impact points in [0,1] are as follows: $\hat{\tau}_1 = 0.2$ (The 75th day), $\hat{\tau}_2 = 0.7$ (The 257th day), $\hat{\tau}_3 = 0.04$ (The 14th day), $\hat{\tau}_4 = 0.87$ (The 316th day),

Table 1: Some of the estimated eigenfunctions of the covariance operator X evaluated at the impact points estimates $\hat{\tau}_r;\ r = 1, \ldots, 5$. For each j, the estimated eigenfunction $\hat{\psi}_j(\cdot)$ is evaluated at the impact points estimates $\hat{\tau}_r;\ r = 1, \ldots, 5$.

j	r				
	1	2	3	4	5
1	0.878	1.135	0.965	1.038	1.049
2	1.247	-1.328	0.797	-0.978	1.153
3	0.460	0.782	0.581	0.575	0.158
4	-0.340	-1.418	-0.950	1.632	0.569
5	-0.560	1.014	-1.389	-0.024	0.328
10	0.717	0.891	-0.100	2.366	-0.398
15	-0.928	-1.180	-0.691	0.894	1.039
20	0.383	-0.723	0.288	2.422	1.415
30	-1.308	0.708	0.634	1.244	1.358
40	-1.522	-1.104	-0.181	-0.408	-0.389
50	-1.173	-0.216	-0.551	0.168	-1.215

and $\hat{\tau}_5 = 0.31$ (The 113th day). Table 1 provides some of the estimated eigenfunctions of the covariance operator X evaluated at the impact points estimates $\hat{\tau}_r;\ r = 1, \ldots, 5$. Based on Theorems 3.3 and 3.4, these estimates can be used asymptotically for making any inference about the true values of the eigenfunctions of the covariance operator X evaluated at the impact points $\tau_r;\ r = 1, \ldots, 5$.

5 Conclusions

Based on Theorems 3.3 and 3.4, the estimators $\psi_j(\hat{\tau}_r)$ and $\hat{\psi}_j(\hat{\tau}_r)$ have the same rates of convergence for estimating $\psi_j(\tau_r)$ and $\psi_j(\hat{\tau}_r)$, respectively. These are important results, because we do not have τ_r and ψ_j in practice and we have to use their estimates instead. The obtained asymptotic properties will be used in many statistical inferences about the functional linear regression model with points of impact, for example for finding the asymmetric distribution of prediction and confidence intervals, and performing hypothesis tests.

Acknowledgments

The authors are grateful to the two referees and Editor-in-chief for their valuable comments and suggestions which led to improvement of this manuscript.

References

Bosq, D. (2000). *Linear Processes in Function Spaces: Theory and Applications. Lecture Notes in Statistics*. Vol. 149, New York: Springer.

Cardot, H. and Johannes, J. (2010). Thresholding projection estimators in functional linear models. *Journal of Multivariate Analysis*, **101**(2):395–408.

Comte, F. and Johannes, J. (2012). Adaptive functional linear regression. *The Annals of Statistics*, **40**(6):2765–2797.

Crambes, C., Kneip, A. and Sarda, P. (2009). Smoothing splines estimators for functional linear regression. *The Annals of Statistics*, **37**(1):35–72.

Delaigle, A. and Hall, P. (2012). Methodology and theory for partial least squares applied to functional data. *The Annals of Statistics*, **40**(1):322–352.

Frank, I.E. and Friedman, J.H. (1993). A statistical view of some chemometrics regression tools. *Technometrics*, **35**(2):109–135.

Hall, P. and Horowitz, J.L. (2007). Methodology and convergence rates for functional linear regression. *The Annals of Statistics*, **35**(1):70–91.

Hall, P. and Hosseini-Nasab, M. (2006). On properties of functional principal components analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **68**(1):109–126.

James, G.M., Wang, J. and Zhu, J. (2009). Functional linear regression that's interpretable. *The Annals of Statistics*, **37**(5A):2083–2108.

Khademnoe, O. and Hosseini-Nasab, S.M.E. (2016). On asymptotic distribution of prediction in functional linear regression. *Statistics*, **50**(5):974–990.

Kneip, A., Poss, D., Sarda, P. (2016). Functional linear regression with points of impact. *The Annals of Statistics*, **44**(1):1–30.

Li, Y. and Hsing, T. (2010). Deciding the dimension of effective dimension reduction space for functional and high-dimensional data. *The Annals of Statistics*, **38**(5):3028–3062.

Liebl, D., Rameseder, S. and Rust, C. (2020). Improving estimation in functional linear regression with points of impact: insights into Google AdWords. *Journal of Computational and Graphical Statistics*, **29**:814-826.

Mckeague, I.W. and Sen, B. (2010). Fractals with point impact in functional linear regression. *The Annals of Statistics*, **38**(4):2559–2586.

Muller, H.-G. and Stadtmuller, U. (2005). Generalized functional linear models. *The Annals of Statistics*, **33**(2):774–805.

Poß, D., Liebl, D., Kneip, A., Eisenbarth, H., Wager, T.D. and Barrett, L.F. (2020). Superconsistent estimation of points of impact in non-parametric regression with functional predictors. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **82**(4):1115–1140.

Ramsay, J.O. and Silverman, B.W. (2005). *Functional Data Analysis*. New York: Springer.

Shi, H. and Cao, J. (2022). Robust functional principal component analysis based on a new regression framework. *Journal of Agricultural, Biological and Environmental Statistics*, **27**(3):523–543.