Statistica Sinica Preprint No: SS-2018-0155	
Title	Inference for Generalized partial functional linear regression
Manuscript ID	SS-2018-0155
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202018.0155
Complete List of Authors	Ting Li and Zhongyi Zhu
Corresponding Author	Zhongyi Zhu
E-mail	zhuzy@fudan.edu.cn
Notice: Accepted version subject to English editing.	

Statistica Sinica

Inference for generalized partial functional linear regression

Ting Li and Zhongyi Zhu

Department of Statistics, Fudan University

Abstract: We study the generalized partial functional linear models and aim to conduct inference, especially hypothesis testings, for these models. A Bahadur representation for both functional and scalar estimators is developed based on the reproducing kernel Hilbert space. We establish an asymptotic independence between the scalar estimators and the estimator of the functional part. A penalized likelihood ratio test is proposed to detect the significant effects of the functional and the scalar covariates on the scalar outcome either simultaneously or separately. The asymptotic normality of the test statistic is established under the null hypothesis. Simulation studies provide a numerical support for the asymptotic properties. Data of air pollution are used to demonstrate our method.

Keywords and phrases: Bahadur Representation; Hypothesis testing; Penalized likelihood ratio test; Reproducing kernel Hilbert space.

1. Introduction

Driven by numerous applications, functional data analysis has gained huge attention in recent years. The functional linear model and its extensions have been massive studied in the literature (Ramsay and Silverman, 2002, 2005). Frequently we have both one covariate vector and one functional variable on each individual subject, and people are interested in assessing the effects of the functional variable and the scalar covariates on a scalar response. Hypothesis testings relating to the functional and scalar parameters simultaneously or separately are of great importance, since they not only provide an overall assessment of the model, but also evaluate the effects of the covariates on the outcome.

Most of the existing testing methods for the partial functional linear models are based on the the functional principal component analysis (FPCA) method. Kong, Staicu, and Maity (2016) and Su, Di, and Hsu (2017) proposed to assess the association between the functional predictor and the response. Yu, Zhang, and Du (2016) turned to test the effect of the parametric component on the response. These testing procedures rely heavily on the success of the FPCA approach, and may be inappropriate if the functional parameter can not be effectively represented by the leading principals of the functional covariates. Also they are not capable of handling binary response variables. Moreover, the above methods mainly deal with testing the linear effect of either the functional predictor or the scalar covariates, while simultaneous testings of functional and scalar parameters receive little attention in the functional data analysis.

To the best of our knowledge, it is the first work to simultaneously test global behaviors of functional and scalar parameters. The aim of this paper is to develop a new method to conduct inference, especially hypothesis testings, on both functional and scalar parameters simultaneously or separately for generalized partial functional linear models under the RKHS framework.

Following Yuan and Cai (2010) and Cai and Yuan (2012), we employ the roughness regularization method to avoid the drawbacks of the FPCA method. Motivated by the seminal work of Shang and Cheng (2015), we establish a Bahadur representation for both scalar and functional estimators using our predefined inner product. Despite conceptual similarity, the model considered in this paper is more comprehensive and informative by additionally incorporating the scalar variables. Moreover, we define a new type of inner product, leading to a different eigensystem. The potential challenge of allowing for the scalar variables arises from dealing with the interactions between the functional covariate and the scalar covariates. To overcome this difficulty, we restrict that the scalar covariates can only linearly associate with the functional process and figure out the decay rates of the corresponding coefficients.

We discover that the scalar estimators and the estimator of the functional part are asymptotically independent under some mild conditions. Cheng and Shang (2015) demonstrated the asymptotic independence between the estimator of a general nonlinear function and the parametric estimators. However, all the covariates are scalar and the results can not apply to functional data.

A penalized likelihood ratio test is also developed to detect significant effects of the functional and scalar covariates on the outcome either simultaneously or separately. Compared with the test in Shang and Cheng (2015) which can only investigate the association between the functional predictor and the response, the proposed test offers more choices. The null limit distribution of the proposed test statistic is shown to be a normal distribution and approximately a chi-square distribution, which enables the easy implementation of the proposed testing procedure. Simulation studies demonstrate that the proposed test has good size and power performance, and show its superiority over other competing methods.

The rest of this paper is organized as follows. Section 2 introduces the model and defines inner products for the parameter spaces. Section 3 shows the Bahadur representation and the asymptotic independence results. The test statistic and its null limit distribution are presented in Section 4.

Simulation studies and real data analysis are conducted in Sections 5 and 6. Section 7 concludes the paper with some further discussion. Additional simulations and all the proofs are provided in the Supplementary Material.

2. Model and Inner Products

The generalized partial functional linear regression model has the form

$$\mu_0(X, Z) = E(Y|X, Z) = F\left(Z^\top\gamma_0 + \int_0^1 X(t)\beta_0(t)dt\right), \tag{2.1}$$

where $Y \in \mathcal{Y} \subseteq \mathbb{R}$ is the response, $X(t)$ is a random function recorded on the interval $\mathbb{I} = [0, 1]$, and $Z \in \mathbb{R}^p$ is a vector of covariates with the dimension p being fixed. The functional coefficient $\beta_0(\cdot)$ is defined on $\mathbb{I} = [0, 1]$ and F is a known link function. We consider $\beta \in H^m(\mathbb{I})$ the m-order Sobolev space defined as

$$\begin{aligned} H^m(\mathbb{I}) = \{\beta : \mathbb{I} \mapsto \mathbb{R}|\ & \beta^{(j)} \text{ is absolutely continuous} \\ & \text{for } j = 0, \ldots, m-1 \text{ and } \beta^{(m)} \in L^2(\mathbb{I})\}. \end{aligned} \tag{2.2}$$

Following Yuan and Cai (2010) and Cai and Yuan (2012), we assume that $m > 1/2$ such that $H^m(\mathbb{I})$ is a RKHS. The full parameter space for $\theta = (\gamma, \beta)$ is $\mathcal{H} = \mathbb{R}^p \times H^m(\mathbb{I})$.

Model (2.1) is more comprehensive and flexible than the standard generalized functional linear model by allowing for the scalar covariates. Different from the general class of semi-nonparametric regression models, the whole curves rather than single points are included in the model. Obviously, the observed curves contain more information than points.

To estimate γ_0 and $\beta_0(t)$, we adopt a general loss function $\ell(y; a)$: $\mathcal{Y} \times \mathbb{R} \mapsto \mathbb{R}$, which can be either a likelihood or a quasi-likelihood function. Dimension-reduction or additional constraints are mandatory due to the infinite dimensionality of $\beta_0(t)$. One popular way is to represent $\beta_0(t)$ as a truncated expansion of certain basis functions, such as those derived from FPCA, B-splines or Fourier basis functions. As pointed out in Ramsay and Silverman (2005), the truncation parameter changes in a discrete manner, which may yield an imprecise control on the model complexity, and can often result in inaccurate functional estimates with hard-to-interpret "artificial" bumps. We choose the roughness penalty approach to avoid these problems. The penalized estimators are obtained by $(\hat{\gamma}_{n,\lambda}, \hat{\beta}_{n,\lambda}) = \arg\sup_{(\gamma,\beta)\in\mathcal{H}} \ell_{n,\lambda}(\theta)$, where

$$\ell_{n,\lambda}(\theta) = \left\{ \frac{1}{n} \sum_{i=1}^{n} \ell(Y_i; Z_i^\top \gamma + \int_0^1 X_i(t)\beta(t)dt) - (\lambda/2)J(\beta, \beta) \right\}, \quad (2.3)$$

and $J(\beta_1, \beta_2) = \int_0^1 \beta_1^{(m)}(t)\beta_2^{(m)}(t)dt$ is a roughness penalty. The roughness penalty is used to control the model complexity via the smoothing parameter λ in a relatively continuous way. This type of penalized estimators is quite common in the literature (Yuan and Cai, 2010; Cai and Yuan, 2012; Du and Wang, 2014; Shang and Cheng, 2015).

We now introduce the inner products and norms of $H^m(\mathbb{I})$ and $\mathcal{H}$. The Bahadur representation in Section 3 is established in terms of the norms we defined. Meanwhile, deriving the null limit distribution of the proposed test statistic in Section 4 relies heavily on these inner products. Let $U = (X, Z) \in \mathcal{U}$. Denote $\dot{\ell}_a(y;a)$, $\ddot{\ell}_a(y;a)$ and $\ell'''_a(y;a)$ as the first-, second- and third-order derivatives of $\ell(y;a)$ with respect to a. Also define $\epsilon = \dot{\ell}_a(Y; Z^\top\gamma_0 + \int_0^1 X(t)\beta_0(t)dt)$ and

$$I(U) = -E(\ddot{\ell}_a(Y; Z^\top\gamma_0 + \int_0^1 X(t)\beta_0(t)dt)|U). \tag{2.4}$$

The inner product for any $\beta_1, \beta_2 \in H^m(\mathbb{I})$ is defined by

$$\langle \beta_1, \beta_2 \rangle_1 = V(\beta_1, \beta_2) + \lambda J(\beta_1, \beta_2), \tag{2.5}$$

where $V(\beta_1, \beta_2) = \int_0^1 \int_0^1 C(s,t)\beta_1(s)\beta_2(t)dsdt$, $C(s,t) = E_X\{B(X)X(t)X(s)\}$, and $B(X) = E\{I(U)|X\}$ acts as a weighting function so that $C(s,t)$ can be viewed as a weighted covariance operator of X. Denote the induced norm by $\|\cdot\|_1$. Yuan and Cai (2010) adopted an inner product without the weighting function $B(X)$, and Shang and Cheng (2015) introduced the fisher information $I(U) = I(X)$ as the weighting function. We modify the weighting function to be the conditional expectation $B(X) = E\{I(U)|X\}$, so that $C(s,t)$ is only a function of X.

For any $\theta_1 = (\gamma_1, \beta_1), \theta_2 = (\gamma_2, \beta_2) \in \mathcal{H}$, the inner product of the full parameter space $\mathcal{H}$ is defined by

$$\begin{aligned}
& \langle(\gamma_1,\beta_1),(\gamma_2,\beta_2)\rangle \\
= \; & E_U\left\{I(U)\left(Z^\top\gamma_1+\int_0^1 X(t)\beta_1(t)dt\right)\left(Z^\top\gamma_2+\int_0^1 X(t)\beta_2(t)dt\right)\right\} \\
& +\lambda J(\beta_1,\beta_2). \qquad (2.6)
\end{aligned}$$

The corresponding norm $\|\theta\|^2=\langle\theta,\theta\rangle$ is well defined under some conditions. Specifically, the positive definiteness of the matrix Ω_1 defined in Assumption 4 ensures the validity of the norm. We can derive the Bahadur representation in terms of this norm, which is rarely studied under the L^2 norm. Cheng and Shang (2015) considered the general nonparametric function and employed a similar inner product. It is more challenging to obtain a well defined inner product allowing for functional data, since it requires more efforts to obtain the expression of Ω_1.

3. Theoretical Results

In this section, we focus on deriving the joint Bahadur representation for the penalized estimators. The joint Bahadur representation greatly facilitates the asymptotic analysis. The joint distribution for the scalar estimators and the estimator of the functional part can be obtained afterwards. We start with some assumptions. Let $a_v \asymp b_v$ if there exist constants $c_1, c_2>0$ such that $c_1\le a_v/b_v\le c_2$, and let $\|\cdot\|_{L^2}$ be the L^2 norm.

Assumption 1. *The weighted covariance operator $C(s,t)$ is continuous on $\mathbb{I} \times \mathbb{I}$. For any $\beta \neq 0 \in H^m(\mathbb{I})$, we have $C\beta \neq 0$ where $(C\beta)(t) = \int_0^1 C(s,t)\beta(t)dt$.*

Assumption 2. *The loss function $\ell(y;a)$ is three times continuously differentiable and strictly concave with respect to a. There exist positive constants C_0, C_1 and C_2 such that,*

$$\begin{aligned} E\left\{\exp\left(\sup_{a\in\mathbb{R}} |\ddot{\ell}_a(y;a)|/C_0\right)|U\right\} &\leq C_1 \quad \text{a.s.}, \\ E\left\{\exp\left(\sup_{a\in\mathbb{R}} |\ell'''_a(y;a)|/C_0\right)|U\right\} &\leq C_1 \quad \text{a.s.}, \end{aligned} \tag{3.1}$$

and $C_2^{-1} \leq I(U) \leq C_2$ a.s.. Also $E(\epsilon|U) = 0$ and $E(\epsilon^2|U) = I(U)$ a.s..

Assumption 1 is an analogy to the positive definiteness of the ordinary covariance operator in FPCA, and enables (2.5) to be a well defined inner product. Assumption 2 is commonly used in semiparametric qusi-likelihood models (Mammen and van de Geer, 1997). Since $\ell(y;a)$ is either a likelihood $\ell(y;a) = \log p(y;F(a))$ where the conditional distribution of y is $p(y;\mu_0(x,z))$, or a qusi-likelihood $\ell(y;a) = \int_y^{F(a)}(y-t)/V(t)dt$ with $V(\mu_0(X,Z)) = \mathrm{Var(Y|X,Z)}$, the link function F is determined by $\ell(y;a)$. For example, in the functional linear model where $\ell(y;a) = (y-a)^2$, F is an identity link function. While in a logistic regression model with $\ell(y;a) = ay - \log(1+\exp(a))$, F has the form $F(a) = \exp(a)/(1+\exp(a))$.

Therefore, the assumptions on the loss function imply the assumptions on the link function F.

Assumption 3. *The space $H^m(\mathbb{I})$ is a RKHS under the inner product (2.5). For any $v \geq 1$ and some constants $a \geq 0$ and $C_\varphi > 0$, there exist eigenfunctions $\{\varphi_v\}_{v\geq 1}$ in $H^m(\mathbb{I})$ such that $\|\varphi_v\|_{L^2} \leq C_\varphi v^a$ for each $v \geq 1$. The eigenfunctions also satisfy $V(\varphi_v, \varphi_u) = \delta_{vu}$ and $J(\varphi_v, \varphi_u) = \rho_v \delta_{vu}$ for any $v, u \geq 1$, where $\rho_v \asymp v^{2k}(k \geq 3/2)$ is a nondecreasing nonnegative sequence, and $\delta_{vu} = 1$ if and only if $v = u$.*

Assumption 3 indicates that all $\beta \in H^m(\mathbb{I})$ have the representation $\beta = \sum_{v=1}^{\infty} V(\beta, \varphi_v)\varphi_v$. The eigensystem construction above plays a role of controlling the local behaviors of the penalized estimates (Gu, 2013). In particular, the eigensystem can be obtained by the pseudo Sacks-Ylvisaker conditions in the Supplementary Material of Shang and Cheng (2015). It is worth mentioning that although the eigensystem has a similar form as that in Shang and Cheng (2015), the final eigenfunctions and eigenvalues are different due to the different inner products.

Assumption 4. *The $p \times p$ matrix $\Omega_1 = E\{I(U)(Z - \mathbf{G}(X))(Z - \mathbf{G}(X))^\top\}$ is positive definite, where $\mathbf{G}(X) = E\{I(U)Z|X\}/B(X)$ is a p-dimensional functional valued vector. Denote $\mathbf{G}(X) = (G_1(X), \ldots, G_p(X))^\top$. For each*

$j = 1, \ldots, p$, there exists $\tilde{\beta}_j$ such that $G_j(X)$ can be expressed as $G_j(X) = \int_0^1 X(t)\tilde{\beta}_j(t)dt$ with $V(\tilde{\beta}_j, \tilde{\beta}_j) < \infty$.

The p-dimensional functional valued vector $\mathbf{G}(X)$ is a projection of Z to X satisfying $E\{I(U)(Z - \mathbf{G}(X))X\} = 0$. Assumption 4 guarantees the positive definiteness of Ω_1 and admits that the projection of Z to X is linear in X. A similar condition is adopted in Shin and Lee (2012).

Assumption 5. *There exist constants $s_1^* \in (0, 1)$ and M_1 such that*

$$E\{\exp(s_1^* \|X\|_{L^2})\} \;\; < \;\; \infty, \tag{3.2}$$

$$E\left\{\left|\int_0^1 X(t)\beta(t)dt\right|^4\right\} \leq M_1 \left[E\left\{\left|\int_0^1 X(t)\beta(t)dt\right|^2\right\}\right]^2, \tag{3.3}$$

for any $\beta \in H^m(\mathbb{I})$.

Assumption 6. *There exists a constant $s_2^* \in (0, 1)$ such that*

$$E\{\exp(s_2^*(Z^\top Z)^{1/2})\} < \infty, \tag{3.4}$$

$$E\{\exp(s_2^*((Z - \langle \boldsymbol{A}, \tau(X)\rangle_1)^\top (Z - \langle \boldsymbol{A}, \tau(X)\rangle_1))^{1/2})\} < \infty, \tag{3.5}$$

where $\tau(X) = \sum_v \frac{X_v}{1+\lambda\rho_v}\varphi_v(t)$ with $X_v = \int_0^1 X(t)\varphi_v(t)dt$, and $\boldsymbol{A} = (A_1, \ldots, A_p)^\top$ with $A_j(t) = \sum_v \frac{V(\tilde{\beta}_j, \varphi_v)}{1+\lambda\rho_v}\varphi_v(t)$. Moreover, for any $\gamma \in \mathbb{R}^p$, there exists a constant M_2 satisfying

$$E(|Z^\top\gamma|^4) \leq M_0[E(|Z^\top\gamma|^2)]^2. \tag{3.6}$$

A detailed discussion of Assumption 5 can be found in Shang and Cheng (2015). The condition (3.3) is commonly seen in roughness penalty methods (Yuan and Cai, 2010; Du and Wang, 2014). Assumption 6 is an analogy to Assumption 5. The condition (3.5) is critical to derive the null limit distribution of the proposed test statistic in Section 4.

Recall that k is specified in Assumption 3, and let $h = \lambda^{1/(2k)}$. The following theorem gives the convergence rate of $\hat{\theta}_{n,\lambda}$.

Theorem 1. *Suppose that Assumptions 1–6 are satisfied. As $n \to \infty$, if the conditions $n^{-1/2}h^{-(a+1)-((2k-2a-1)/4m)}(\log n)^2(\log\log n)^{1/2} = o(1)$, $n^{-1/2}h^{-1} = o(1)$ and $h = o(1)$ hold, we have*

$$\|\hat{\theta}_{n,\lambda} - \theta_0\| = O_p((nh)^{-1/2} + h^k). \tag{3.7}$$

The convergence rate in (3.7) is a nonparametric rate, which is commonly seen in the smoothing spline literature (Gu, 2013). However, γ and β are supposed to be estimated at parametric and nonparametric rates, respectively. The parametric convergence rate of $\hat{\gamma}_{n,\lambda}$ is carried out under some additional mild conditions in Theorem 3.

We are now able to establish the joint Bahadur representation for both the functional and the scalar estimators. Let $S_{n,\lambda}(\theta)$ denote the first Fréchet derivative of $\ell_{n,\lambda}(\theta)$ with respect to θ.

Theorem 2. *Under the conditions in Theorem 1, if* $\log(h^{-1}) = O(\log n)$ *as* $n \to \infty$*, then we have*

$$\|\hat{\theta}_{n,\lambda} - \theta_0 - S_{n,\lambda}(\theta_0)\| = O_p(a_n), \tag{3.8}$$

where $a_n = n^{-1/2}h^{-(4ma+6m-1)/4m}r_n(\log n)^2(\log\log n)^{1/2} + C_l h^{-1/2}r_n^2$, $r_n = (nh)^{-1/2} + h^k$ *and* $C_l = \sup_{u\in\mathcal{U}} E\{\sup_{a\in\mathbb{R}} |\ell_a'''(Y;a)||U = u\}$.

We obtain the same rate a_n as that in Shang and Cheng (2015), but additionally allow for scalar estimators. The Bahadur representation for generalized functional linear models is hence extended to generalized partial functional linear models. Such an extension benefits from the additional Assumptions 4 and 6, and it requires more efforts to derive the theoretical properties. Specifically, Assumptions 4 and 6 contribute to representing $E\{I(U)Z\int_0^1 X(t)\beta(t)dt\}$ via the inner product defined in (2.5). In particular, the rate a_n can be of order $o(n^{-1/2})$ under the specific conditions such that $a = 1, k = m + 1$ with $m > 2$, and $h = O(n^{-1/(2k)})$.

The Bahadur representation greatly facilitates the study of the joint limit distribution for the scalar estimators and the estimator of the functional part. We define a linear operator R by

$$\langle R_u, \theta\rangle = z^\top\gamma + \int_0^1 x(t)\beta(t)dt \quad \text{for any } u \in \mathcal{U} \text{ and } \theta \in \mathcal{H}. \tag{3.9}$$

Theorem 3 derives the joint limit distribution of $\hat{\gamma}_{n,\lambda}$ and $\int_0^1 x_0(t)\hat{\beta}_{n,\lambda}(t)dt$

for any $x_0 \in L^2(\mathbb{I})$. Define $\tilde{x}_0 = x_0 \cdot {\sigma_{x_0}}^{-1}$, where $\sigma_{x_0}^2 = \sum_{v=1}^{\infty} \frac{|x_v^0|^2}{(1+\lambda\rho_v)^2}$ and $x_v^0 = \int_0^1 x_0(t)\varphi_v(t)dt$.

Theorem 3. *Suppose that the conditions in Theorem 2 are satisfied. Assume $\|R_{u^*}\| = O(1)$ for any $u^* = (\tilde{z}, \tilde{x}_0)$ where $\tilde{z} \in \mathbb{R}^p$, and $E\{\exp(s^*|\epsilon|)\} < \infty$ for some $s^* > 0$. Besides, there exists $b \in ((2a+1)/2k, a/k+1]$ satisfying*

$$\sum_v |V(\tilde{\beta}_j, \varphi_v)|^2 \rho_v^b < \infty \quad \textit{for any } j = 1, \cdots, p. \tag{3.10}$$

Meanwhile, if $na_n^2 = o(1)$, $nh^{4k} = o(1)$ and $nh^{2a+1}(\log n)^{-4} \to \infty$ hold, and $\beta_0 = \sum_v b_v \varphi_v$ satisfies the condition $\sum_v b_v^2 \rho_v^2 < \infty$. Then as $n \to \infty$ we have

$$\begin{pmatrix} \sqrt{n}(\hat{\gamma}_{n,\lambda} - \gamma_0) \\ \frac{\sqrt{n}}{\sigma_{x_0}}(\int_0^1 x_0(t)\hat{\beta}_{n,\lambda}(t)dt - \int_0^1 x_0(t)\beta_0(t)dt) \end{pmatrix} \xrightarrow{d} N(0, \Psi),$$

where

$$\Psi = \begin{pmatrix} \Omega_1^{-1} & 0 \\ 0 & 1 \end{pmatrix}.$$

Condition (3.10) is essential to obtain the asymptotic independence between the scalar estimators and the estimator of the functional part. It is also vital to guarantee the $\sqrt{n}$ consistency of the parametric estimators. It controls the decay rates of the coefficients for the projection $\mathbf{G}(X)$. To be more specific, the coefficients $V(\tilde{\beta}_j, \varphi_v)$ are required to converge to zero at a faster rate than v^{-kb} since $\rho_v \asymp v^{2k}$. It is a nontrivial extension to the results in Cheng and Shang (2015) by dealing with the whole curves rather than single points.

Theorem 3 helps to construct the joint confidence interval for the scalar estimates and the estimate of the functional part by constructing marginal ones. Also, it will simplify the construction of the prediction interval for a new response with given new covariates.

4. Hypothesis Testing

In this section, we develop a novel test to investigate the effects of functional and scalar covariates on the response, such as testing the significance of a given model, testing the effect of the functional covariate on the response, and testing the effects of the scalar covariates on the outcome.

Consider the following hypothesis

$$H_0 : \theta = \theta_0 \quad \text{v.s.} \quad H_1 : \theta \in \mathcal{H} - \theta_0.$$

Without loss of generality, let $\theta_0 = (\gamma_0, \beta_0) = 0$. It corresponds to testing the significance of the model. We propose a penalized likelihood ratio test for the hypothesis above as follows,

$$T_P = -2n\{\ell_{n,\lambda}(\theta_0) - \ell_{n,\lambda}(\hat{\theta}_{n,\lambda})\}, \tag{4.1}$$

where $\hat{\theta}_{n,\lambda}$ is the maximizer of $\ell_{n,\lambda}(\theta)$ over $\mathcal{H}$. The following theorem states the null limit distribution of the proposed test statistic.

Theorem 4. *Assume that the conditions in Theorem 2 are satisfied, and as $n \to \infty$, h satisfies the following conditions:*

$$nh^{2k+1} = O(1),\ nh \to \infty,\ n^{1/2}a_n = o(1),\ nr_n^3 = o(1),$$

$$n^{1/2}h^{-(a+1/2+(2k-2a-1)/(4m))}r_n^2(\log n)^2(\log\log n)^{1/2} = o(1),$$

$$n^{1/2}h^{-(2a+1+(2k-2a-1)/(4m))}r_n^3(\log n)^3(\log\log n)^{1/2} = o(1).$$

Besides, there exists a positive constant $M_4 > 0$ such that $E\{\epsilon^4|U\} \le M_4$ a.s., and condition (3.10) holds. Then under H_0, as $n \to \infty$, we have

$$\sigma^2 T_P \xrightarrow{d} N(u_n + p\sigma^2, 2u_n + 2p\sigma^2),$$

where $u_n = h^{-1}\sigma_1^4/\sigma_2^2$, $\sigma^2 = \sigma_1^2/\sigma_2^2$ and $\sigma_l^2 = h\sum_v(1+\lambda\rho_v)^{-l}$ for $l = 1, 2$.

The proof of Theorem 4 relies on the Bahadur representation and the inner products defined in (2.5) and (2.6). Except for similar assumptions about h and n in Shang and Cheng (2015), extra conditions on Z and the relationship between Z and $X(t)$ are included to derive the null limit distribution. Specifically, Assumptions 4 and 6 play an important role in deriving the Bahadur representation. The condition (3.10) here is utilized to guarantee that the projection $\mathbf{G}(X)$ can be approximate by an inner product $\langle \boldsymbol{A}, \tau(X)\rangle_1$ satisfying $E\{I(U)(Z - \langle \boldsymbol{A}, \tau(X)\rangle_1)X\} = 0$, where $\boldsymbol{A}$ and $\tau(X)$ are defined in Assumption 6.

By observing the mean and variance of the null limit distribution, we can find that $\sigma^2 T_P$ is approximately distributed as a chi-square distribution $\chi^2_{u_n+p\sigma^2}$. The degree of freedom $p\sigma^2 + u_n$ is different from the degree of freedom u_n in Shang and Cheng (2015). Notice that $p\sigma^2$ is caused by the scalar covariates. In practice, we can construct the eigensystem $(\hat{\rho}_v, \hat{\varphi}_v)$ similar to the procedures in Section S.5 of Shang and Cheng (2015), and estimate σ_l^2 by using the leading $O(h^{-1})$ eigenvalues.

Testing the associations between the covariates and the response offers a comprehensive way to decide whether a given variable should be involved in the model. In addition to testing $H_0 : \beta = 0$ and $\gamma = 0$, two other testings $H_0 : \beta = 0$ and $H_0 : \gamma = 0$ are often of interest. These respectively correspond to testing the significance of the functional variable and that of the scalar covariates. The test statistic can be modified to

$$T_P = -2n\{\ell_{n,\lambda}(\hat{\theta}_0) - \ell_{n,\lambda}(\hat{\theta}_{n,\lambda})\}, \tag{4.2}$$

where $\hat{\theta}_0$ is the maximizer under the null hypothesis. We present the null limit distributions of the penalized likelihood ratio test statistic for the two cases in the following corollary.

Corollary 1. *(a) For $H_0 : \beta = 0$, if the conditions in Theorem 4 hold, then $\sigma^2 T_P \xrightarrow{d} \chi^2_{u_n}$.*

(b) For $H_0 : \gamma = 0$, with the same conditions in Theorem 3, we have $T_P \xrightarrow{d} \chi^2_p$.

The results in Corollary 1 can be easily verified. The null limit distribution derived in (a) can be obtained similar to the procedures in Shang and Cheng (2015). The null limit distribution in (b) can be derived directly from Theorem 3. It is easy to extend the result of (b) to assess the effects of some selected p_1 covariates from the p covariates, which results in a p_1 degree of freedom.

5. Simulation Studies

In this section, we investigate the finite sample performance of the proposed method based on three commonly used testing problems: testing the significance of a given model, testing the effect of the functional covariate, and testing the effects of the scalar covariates.

Simulated data are generated from two widely used models. The first one is the partial functional linear model (PFLM)

$$Y = Z^\top \gamma + \int_0^1 X(t)\beta_0(t)dt + \epsilon,$$

and the second is the partial functional logistic regression model (PFLGRM)

$$P(Y = 1|X, Z) = \frac{\exp(Z^\top \gamma + \int_0^1 X(t)\beta_0(t)dt)}{1 + \exp(Z^\top \gamma + \int_0^1 X(t)\beta_0(t)dt)}$$

for $Y \in \{0,1\}$. We adopt the generalized cross-validation (GCV) to select the roughness penalty parameter λ. The nominal significance level is chosen to be 5%. Sample size $n \in \{100, 500\}$ and 1000 replications are considered for each case throughout the simulation studies.

Case 1: Testing $H_0 : \beta = 0$ and $\gamma = 0$. For $t \in [0,1]$, the functional process $X_i(t) = \sum_{j=1}^{100} \sqrt{\lambda_j}\eta_{ij}V_j(t)$ is a Brownian motion, where $\eta_{ij} \sim N(0,1)$, $\lambda_j = (j-0.5)^{-2}\pi^{-2}$, and $V_j(t) = \sqrt{2}\sin((j-0.5)\pi t)$. Each $X_i(t)$ is generated at 200 evenly spaced points on $[0,1]$. The true functional parameter is chosen in the same way as in Hilgert, Mas, and Verzelen (2013) where

$$\beta_0^{B,\xi}(t) = \frac{B}{\sqrt{\sum_{k=1}^{\infty} k^{-2\xi-1}}} \sum_{j=1}^{100} j^{-\xi-0.5} V_j(t).$$

Notice that B and ξ represent the signal strength and the smoothness, respectively. The two parameters are set to be $B \in \{0, 0.1, 0.5, 1\}$ and $\xi \in \{0.1, 0.5\}$. Meanwhile, each element of $Z_i \in \mathbb{R}^2$ is generated from the standard normal distribution $N(0,1)$. The true value of $\gamma = (\gamma_1, \gamma_2)$ is chosen to be $(0,0), (0.1,0.1), (0.2,0.2)$ or $(0.3,0.3)$. When data are from PFLM, we generate additional $\epsilon_i \sim N(0,1)$, for $i = 1, \cdots, n$.

Notice that when $B = 0$ and $(\gamma_1, \gamma_2) = (0,0)$, we obtain the sizes. Table 1 presents the empirical rejection rates under the PFLM and PFLGRM settings. It shows that the proposed test is valid in terms of achieving desirable sizes, and its powers increase as the increases of the signal strength,

Table 1: Sizes and powers when testing $H_0 : \beta = 0$ and $\gamma = 0$.

	n			$\xi = 0.1$			$\xi = 0.5$		
		(γ_1, γ_2)	$B = 0$	$B = 0.1$	$B = 0.5$	$B = 1$	$B = 0.1$	$B = 0.5$	$B = 1$
PFLM	100	(0.0,0.0)	5.5	8.5	19.8	64.4	9.6	56.9	99.3
		(0.1,0.1)	20.7	23.5	36.5	72.6	21.7	64.5	99.1
		(0.2,0.2)	65.9	63.4	75.7	91.7	65.7	86.7	99.6
		(0.3,0.3)	93.9	95.5	96.5	98.8	94.4	98.6	100
	500	(0.0,0.0)	5.2	9.4	72.9	100	16.7	99.6	100
		(0.1,0.1)	75.6	74.9	96.9	100	79.3	100	100
		(0.2,0.2)	100	100	100	100	100	100	100
		(0.3,0.3)	100	100	100	100	100	100	100
PFLGRM	100	(0.0,0.0)	5.2	4.7	7.1	14.7	5.8	11.3	42.4
		(0.1,0.1)	7.8	7.3	9.1	19.2	8.2	15.2	46.4
		(0.2,0.2)	15.1	15.2	19.0	27.5	15.6	26.6	57.9
		(0.3,0.3)	31.7	31.6	34.7	43.3	31.4	42.2	65.1
	500	(0.0,0.0)	5.1	5.2	18.8	66.8	6.9	58.2	99.8
		(0.1,0.1)	19.8	20.5	36.1	79.4	20.8	69.7	100
		(0.2,0.2)	69.6	69.9	80.0	95.7	69.1	94.0	100
		(0.3,0.3)	97.6	97.6	98.6	99.6	97.8	99.7	100

smoothness of the functional parameter, and the sample size. Meanwhile, the powers approach to one at $n = 500$.

Case 2: Testing $H_0 : \beta = 0$. In this case, we compare our test with some applicable methods of Kong, Staicu, and Maity (2016) and Su, Di, and Hsu (2017) in terms of sizes and powers under the PFLM setting. The methods proposed by Kong, Staicu, and Maity (2016) are based on FPCA, and the number of functional components is selected such that the cumulative percentage of variance (PVE) explained is 95%. Su, Di, and Hsu (2017) introduced the percentage of association-variance explained (PAVE) to order and select principal components after fitting the model with a high

PVE. We choose PVE to be 99% in the pre-fitting step, and choose PAVE to be 95% to select principal components afterwards. Samples are generated in the same way as in *Case 1*. Since the results have similar patterns for different values of (γ_1, γ_2), here we only report the sizes and powers when $(\gamma_1, \gamma_2) = (0.3, 0.3)$ for the sake of a concise presentation.

Recall that T_P denotes the proposed penalized likelihood ratio test. Let T_S, T_W, T_L and T_F denote the score test, Wald test, modified likelihood ratio test and F test in Kong, Staicu, and Maity (2016), and T_W^* be the test method of Su, Di, and Hsu (2017). Table 2 summaries the results for the two generative models. Under the PFLM setting, it is obvious that T_P performs the best in nearly all considered setups. All the methods have comparable sizes around the nominal significance level 5%. However, the proposed test generally performs better than the other tests with larger powers, especially for weak signals. This is mainly because the roughness penalty controls the model complexity in a continuous way, while the truncation parameter of FPCA may yield an imprecise control on the model complexity. Under the PFLGRM setting, the proposed test also has reasonably good performance. The sizes are around 5% and we have larger powers for stronger signals and larger sample sizes.

Case 3: Testing $H_0 : \gamma = 0$. In this case, we apply the proposed test

Table 2: Sizes and powers when $H_0 : \beta = 0$.

	n			$\xi = 0.1$			$\xi = 0.5$		
			$B = 0$	$B = 0.1$	$B = 0.5$	$B = 1$	$B = 0.1$	$B = 0.5$	$B = 1$
PFLM	100	T_P	5.6	21.1	44.7	89.9	21.6	81.9	99.7
		T_S	5.1	5.8	19.0	58.7	6.6	52.0	98.9
		T_W	5.5	6.2	19.7	59.2	7.2	53.4	99.2
		T_L	5.5	6.3	19.9	59.3	7.2	53.7	99.2
		T_F	4.9	5.6	18.4	58.1	6.2	51.6	98.8
		T_W^*	5.7	5.8	15.0	50.8	5.5	44.9	97.2
	500	T_P	5.4	23.9	91.2	100	35.2	100	100
		T_S	5.4	7.2	73.8	100	14.2	100	100
		T_W	5.6	7.3	74.1	100	14.4	100	100
		T_L	5.6	7.3	74.1	100	14.4	100	100
		T_F	5.4	7.0	73.5	100	14.1	100	100
		T_W^*	4.8	5.9	64.2	100	11.5	99.7	100
PFLGRM	100	T_P	5.1	5.2	7.7	22.1	5.6	17.1	55.2
	500	T_P	5.2	5.4	26.5	78.4	8.8	71.5	100

to test the effects of the scalar covariates. For comparison, the method of Yu, Zhang, and Du (2016) based on FPCA approach, denoted as T_n, is considered in the context of PFLM setting. The number of functional components is selected such that the cumulative PVE explained is 95%.

We adopt similar data settings as those in Yu, Zhang, and Du (2016). Specifically, each $X_i(t) = \sum_{k=1}^{50} \xi_k v_k(t)$ is generated at 200 evenly spaced points on $[0, 1]$, where $\xi_k \sim N(0, \tilde{\sigma}_k^2)$ with $\tilde{\sigma}_k^2 = ((k - 0.5)\pi)^{-2}$ and $v_k(t) = \sqrt{2}\sin((k-0.5)\pi t)$. The scalar covariate $Z \in \mathbb{R}^1$ is generated from $N(0, 1)$. The error terms are also from $N(0, 1)$ under the PFLM setting. The scalar coefficient $\gamma \in \mathbb{R}^1$ takes value from $\{0, 0.5, 1, 2, 4, 6\}/\sqrt{n}$ and the true functional coefficient is $\beta_0 = \sqrt{2}\sin(7\pi t/2) + 3\sqrt{2}\sin(9\pi t/2)$. The correlations

Table 3: Sizes and powers when testing $H_0 : \gamma = 0$.

			PFLM $C = \sqrt{n}\beta$						PFLGRM $C = \sqrt{n}\beta$					
ρ	n		0	0.5	1	2	4	6	0	0.5	1	2	4	6
0	100	T_P	4.8	8.0	16.7	50.2	96.5	100	5.3	6.2	8.5	15.9	46.7	78.4
		T_n	4.9	8.0	16.9	51.0	96.4	100	-	-	-	-	-	-
	500	T_P	4.9	7.3	15.3	53.2	98.6	100	4.9	6.3	8.9	17.3	46.6	82.5
		T_n	5.1	7.2	14.8	51.5	98.0	100	-	-	-	-	-	-
0.2	100	T_P	5.5	8.0	17.9	48.8	96.0	100	5.4	6.9	9.7	21.2	53.1	82.5
		T_n	6.0	11.0	23.9	57.8	97.5	100	-	-	-	-	-	-
	500	T_P	5.2	8.2	17.0	52.0	97.2	100	5.7	7.4	10.6	22.6	59.8	88.9
		T_n	8.8	17.0	31.0	70.5	99.3	100	-	-	-	-	-	-

between Z and ξ_k's are $Corr(Z, \xi_j) = \rho^{|j-5|+1}$, for $j = 2, \cdots, 8$. We set $\rho = 0$ and 0.2.

Table 3 contains the sizes and powers under $H_0 : \gamma = 0$. For the two models, sizes of the proposed test are close to the nominal level 5% and powers approach to 1 as the increase of the sample size and the signal strength. It also shows that higher correlations between X and Z inflate the Type I errors of T_n, while the sizes of T_P maintain well around 5%.

In general, the above simulation results give evidence that the sizes of the proposed test are reasonably controlled around the nominal level and the powers increase as sample size and signal strength become larger, which confirms our theoretical results. Moreover, to explore the effects of the observation errors of the functional trajectory numerically, we also conduct simulation studies when $X(t)$'s are observed with measurement errors in Section S5 of the Supplementary Material. If the errors are small and dense measurements are available on each curve, the sizes and powers behave similar to the sizes and powers when $X(t)$'s are fully observed. However, the proposed test is likely to lose power at a slow rate as the errors become larger.

6. Application to the Air Pollution Data

We apply our method to learn the effects of PM2.5 and other scalar factors on the non-accidental mortality rate across different cities in the U.S.. The dataset is obtained from the National Mortality, Morbidity, and Air Pollution Study, which contains air pollution measurements and mortality counts from U.S. cities collected during the census in year 2000. Similar to Kong et al. (2016), the scalar covariates are proportion of urban population (Purban), proportion of the population with at least a high school diploma (Phigh), proportion of the population with at least a university diploma (Pdeg), proportion of the population below the poverty line (Ppoverty), proportion of household owners (Powner), land area per individual (perland), and water area per individual (perwater). We focus on daily concentration measurements of PM2.5 from April 1 to August 31 in 2000. The response of interest is the log-transformed total nonaccidental mortality rate in the following month, September 2000, among individuals of age 65 and older, who account for the majority of non-accidental deaths. A total of 60 cities are included in the study after removing cities with more than ten consecutive missing measurements of PM2.5. We consider the partial functional linear regression model

$$Y = Z^{\top}\gamma + \int_{\text{Apr.1st}}^{\text{Aug.31th}} X(t)\beta(t)dt + \epsilon, \tag{6.1}$$

where Y is the log-transformed total non-accidental mortality rate, $X(t)$ denotes measurements of PM2.5, and $Z \in \mathbb{R}^7$ contains scalar covariates.

We first investigate the significance of model (6.1) by testing $H_0 : \gamma = 0$ and $\beta = 0$, and obtain a p-value 0.0001 for T_P, which implies that the proposed model is of significance at 95% nominal level.

For the null hypothesis $H_0 : \gamma = 0$, we include the test procedure T_n introduced by Yu, Zhang, and Du (2016) for comparison. The resulting p-value of T_P is 0.0004 and that of T_n is 0.0003. It suggests that there are significant scalar variables. We then further explore the association between a given scalar variable and the log-transformed total nonaccidental mortality rate. Estimates and p-values are summarized in Table 4. Our method for Phigh, Pdeg and Ppoverty gives p-values of 0.0182, 0.0103 and 0.0528, respectively, which provides evidence for the effects of Phigh and Pdeg on the mortality rate at 95% nominal level and the effect of Ppoverty at 90% nominal level. Moreover, the results of T_n show that Phigh, Pdeg

Table 4: Estimates and p-values when testing the effects of the scalar variables. T_P denotes the proposed test, and T_n is the method of Yu, Zhang, and Du (2016).

		Purban	Phigh	Pdeg	Ppoverty	Powner	perland	perwater
T_P	Estimate	-0.0576	0.5929	-0.5949	0.4924	0.2047	-0.1548	0.1261
	p-value	0.5438	0.0182	0.0103	0.0528	0.5631	0.3485	0.2996
T_n	Estimate	-0.0859	0.5997	-0.6042	0.4795	0.1891	-0.1493	0.1492
	p-value	0.3548	0.0008	0.0010	0.0044	0.3722	0.1377	0.1507

and Ppoverty are all effective variables on the mortality rate at 95% nominal level. The two methods give similar estimation and significance results.

We also apply the proposed test to explore the association between PM2.5 and the non-accidental mortality rate. Except for the proposed T_P, the score test (T_S), Wald test (T_W), modified likelihood ratio test (T_L) and F test (T_F) in Kong, Staicu, and Maity (2016), together with the test method of Su, Di, and Hsu (2017) (T_W^*) have been implemented. Table 5 presents the p-values of the aforementioned methods, demonstrating a significant effect of PM2.5 on the non-accidental mortality rate. This coincides with the findings of Kong et al. (2016). Specifically, the p-value of the proposed test is 0.0264, while the p-values of all the other competing methods are around 0.1.

Table 5: p-values when testing the effect of PM2.5.

Method	T_P	T_S	T_W	T_L	T_F	T_W^*
p value	0.0264	0.0906	0.0954	0.0896	0.1051	0.0914

We conclude that PM2.5, Phigh, Pdegree and and Ppoverty have significant impacts on the mortality rate of elder residents in U.S. cities.

7. Discussion

In this paper, we proposed a penalized likelihood ratio test for the generalized partial linear models based on a Bahadur representation for both functional and scalar estimators, and showed that the scalar estimators are asymptotically independent with the estimator of the functional part. A primary advantage of the proposed test is that it allows simultaneous testing as well as separate testings to both functional and scalar parameters. The empirical analysis of the proposed test demonstrated that the proposed test behaves well by respecting sizes and having good powers.

Our methodology and theoretical results are based on the assumption that $X(t)$ is smooth and fully observed without noise. A natural but nontrivial extension is to deal with intermittently and noisily observed curves. When $X(t)$ is observed with measurement errors, a popular approach is to obtain an estimate of $X(t)$ using a nonparametric method and then treat the estimate as the fully observed functional variable. One may refer to Hall, Müller, and Wang (2006) for more details about the procedure. In practice, the pre-smoothing step is applicable to the considered problem, especially when the variance of the measurement errors is small and dense measurements are available on each curve, which is evidenced by additional simulations in Section S5 of the Supplementary Material.

Although the effects of measurement errors on functional linear models based on FPCA has been addressed in the literature (Zhang and Chen, 2007; Li, Wang, and Carroll, 2010; Wong, Li, and Zhu, 2018), to the best of our knowledge, there has been little work about how measurement errors influence the theoretical results under the RKHS framework. Different from the FPCA method with noisily observed functional variables, where the estimation and inference procedures are similar to the parametric models after truncation, the functional parameter in this paper is represented through the eigen-system defined in Assumption 3 without truncation throughout the proofs. We obtain the theoretical results by virtue of the Fréchet derivatives defined on the Banach space other than the derivatives defined on the Euclidean space in FPCA. Moreover, technical proofs rely on the inner products defined in (2.5) and (2.6). The convergence rate and the Bahadur representation are derived through these inner products, which involve the fully observed trajectory, instead of the L^2 norm used in FPCA. Therefore, the techniques applied in the FPCA method to deal with measurement errors are not fully applicable under the RKHS framework.

In Section S4 of the Supplementary Material, we discuss in more details on the potential challenges to deriving the theory when the functional covariate is observed intermittently and with errors. We will pursue this direction in our future work.

Supplementary Material

Additional simulation results, calculations of some linear operators and all the technical proofs are included in the Supplementary Material.

Acknowledgment

The authors thank the Editor, the Associate Editor and two referees for their insightful comments that have significantly improved the paper. Zhu's research was partially supported by NSFC 11671096, NSFC 11690013 and NSFC 11731011.

References

Cai, T. T. and M. Yuan (2012). Minimax and adaptive prediction for functional linear regression. *J. Amer. Statist. Assoc. 107*, 1201–1216.

Cheng, G. and Z. Shang (2015). Joint asymptotics for semi-nonparametric regression models with partially linear structure. *Ann. Statist. 43*, 1351–1390.

Du, P. and X. Wang (2014). Penalized likelihood functional regression. *Statist. Sinica 24*, 1017–1041.

Gu, C. (2013). *Smoothing spline ANOVA models.* Springer Science & Business Media.

Hall, P., H.-G. Müller, and J.-L. Wang (2006). Properties of principal component methods for functional and longitudinal data analysis. *Ann. Statist. 34*, 1493–1517.

Hilgert, N., A. Mas, and N. Verzelen (2013). Minimax adaptive tests for the functional linear model. *Ann. Statist. 41*, 838–869.

Kong, D., A.-M. Staicu, and A. Maity (2016). Classical testing in functional linear models. *J. Nonparametr. Stat. 28*, 813–838.

Kong, D., K. Xue, F. Yao, and H. H. Zhang (2016). Partially functional linear regression in high dimensions. *Biometrika 103*, 147–159.

Li, Y., N. Wang, and R. J. Carroll (2010). Generalized functional linear models with semiparametric single-index interactions. *Journal of the American Statistical Association 105*(490), 621–633.

Mammen, E. and S. van de Geer (1997). Penalized quasi-likelihood estimation in partial linear models. *Ann. Statist. 25*, 1014–1035.

Ramsay, J. O. and B. W. Silverman (2002). *Applied functional data analysis: methods and case studies*. Citeseer.

Ramsay, J. O. and B. W. Silverman (2005). *Functioanl data analysis (2nd edition)*, Volume 3. Springer.

Shang, Z. and G. Cheng (2015). Nonparametric inference in generalized functional linear models. *Ann. Statist. 43*, 1742–1773.

Shin, H. and M. H. Lee (2012). On prediction rate in partial functional linear regression. *J. Multivariate Anal. 103*, 93–106.

Su, Y.-R., C.-Z. Di, and L. Hsu (2017). Hypothesis testing in functional linear models. *Biometrics 73*, 551–561.

Wong, R. K., Y. Li, and Z. Zhu (2018). Partially linear functional additive models for multivariate functional data. *Journal of the American Statistical Association*, 1–13.

Yu, P., Z. Zhang, and J. Du (2016). A test of linearity in partial functional linear regression. *Metrika 79*, 953–969.

Yuan, M. and T. T. Cai (2010). A reproducing kernel hilbert space approach to functional linear regression. *Ann. Statist. 38*, 3412–3444.

Zhang, J.-T. and J. Chen (2007). Statistical inferences for functional data. *The Annals of Statistics 35*(3), 1052–1079.

Department of Statistics, Fudan University, Shanghai 200433, China.

E-mail: tingli16@fudan.edu.cn

Department of Statistics, Fudan University, Shanghai 200433, China.

E-mail: zhuzy@fudan.edu.cn