Statistica Sinica **34** (2024), 1973-1995
doi:https://doi.org/10.5705/ss.202022.0083

A PROJECTION-BASED DIAGNOSTIC TEST FOR GENERALIZED FUNCTIONAL REGRESSION MODELS

Guizhen Li[1], Mengying You[2], Ling Zhou[1], Hua Liang[3] and Huazhen Lin*[1]

[1]*Southwestern University of Finance and Economics*
[2]*Shanghai University of International Business and Economics*
and [3]*George Washington University*

Abstract: We propose a novel diagnostic test to check the goodness-of-fit for generalized functional regression models. The proposed test does not require a specification of the distribution, and can be applied to commonly employed functional regression models. Because it is based on independence in distribution, it includes mean-based and higher-order moment-based tests as special cases. In particular, we overcome the problem of the infinite dimensionality of the functional data by projecting functions along certain directions. Moreover, to avoid bias caused by the subjective selection of these directions, we integrate over the directions along which the functional variables project. As a result, the proposed test simultaneously enhances the local power and overcomes the infinite-dimensionality problem. A simple implementation procedure is developed. The performance of the proposed test is evaluated theoretically and using simulation studies. We apply the proposed procedure to analyze Canadian weather data and Chinese air pollution data, resulting in several interesting models that achieve higher interpretability and estimation accuracy than those of existing methods.

Key words and phrases: Distribution free, generalized functional regression, goodness-of-fit, local power, projection-based distribution test.

1. Introduction

Functional data analysis (FDA) has attracted considerable attention since the seminal work of Ramsay (1982). Linear, nonlinear, nonparametric, and semiparametric models for analyzing functional data have been proposed, including those of Kokoszka and Reimherr (2017); Horváth and Kokoszka (2012); Ramsay and Silverman (2002); Hsing and Eubank (2015), and Ferraty and Vieu (2006), leading to the development of various functional regression techniques (Yao, Müller and Wang (2005); Li and Hsing (2010); Li, Wang and Carroll (2010)), and their applications (Horváth and Kokoszka (2012)).

Checking the goodness of fit for a functional regression was first investigated by Cardot, Ferraty and Sarda (2003), prompting further research on model checking for functional regressions; for example, see Kokoszka et al. (2008),

*Corresponding author.

Chiou and Müller (2007), García-Portugués, González-Manteiga and Febrero-Bande (2014), Cuesta-Albertos et al. (2019), Lei (2014), Patilea, Sánchez-Sellero and Saumard (2016), and Lee, Zhang and Shao (2020) for functional linear regression (FLR) models, and McLean, Hooker and Ruppert (2015) for functional generalized additive models.

Although existing goodness-of-fit methods have certain useful properties, such as computational efficiency for parametric functional regression models, or avoiding imposing error distributions, they have limitations. For example, some methods may inherit the "curse-of-dimensionality" problem, as in nonparametric regression, from evaluating the difference between the conditional expectation under the null and alternative hypotheses, and the expectation of the residual under the null hypothesis; for example, see Delsol, Ferraty and Vieu (2011) and Chiou and Müller (2007). Other methods may produce intermittent quantities, causing the selection of user-chosen quantities, such as bandwidths (Patilea, Sánchez-Sellero and Saumard (2016); Lei (2014)). To ensure freedom from the curse of dimensionality, Patilea, Sánchez-Sellero and Saumard (2016) propose a nonparametric test based on a quadratic form, with univariate nearest-neighbor smoothing, for either multidimensional or functional covariates. Their test statistics converge to a standard normal distribution under the null hypothesis, and exhibit good finite-sample performance. However, their test's local power depends on the user-chosen parameter, namely, the bandwidth of the kernel, and achieves only $O((nh^{1/2})^{-1/2})$, with n and h being the sample size and the bandwidth, respectively. In addition, with the exception of Chiou and Müller (2007) and McLean, Hooker and Ruppert (2015), existing works focus on linear functional regression models or specific error distributions. In particular, for Gaussian error distributions, Lei (2014) proposes an exponential scan test, which shows to be uniformly powerful over a certain class of smooth alternatives if the signal-to-noise ratio exceeds the detection boundary.

In addition to the aforementioned limitations, a common problem with these methods is that they focus on modeling/testing the conditional mean of the response variable, given the covariates. Suppose $Y = \mathbb{E}(Y \mid X) + \varepsilon$, where covariate X is function-valued or vector-valued, and ε is the unpredictable part of Y given X. The following hypothesis is commonly considered in the literature:

$$H_0 : \mathbb{E}(\varepsilon \mid X) = 0 \quad \text{almost surely (a.s.)}, \tag{1.1}$$

against the nonparametric alternative $\text{Prob}\{\mathbb{E}(\varepsilon \mid X) = 0\} < 1$. To maintain the local power with the classic parametric rate, $O(n^{-1/2})$, and to avoid imposing an error distribution assumption, Lee, Zhang and Shao (2020) propose a nonparametric test that uses the functional martingale difference divergence to fully characterize the conditional mean dependence of the response and the covariates, both of which can be function-valued or vector-valued.

The mean-based test does not consider the higher-order conditional moment, which is often of interest for functional data. Notably, the second-order covariance function is an essential feature in FDA (Ramsay and Silverman (2002); Wang, Chiou and Müller (2016)). However, the test given in (1.1) cannot check the goodness of fit for a functional regression model with a covariate-dependence second-order moment, as we observe in Table 4 for the example of Chinese air pollution. Specifically, the mean-based hypothesis (1.1) does not detect a relationship between the air quality index $Y_i(t)$ and PM2.5, whereas the proposed distribution-based test suggests that the variance of the air quality index depends on PM2.5. On the other hand, our theoretical and numerical results show that a moment-based test is more powerful than a distribution-based test. Motivated by these issues, we consider a generalized functional regression test (GFR-test) that includes moment-based tests and distribution-based tests as special cases, that is,

$$H_0 : \mathbb{E}(\mathcal{L}(\varepsilon) \mid X) = \mathbb{E}(\mathcal{L}(\varepsilon)), \tag{1.2}$$

against the nonparametric alternative $\text{Prob}\{\mathbb{E}(\mathcal{L}(\varepsilon) \mid X) = \mathbb{E}(\mathcal{L}(\varepsilon))\} < 1$, where $\mathcal{L}$ is a certain prespecified function. For instance, the proposed test is a mean-based test when $\mathcal{L}(\varepsilon) = \varepsilon$, and is a variance-based test when $\mathcal{L}(\varepsilon) = \varepsilon^2$. For $\mathcal{L}(\varepsilon) = \{I(\varepsilon < v) : v \in \mathcal{R}\}$, where $I(\cdot)$ is an indicator function, the proposed test becomes the following distribution-based test:

$$H_0 : F_{\varepsilon|X}(v) = F_\varepsilon(v) \text{ a.s. } \forall v \in \mathcal{R}, \tag{1.3}$$

against the nonparametric alternative $\text{Prob}\{F_{\varepsilon|X}(v) = F_\varepsilon(v)\} < 1$, for some $v \in \mathcal{R}$, where $F_\varepsilon(\cdot)$ and $F_{\varepsilon|X}(\cdot)$ denote the distribution function of the random variable ε and the conditional distribution function of ε given X, respectively.

Within the test framework (1.2), we start with the conditional mean test. Only if the mean-based test is not rejected do we then apply the independence test to test whether heterogeneous higher-order moments exist. Following this strategy, we obtain more precise and insightful information about the model structure and avoid calculating redundant test statistics. For example, as shown in Table 3 for the Canadian weather data, both the mean-based test and the distribution-based test suggest a correlation between rainfall and temperature. Furthermore, we find that none of the heterogeneous variance models are rejected, whereas all regression models without an analysis of variance (ANOVA) or a heterogeneous variance structure are rejected. This result implies heterogeneity in the rainfall of different climatic zones, and that the variance depends on temperature. This finding is consistent with the conclusion of Patilea, Sánchez-Sellero and Saumard (2016), who use ANOVA models to take into account the variance heterogeneity among climatic zones. Notably, heterogeneous variance models are more insightful and efficient than ANOVA models in terms of interpretability and accuracy, especially when the number of factor groups increases.

In this paper, we provide a unified test framework (1.2) for generalized functional regression models that allows nonlinear functional regression models, and hence, includes numerous such models as special cases, as described in Section 2. Furthermore, the framework accommodates not only the independence test, which is distribution based, but also mean and higher-order moment-based tests, enabling us to compare the mean-based tests and the independence tests. Specifically, the theoretical and numerical results show that a moment-based test is more powerful than an independence test. This finding is understandable, because a moment-based test will not be rejected if the distribution-based test is not rejected, yielding a smaller alternative space for moment-based testing than that of distribution-based testing.

Although some works exist on statistical independence tests for traditional regression models with scalar or vector variables of finite dimension, for example, Neumeyer (2009) and Dhar, Bergsma and Dassios (2018), to the best of our knowledge, there is no statistical independence test for functional regression models. For such models, the challenge when testing (1.2) is that the conditional variable $X(\cdot)$, which determines the conditional moment or distribution, is a function of infinite dimensionality. In this paper, inspired by Escanciano (2006) and motivated by the established equivalency between $\mathbb{E}(\mathcal{L}(\varepsilon)I(X < u)) = \mathbb{E}(\mathcal{L}(\varepsilon))F_X(u), \forall u$ and $\mathbb{E}\left[\mathcal{L}(\varepsilon)I(\langle X, \boldsymbol{\alpha}\rangle_m \leq u)\right] = \mathbb{E}(\mathcal{L}(\varepsilon))F_{\langle X,\boldsymbol{\alpha}\rangle_m}(u)$ for any $\boldsymbol{\alpha}$ and u, given in Proportion 1, we propose a Crámer-von Mises-type test for (1.2), and overcome the infinite dimensionality problem of the functional data by projecting the function along various directions. Moreover, to avoid the bias from the subjective selection of the directions, we integrate over all directions. The proposed approach is both robust and powerful, because it is constructed based on the distribution, but without prespecifying its form. In particular, the proposed test is shown to achieve the parametric order $O(n^{-1/2})$ for the local power, which even tests based on the conditional moment approach do not attain (Patilea, Sánchez-Sellero and Saumard (2016)). A simple implementation procedure is also developed.

The remainder of the paper is organized as follows. Section 2 presents the Crámer-von Mises-type test for a general model with functional data. Section 3 presents the asymptotic distributions of the proposed test statistics. An implementation procedure is introduced in Section 4. The performance of the proposed statistics is assessed using simulation studies in Section 5. In Section 6, we apply our proposed method to data on Canadian weather and Chinese air pollution, resulting in several interesting models. We provide concluding remarks in Section 7. Technical details, including notation, conditions, and all proofs are relegated to the Supplementary Material.

2. Model and Method

Denote $\langle X, \beta \rangle = \int X(t)\beta(t)dt$. We consider the following model:

$$Y = g\left(a, \boldsymbol{Z}, \langle X_1, \beta_1 \rangle, \ldots, \langle X_d, \beta_d \rangle, \epsilon\right), \tag{2.1}$$

where $g(\cdot)$ is a known link function, $\boldsymbol{Z}$ is a vector of covariates, X_j are functional covariates with mean zero, and both $\boldsymbol{Z}$ and $\{X_j\}_{j=1}^d$ are independent of the random error ϵ. Here, Y is a scalar and the error ϵ has mean zero. Note that Y can also be a function-valued response, $Y(t)$, in which case, model (2.1) is then rewritten as

$$Y(t) = g\left(a(t), \boldsymbol{Z}(t), \langle X_1, \beta_1(\cdot, t) \rangle, \ldots, \langle X_d, \beta_d(\cdot, t) \rangle, \epsilon(t)\right), \tag{2.2}$$

where $\langle X_j, \beta_j(\cdot, t) \rangle = \int X_j(s)\beta_j(s, t)ds$, and $\epsilon(t)$ is a mean zero process.

Models (2.1) and (2.2) cover most functional regression models as special cases. First, they include functional linear models (FLMs), including those with a scalar response (FLMsR): $Y = \langle X, \beta \rangle + \epsilon$, a function-valued response (FLMfR): $Y(t) = \langle X, \beta(\cdot, t) \rangle + \epsilon(t)$, and a concurrent response (FLMcR): $Y(t) = X(t)\beta(t) + \epsilon(t)$; see Cai and Hall (2006); Fan and Zhang (2000), and Hall and Horowitz (2007). Second, they include the partially functional linear varying-coefficient models of Feng and Xue (2016) and Li, Huang and Zhu (2017): $Y = \sum_{d=1}^D Z_d\beta_d(u) + \langle X, \alpha \rangle + \epsilon$ and $Y(t) = \sum_{d=1}^D Z_d\beta_d(t) + \langle X, \alpha \rangle + \epsilon(t)$. Third, they include the generalized functional linear models of Muller and Stadtmuller (2005) and McLean et al. (2014): $Y = g\left(\alpha + \langle X, \beta \rangle\right) + \epsilon$. Fourth, they include the multiple index functional regression models (MiFRMs) of Chen, Hall and Muller (2011), Ma (2016), Ding et al. (2017), and Tang et al. (2021): $Y = \sum_{d=1}^D g_d(\langle X, \beta_d \rangle) + \epsilon$, and $Y = g_1(\langle X, \beta_1 \rangle + \boldsymbol{Z}^T\boldsymbol{a}_1) + g_2\left(\langle X, \beta_2 \rangle + \boldsymbol{Z}^T\boldsymbol{a}_2\right)\epsilon$, which are special examples of our models when the link functions are specified by their estimates.

For the test problem (1.2), we express the scalar residual $\varepsilon(\boldsymbol{X}; \boldsymbol{\beta})$ and the function-valued residual $\varepsilon^f(\boldsymbol{X}; \boldsymbol{\beta})(t)$ based on models (2.1) and (2.2), respectively, as

$$\begin{aligned} \varepsilon(\boldsymbol{X}; \boldsymbol{\beta}) &\hat{=} m\left(Y, a, \boldsymbol{Z}, \langle X_1, \beta_1 \rangle, \ldots, \langle X_d, \beta_d \rangle\right) \text{ and} \\ \varepsilon^f(\boldsymbol{X}; \boldsymbol{\beta})(\cdot) &\hat{=} m\left(Y(\cdot), a(\cdot), \boldsymbol{Z}(\cdot), \langle X_1, \beta_1(\cdot) \rangle, \ldots, \langle X_d, \beta_d(\cdot) \rangle\right), \end{aligned}$$

where $\boldsymbol{X} = (X_1, \ldots, X_d)^T$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_d)^T$, $m(\cdot)$ is a known function determined by $g(\cdot)$, and the superscript f indicates that the variable is function-valued. For example, for the FLMsR model, the residual $\varepsilon(X_i; \beta) := Y_i - \int_0^1 \beta(t)X_i(t)dt$; for the FLMfR model, the residual $\varepsilon^f(X_i; \beta) := Y_i(\cdot) - \int_0^1 \beta(\cdot, s)X_i(s)ds$. Because the purpose of this study is to determine the effect of functional covariates, henceforth, we disregard the dependence of $\varepsilon(\boldsymbol{X}; \boldsymbol{\beta})$ and $\varepsilon^f(\boldsymbol{X}; \boldsymbol{\beta})(t)$ on $(a, \boldsymbol{Z})$, for notational simplicity.

Next, the GFR-test problem (1.2) for model (2.1) is rewritten as

$$\mathcal{H}_0 : \text{Prob}\left[\mathbb{E}\{\mathcal{L}(\varepsilon(\boldsymbol{X};\boldsymbol{\beta})) \mid \boldsymbol{X} = \boldsymbol{x}\} = \mathbb{E}\{\mathcal{L}(\varepsilon(\boldsymbol{X};\boldsymbol{\beta}))\}\right] = 1,$$
$$\text{for some functions } \boldsymbol{\beta}(\cdot),$$

against the alternative hypothesis

$$\mathcal{H}_1 : \text{Prob}\left[\mathbb{E}\{\mathcal{L}(\varepsilon(\boldsymbol{X};\boldsymbol{\beta})) \mid \boldsymbol{X} = \boldsymbol{x}\} = \mathbb{E}\{\mathcal{L}(\varepsilon(\boldsymbol{X};\boldsymbol{\beta}))\}\right] < 1,$$
$$\text{for any function } \boldsymbol{\beta}(\cdot).$$

The test problem is defined for model (2.2), with $\varepsilon(\boldsymbol{X};\boldsymbol{\beta})$ replaced with $\varepsilon^f(\boldsymbol{X};\boldsymbol{\beta})(t)$.

The projection-based distribution-free test statistic. Here, we construct the projection-based distribution-free (PD) test statistic, and show the associated theory mainly for a scalar ε. The test and theory for a function-valued $\varepsilon(t)$ are similar, and are discussed next. Throughout this paper, we assume that all covariates $\{X_{i,j}(\cdot)\}_{j=1}^d$ have mean zero. We use $\langle\cdot,\cdot\rangle$ to denote the inner product in $L^2[0,1]$, that is, $\langle W_1, W_2\rangle = \int_0^1 W_1(t)W_2(t)dt$, $\forall\, W_1, W_2 \in L^2[0,1]$.

We assume that the covariance function of $X_{i,j}(\cdot)$ is $\Sigma_j(s,t) = \text{cov}(X_{i,j}(s), X_{i,j}(t))$, for $j = 1,\ldots,d$. Mercer's theorem implies that the spectral decomposition of Σ_j leads to $\Sigma_j(s,t) = \sum_{k=1}^\infty \theta_{j,k}\phi_k(s)\phi_k(t)$, with uniform convergence, where $\theta_{j,k}$ are the eigenvalues and ϕ_k are the corresponding orthonormal eigenfunctions (Wang, Chiou and Müller (2016)). According to the Karhunen–Loève (KL) theorem, we have $X_{i,j}(t) = \sum_{k=1}^\infty \langle X_{i,j}, \phi_k\rangle\phi_k(t) = \sum_{k=1}^\infty \xi_{ij,k}\phi_k(t)$, where $\xi_{ij,k}$ $\xi_{ij,k}$ are pairwise, uncorrelated, mean-zero functional principal component scores (FPCs) of $X_{i,j}$, with variance $Var(\xi_{ij,k}) = \theta_{j,k}$. Furthermore, for a nonrandom p-dimensional vector $\boldsymbol{\alpha}_j = \{\alpha_{j,k}\}_{k=1}^p \in \mathcal{R}^p$, we define the product of the covariate function $X_{i,j}(t)$ and $\boldsymbol{\alpha}_j$ as the product of $X_{i,j}(t)$ and an element of $L^2[0,1]$, for example, $M_j(t) = \sum_{k=1}^\infty \langle M_j, \phi_k\rangle\phi_k(t)$, with coordinates $\langle M_j, \phi_k\rangle = \alpha_{j,k}$ for $k \le p$ and $\langle M_j, \phi_k\rangle = 0$ for $k > p$ in the basis $\mathcal{B} = \{\phi_k(\cdot), k \ge 1\}$, that is, $\langle X_{i,j}, \boldsymbol{\alpha}_j\rangle_m := \langle X_{i,j}, M_j\rangle = \sum_{k=1}^p \xi_{ij,k}\alpha_{j,k}$. Using this definition, we overcome the problem of the infinite dimensionality of the function $X_{i,j}(t)$ by projecting function $X_{i,j}(t)$ along the direction $\boldsymbol{\alpha}_j$. As long as $\boldsymbol{\alpha}_j$ includes all directions, from Proposition 1, all information about $X_{i,j}(t)$ is captured by $\langle X_{i,j}, \boldsymbol{\alpha}_j\rangle_m, \forall\boldsymbol{\alpha}_j$. Let $\boldsymbol{X}_i = \{X_{i,j}(\cdot)\}_{j=1}^d$, $\boldsymbol{\beta} = \{\beta_j(\cdot)\}_{j=1}^d$ and $\boldsymbol{\alpha} = \{\boldsymbol{\alpha}_j\}_{j=1}^d$. Denote $\langle \boldsymbol{X}_i, \boldsymbol{\alpha}\rangle_m := \sum_{j=1}^d \langle X_{i,j}, \boldsymbol{\alpha}_j\rangle_m$ and $F_{\langle \boldsymbol{X},\boldsymbol{\alpha}\rangle_m}(u) = \text{Prob}\,(\langle \boldsymbol{X}_i, \boldsymbol{\alpha}\rangle_m \le u)$. The following proposition plays a key role in motivating our methods.

Proposition 1. *We use $U, X_j \in L^2[0,1]$, for $j = 1,\ldots,d$, as random functions. For any p, and $\boldsymbol{\gamma} \in \mathcal{R}^p$, we denote $F_{\langle U,\boldsymbol{\gamma}\rangle_m}(u) = Prob\,(\langle U_i, \boldsymbol{\gamma}\rangle_m \le u)$, and $F_{\langle U,\boldsymbol{\gamma}\rangle_m|\boldsymbol{X}}(u) = \mathbb{E}\left[I\left(\langle U_i, \boldsymbol{\gamma}\rangle_m \le u\right) \mid \boldsymbol{X}\right]$. Next, the following statements (a) and (b) are equivalent: (a) $F_{\langle U,\boldsymbol{\gamma}\rangle_m|\boldsymbol{X}}(\cdot) = F_{\langle U,\boldsymbol{\gamma}\rangle_m}(\cdot)$ a.s., and (b) $F_{\langle U,\boldsymbol{\gamma}\rangle_m|\langle \boldsymbol{X},\boldsymbol{\alpha}\rangle_m}(\cdot) =$*

$F_{\langle U,\boldsymbol{\gamma}\rangle_m}(\cdot)$ *a.s.* $\forall p \geq 1$, $\forall \boldsymbol{\alpha}_j \in \mathbb{S}^p$, *for* $j = 1, \ldots, d$, *where* $\mathbb{S}^p = \{\boldsymbol{\alpha} \in \mathcal{R}^p : \|\boldsymbol{\alpha}\|_2 = 1\}$ *denotes the unit hypersphere in* $\mathcal{R}^p$.

Proposition 1 indicates that the test problem (1.2), $\mathbb{E}\left[\mathcal{L}(\varepsilon)I(\boldsymbol{X} \leq \boldsymbol{u})\right] = \mathbb{E}(\mathcal{L}(\varepsilon))F_{\boldsymbol{X}}(\boldsymbol{u})$, for any $\boldsymbol{u}$, is equal to the projected test problem $\mathbb{E}[\mathcal{L}(\varepsilon)I(\langle \boldsymbol{X}, \boldsymbol{\alpha}\rangle_m \leq u)] = \mathbb{E}(\mathcal{L}(\varepsilon))F_{\langle \boldsymbol{X},\boldsymbol{\alpha}\rangle_m}(u)$, $\forall \boldsymbol{\alpha}_j \in \mathbb{S}^p$, for $j = 1, \ldots, d$ and any u.

To explicitly express our proposed PD test statistics, we consider three special formulae of $\mathcal{L}(\varepsilon)$. For $\mathcal{L}(\varepsilon) = \{I(\varepsilon < v) : v \in \mathcal{R}\}$, let the empirical version of $F_\varepsilon(v)$ be $F_{n,\varepsilon}(v) = n^{-1}\sum_{i=1}^n I(\varepsilon(\boldsymbol{X}_i;\boldsymbol{\beta}) \leq v)$. For a scalar $\varepsilon(\boldsymbol{X}_i;\boldsymbol{\beta})$, we define the independence test statistic

$$M_{n,\mathtt{F}}(\boldsymbol{\beta};\boldsymbol{\alpha},u,v) = n^{-1/2}\sum_{i=1}^n \left[I\left(\varepsilon(\boldsymbol{X}_i;\boldsymbol{\beta}) \leq v\right) - F_{n,\varepsilon}(v)\right] I\left(\langle \boldsymbol{X}_i, \boldsymbol{\alpha}\rangle_m \leq u\right).$$

To avoid a subjective selection of $\boldsymbol{\alpha}$, which may cause the test to be inconsistent (Escanciano (2006)), we consider integrating all possible $\boldsymbol{\alpha}$. In particular, we consider the following PD independence test statistic

$$T_{n,\mathtt{F}}(\boldsymbol{\beta}) = \int_{\mathbb{S}^{pd}} \iint_{\mathcal{R}^2} \left(M_{n,\mathtt{F}}(\boldsymbol{\beta};\boldsymbol{\alpha},u,v)\right)^2 F_{n,<\boldsymbol{X},\boldsymbol{\alpha}>_m}(du) \times F_{n,\varepsilon}(dv)d\boldsymbol{\alpha},$$

where $F_{n,\langle \boldsymbol{X},\boldsymbol{\alpha}\rangle_m}(u)$ is the empirical version of $F_{\langle \boldsymbol{X},\boldsymbol{\alpha}\rangle_m}(u)$.

Remark 1. For a function-valued $\varepsilon^f(\boldsymbol{X};\boldsymbol{\beta})(t)$, we induce another p-dimensional vector $\boldsymbol{\gamma} := \{\gamma_j\}_{k=1}^p \in \mathcal{R}^p$ to project $\varepsilon^f(\boldsymbol{X};\boldsymbol{\beta})(t)$ along the direction $\boldsymbol{\gamma}$. Specifically, using the KL expression, we have $\epsilon^f(\boldsymbol{X}_i;\boldsymbol{\beta})(\cdot) = \sum_{k=1}^\infty e_{i,k}\phi_k(\cdot)$. For any fixed $\boldsymbol{\gamma}$, the projection of the residual $\langle \varepsilon^f(\boldsymbol{X}_i;\boldsymbol{\beta}), \boldsymbol{\gamma}\rangle_m$ is a scalar, with an empirical marginal distribution of the form $F_{n,\langle \varepsilon^f,\boldsymbol{\gamma}\rangle_m}(v) = n^{-1}\sum_{i=1}^n I(\langle \varepsilon^f(\boldsymbol{X}_i;\boldsymbol{\beta}), \boldsymbol{\gamma}\rangle_m \leq v) = n^{-1}\sum_{i=1}^n I\left(\sum_{k=1}^p e_{i,k}\gamma_k \leq v\right)$. Consequently, the PD independence test for the function-valued response model (2.2) is constructed similarly to the test for the scalar response model (2.1), as shown in the Supplementary Material.

When $\mathcal{L}(\varepsilon) = \varepsilon$, a conditional mean-based test for hypothesis (1.1) is

$$T_{n,\mathtt{M}}(\boldsymbol{\beta}) = \int_{\mathbb{S}^{pd}} \int_{\mathcal{R}} \left(M_{n,\mathtt{M}}(\boldsymbol{\beta};\boldsymbol{\alpha},u)\right)^2 F_{n,\langle \boldsymbol{X},\boldsymbol{\alpha}\rangle_m}(du)d\boldsymbol{\alpha},$$

with $M_{n,\mathtt{M}}(\boldsymbol{\beta};\boldsymbol{\alpha},u) = n^{-1/2}\sum_{i=1}^n \varepsilon(\boldsymbol{X}_i;\boldsymbol{\beta}) \times I\left(\langle \boldsymbol{X}_i, \boldsymbol{\alpha}\rangle_m \leq u\right)$. García-Portugués, González-Manteiga and Febrero-Bande (2014) also consider this conditional mean-based test, but without providing theoretical justifications. Note that if the null hypothesis (1.3) cannot be rejected based on $T_{n,\mathtt{F}}(\boldsymbol{\beta})$, then the null hypothesis (1.1) cannot be rejected by $T_{n,\mathtt{M}}(\boldsymbol{\beta})$. Thus, a conditional mean-based test is more powerful than a distribution-based test, which is supported by our theoretical and numerical results.

When $\mathcal{L}(\varepsilon) = \varepsilon^2$, we use $M_{n,\mathtt{V}}(\boldsymbol{\beta};\boldsymbol{\alpha},u) = n^{-1/2}\sum_{i=1}^n \left(\varepsilon^2(\boldsymbol{X}_i;\boldsymbol{\beta}) - \sigma_n^2\right) I\left(\langle \boldsymbol{X}_i, \boldsymbol{\alpha}\rangle_m \leq u\right)$, $T_{n,\mathtt{V}}$, with $\sigma_n^2 = n^{-1}\sum_{i=1}^n \varepsilon^2(\boldsymbol{X}_i;\boldsymbol{\beta})$. Next, the variance-based

test takes the form

$$T_{n,\mathtt{V}}(\boldsymbol{\beta}) = \int_{\mathbb{S}^{pd}} \int_{\mathcal{R}} (M_{n,\mathtt{V}}(\boldsymbol{\beta}; \boldsymbol{\alpha}, u))^2 F_{n,\langle \boldsymbol{X}, \boldsymbol{\alpha} \rangle_m}(du) d\boldsymbol{\alpha}.$$

Note that for $\mathcal{L}(\varepsilon) = \varepsilon^r, r \geq 2$, that is, a higher-order moment-based test, the calculation of higher-order moments of the residual is usually unstable.

The calculation of $T_{n,\mathtt{F}}$, $T_{n,\mathtt{M}}$, and $T_{n,\mathtt{V}}$ depends on the residual $\varepsilon(\boldsymbol{X}_i; \boldsymbol{\beta})$, which involves unknown coefficient functions $\boldsymbol{\beta}$. To make this feasible, we replace $\boldsymbol{\beta}$ with its estimator $\hat{\boldsymbol{\beta}}$. As stated in Section 3, under some conditions, this substitution does not affect the local power up to the order. For completeness, we briefly introduce the estimation for $\boldsymbol{\beta}$ in model (2.1).

We define the loss functions as $\ell(\varepsilon(\boldsymbol{X}; \boldsymbol{\beta}))$, where $\ell(x)$ is a nonnegative known function of x, such as the least-squares solution $\ell(x) = x^2$. We establish the covariance of $X_{i,j}(t)$ as $\mathbb{E}\left(X_{i,j}(t) X_{i,j}(s)\right) = \sum_k \theta_{j,k} \phi_k(s) \phi_k(t)$, for $j = 1, \ldots, d$, and $X_{i,j}(t)$ has the empirical expression $\hat{X}_{i,j}(t) = \sum_k \hat{\xi}_{ij,k} \hat{\phi}_k(t)$ $\hat{\xi}_{ij,k}$, with $n^{-1} \sum_{i=1}^n \hat{X}_{i,j}(t) \hat{X}_{i,j}(s) = \sum_k \hat{\theta}_{j,k} \hat{\phi}_j(s) \hat{\phi}_k(t)$, for $j = 1, \ldots, d$. We assume that $\beta_j(t) = \sum_k b_{j,k} \phi_k(t)$, $j = 1, \ldots, d$ and denote $\boldsymbol{\theta} = (a, \boldsymbol{b}_1^T, \ldots, \boldsymbol{b}_d^T)^T$, where $\boldsymbol{b}_j = \{b_{j,k}\}_{k=1}^K$, for $j = 1, \ldots, d$. Next, we estimate $\boldsymbol{\theta}$ by solving the equation $\boldsymbol{U}(\boldsymbol{\theta}) := \sum_{i=1}^n \dot{\ell}(m(Y_i, \hat{\boldsymbol{\eta}}_i)) \hat{\boldsymbol{D}}_i = 0$, with respect to $\boldsymbol{\theta}$, where $\boldsymbol{\eta}_i = (a, \boldsymbol{b}_1^T \boldsymbol{\xi}_{i1}, \ldots, \boldsymbol{b}_d^T \boldsymbol{\xi}_{id})$, $\hat{\boldsymbol{\eta}}_i = (a, \boldsymbol{b}_1^T \hat{\boldsymbol{\xi}}_{i1}, \ldots, \boldsymbol{b}_d^T \hat{\boldsymbol{\xi}}_{id})$, $\dot{\ell}(x) = d\ell(x)/dx$ is the first derivative of $\ell(x)$, $\hat{\boldsymbol{D}}_i = (m_0(Y_i, \hat{\boldsymbol{\eta}}_i), m_1(Y_i, \hat{\boldsymbol{\eta}}_i)(\hat{\boldsymbol{\xi}}_{i1})^T, \ldots, m_d(Y_i, \hat{\boldsymbol{\eta}}_i)(\hat{\boldsymbol{\xi}}_{id})^T)^T$, $m_0(Y_i, \boldsymbol{\eta}_i) = \partial m(Y_i, \boldsymbol{\eta}_i)/\partial a$, $m_j(Y_i, \boldsymbol{\eta}_i) = \partial m(Y_i, \boldsymbol{\eta}_i)/\partial(\boldsymbol{b}_j^T \boldsymbol{\xi}_{ij})$, and $\boldsymbol{\xi}_{ij} = \{\xi_{ij,k}\}_{k=1}^K$, for $j = 1, \ldots, d$, $\hat{\boldsymbol{D}}_i$. We denote the solution to $\boldsymbol{U}(\boldsymbol{\theta}) = 0$ as $\hat{\boldsymbol{\theta}}$. Next, we estimate $\beta_j(t)$ as $\hat{\beta}_j(t) = \sum_{k=1}^K \hat{b}_{j,k} \hat{\phi}_k(t)$, for $j = 1, \ldots, d$, the validity of which we justified in Section 3.

3. Theoretical Property for the PD-Test Statistic

Here, we focus on the PD distribution-based test, and leave the mean-based test to the Supplementary Material. Other specific formulae of $\mathcal{L}(\cdot)$ may be obtained similarly. Let $q = Kd + 1$. W denote $\boldsymbol{\varpi} = (\varpi_i)_{i=1}^n := (\dot{\ell}(m(Y_i, \boldsymbol{\eta}_i)))_{i=1}^n$, $\boldsymbol{V} = \mathrm{Diag}\{\varpi_1^2, \ldots, \varpi_n^2\}$, $\boldsymbol{D} = \{\boldsymbol{D}_i\}_{i=1}^n$ and $\tilde{\boldsymbol{D}} = \{(\ddot{\ell}(m(Y_i, \boldsymbol{\eta}_i)))^{1/2} \boldsymbol{D}_i\}_{i=1}^n$ as $n \times q$-dimensional matrices, $\boldsymbol{\Gamma} = \lim_{n \to \infty} \boldsymbol{D}^T \boldsymbol{V} \boldsymbol{D}/n := \{\Gamma_{k,l}\}_{1 \leq k,l \leq q} \tilde{\boldsymbol{\Gamma}}$, $\tilde{\boldsymbol{\Gamma}} = \lim_{n \to \infty} (\tilde{\boldsymbol{D}}^T \tilde{\boldsymbol{D}}/n) = (\tilde{\Gamma}_{k,l})_{1 \leq k,l \leq q}$, $\Xi = \tilde{\boldsymbol{\Gamma}}^{-1} = (\zeta_{j,k})_{1 \leq j,k \leq q}$, and $\sum_{k=1}^q \zeta_{j,k}^{(1/2)} \zeta_{k,l}^{(1/2)} = \zeta_{j,l}$. The following conditions are needed to establish the asymptotic properties for model (2.1).

(C1) $\{(Y_i, \{X_{i,j}\}_{j=1}^d)\}_{i=1}^n$ are independent and identically distributed (i.i.d.) random vectors with $0 < \mathbb{E}|Y_i| < \infty$.

(C2) ϵ_i has a zero mean and is independent of $\{X_{i,j}(\cdot)\}_{j=1}^d$, and $X_{i,j}(\cdot) \in L^2[0,1]$, for $j = 1, \ldots, d$. The covariance function of each $X_{i,j}$ is positive definite with a spectral decomposition $\Sigma_j(s,t) = \sum_{k=1}^\infty \theta_{j,k} \phi_k(s) \phi_k(t)$, where $\theta_{j,1} > \theta_{j,2} > \cdots$ and $C^{-1} k^{-\alpha} \leq \theta_{j,k} \leq C k^{-\alpha}$, and $\theta_{j,k} - \theta_{j,k+1} \geq C^{-1} k^{-\alpha-1}$,

for $\alpha > 1$, $j = 1, \ldots, d, k \geq 1$. Furthermore, the true coefficient function $\beta_j(t) = \sum_{k=1}^{\infty} b_{j,k}\phi_k(t)$, with $|b_{j,k}| \leq C_1 k^{-\kappa}$, for $k \geq 1$, $\kappa \geq \alpha + 2$.

(C3) $\mathbb{E}(\varpi_i \boldsymbol{D}_i \mid \boldsymbol{X}_i) = \mathbf{0}$. There exist constants c_2, c_l, and c_u satisfying $\mathbb{E}(\varpi_i^2) < c_2 < \infty$, and $0 < c_l \leq \inf \ddot{\ell}(m(Y_i, \boldsymbol{\eta}_i)) \leq \sup \ddot{\ell}(m(Y_i, \boldsymbol{\eta}_i)) \leq c_u < \infty$.

(C4) $m(\cdot)$ has continuous bounded first-order derivatives, $\|\boldsymbol{D}\|_\infty \leq c < \infty$, and $\tilde{\boldsymbol{\Gamma}}$ is positive definite and has a bounded maximum eigenvalue.

(C5) $Kn^{-1/(2\kappa+\alpha-1)}$ is bounded away from zero and infinity as $n \to \infty$. The following equations hold: $\sum_{k_1,k_2,k_3,k_4=1}^{q} \mathbb{E}(D_{i,k_1}D_{i,k_2}D_{i,k_3}D_{i,k_4}\zeta_{k_1,k_2}\zeta_{k_3,k_4}) = o(n/q^2)$, $\sum_{k_1,\ldots,k_8=1}^{q} \mathbb{E}(D_{i,k_1}D_{i,k_3}D_{i,k_5}D_{i,k_7})\,\mathbb{E}(D_{i,k_2}D_{i,k_4}D_{i,k_6}D_{i,k_8})\,\zeta_{k_1,k_2}\zeta_{k_3,k_4}\zeta_{k_5,k_6}\zeta_{k_7,k_8} = o(n^2q^2)$, where $q = K(d+1)$.

(C6) $\mathcal{F}(\boldsymbol{X}_i(t))$ is a measurable function of $\{X_{i,j}(t)\}_{j=1}^{d}$ satisfying $0 < \sup_{0<t<1} \mathbb{E}[\mathcal{F}(\boldsymbol{X}_i(t))] < \infty$, and the link function g_l defined in (3.1) under the alternative hypothesis has continuous bounded first-order derivatives.

(C7) $\{\varrho_i, I = 1, \ldots, n\}$ are i.i.d. with mean zero and variance one. For all i, ϱ_i is independent of $(Y_i, \boldsymbol{X}_i)$. Furthermore, $\epsilon_i\varrho_i$ and ϵ_i have the same distribution function.

Conditions (C1)–(C4) are general conditions that are readily satisfied in practice. Condition (C5) ensures that $\hat{\boldsymbol{\beta}}$ does not affect the convergence rate. In particular, under conditions (C3)–(C5), we have $n(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^\top \tilde{\boldsymbol{\Gamma}}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = O_p(q)$. Furthermore, following expression (17) and the last second expression on Page 2434 in Dou, Pollard and Zhou (2012), $q \hat{=} K(d+1) \asymp n^{1/(2\kappa+\alpha-1)}$ leads to the estimation error $\int_0^1 (\hat{\beta}_j(t) - \beta_j(t))^2 dt = O_p(n^{-(2\kappa-2)/(\alpha+2\kappa-1)})$. Following expression (4.6) in Cai and Hall (2006), $q \asymp n^{1/(2\kappa+\alpha-1)}$ leads to the prediction error $\mathbb{E}(\langle x, \hat{\beta}_j\rangle - \langle x, \beta_j\rangle)^2 = O(n^{-(2\kappa+\alpha-2)/(\alpha+2\kappa-1)})$, for any fixed function $x(t) = \sum_{k=1}^{\infty} x_k\phi_k(t)$, with $|x_k| \leq Ck^{-\alpha/2}$, for $j = 1, \ldots, d$. When the prediction error and estimation error have the above rate, substituting $\boldsymbol{\beta}_0$ with $\hat{\boldsymbol{\beta}}$ does not change the convergence property of the proposed test statistics or the order of the local power. Furthermore, with $K \asymp n^{1/(2\kappa+\alpha-1)}$, the estimation rate of $\hat{\beta}_j$ is not of the optimal rate $n^{-(2\kappa-1)/(\alpha+2\kappa)}$, which requires $K_{opt} \asymp n^{1/(\alpha+2\kappa)}$. Instead, we require $K \asymp n^{1/(2\kappa+\alpha-1)}$, which is larger than K_{opt}. We require a larger K to ensure a parametric order for the local power of the proposed statistics, which is standard in the semiparametric literature, where a smaller bandwidth or a larger number of principal components is required to reduce the bias and obtain the parametric convergence rate for the parameters. Moreover, when the proposed model (2.1) degenerates to the FLMsR, condition (C5) is identical to the conditions in Muller and Stadtmuller (2005). Condition (C6) is a general condition for the alternative hypothesis. Condition (C7) is imposed to ensure the validity of the bootstrap procedure. Specifically, if ϵ follows a symmetric distribution, then a two-point

distribution $\varrho_i = -1/1$ with probability 0.5 satisfies condition (C7). For any r, $P(\epsilon_i\varrho_i < r) = 0.5P(\epsilon_i < r) + 0.5P(\epsilon_i > -r) = p(\epsilon_i < r)$.

Remark 2. In the Supplementary Material, we list the conditions (C1f–C5f) for the function-valued outcome model. By imposing a direction $\boldsymbol{\gamma}$, we project the function $\epsilon^f(\boldsymbol{X};\boldsymbol{\beta})(t)$ along $\boldsymbol{\gamma}$, that is, $\langle\epsilon^f(\boldsymbol{X};\boldsymbol{\beta}),\boldsymbol{\gamma}\rangle_m$, which is a scalar. Because we allow the dimension p to diverge to infinity for the scalar response model, the theoretical results of the proposed statistics for the function-valued response model are similar to those for the scalar response model. Therefore, conditions (C1f)–(C5f) are similar to conditions (C1)–(C5), with the exception that the number of parameters in the function-valued response is $q_f \hat{=} K^2d + K$, owing to the approximation of $\beta_j(s,t)$, which is larger than $q \hat{=} Kd + K$ for $\beta(t)$ in the scalar response model. To ensure that the proposed statistics converge in distribution and have a parametric order of local power, we require a stronger condition, $\kappa \geq \alpha + 3$, for $\beta_j(s,t)$ in the function-valued response model, instead of $\kappa \geq \alpha + 2$ required in condition (C2).

We define $\Delta(\boldsymbol{X}_i;\boldsymbol{\alpha},u) = I\left(\langle\boldsymbol{X}_i,\boldsymbol{\alpha}\rangle_m \leq u\right) - F_{\langle\boldsymbol{X},\boldsymbol{\alpha}\rangle_m}(u)$ $\Delta(\cdot;\cdot)$ and $\boldsymbol{\alpha}_{r,\bullet} = \{\boldsymbol{\alpha}_{r,j}\}_{j=1}^d$, with $\boldsymbol{\alpha}_{r,j} \in \mathcal{R}^p$. Next, we establish the asymptotic distribution of the proposed test statistic if the true parameters are known for model (2.1).

Theorem 1. *Under conditions* (C1) *and* (C2) *and the null hypothesis* (1.3), *if* $p = o(n)$, *for any* $m \in \mathcal{R}$, *Prob*$(T_{n,\mathtt{F}}(\boldsymbol{\beta}_0) < m) -$ *Prob*$(T^0_{\infty,F} < m) \to 0$, *where* $T^0_{\infty,F} := \int_{\mathbb{S}^{pd}}\iint_{\mathcal{R}^2}(M^0_{\infty,F}(\boldsymbol{\alpha},u,v))^2F_\varepsilon(dv)F_{\langle\boldsymbol{X},\boldsymbol{\alpha}\rangle_m}(du)d\boldsymbol{\alpha}$, *and* $M^0_{\infty,F}(\cdot,\cdot,\cdot)$ *is a Gaussian process with zero mean and covariance function* $K((\boldsymbol{\alpha}_{1,\bullet},u_1,v_1),(\boldsymbol{\alpha}_{2,\bullet},u_2,v_2)) = \{\mathbb{E}\left[I\left(\epsilon_i \leq v_1\right)I\left(\epsilon_i \leq v_2\right)\right] - F_\epsilon(v_1)F_\epsilon(v_2)\}$ $\mathbb{E}(\Delta(\boldsymbol{X}_i;\boldsymbol{\alpha}_{1,\bullet},u_1)$ $\Delta(\boldsymbol{X}_i;\boldsymbol{\alpha}_{2,\bullet},u_2))$.

Corollary S.2 in the Supplementary Material explores the asymptotic distribution of the proposed mean test $T_{n,\mathtt{M}}(\boldsymbol{\beta}_0)$. Directly comparing $T^0_{\infty,F}$ and $T^0_{\infty,M}$ is not possible because the former depends on the distribution of ϵ, whereas the latter depends on its moment. However, under certain situations, a strict inequality holds between the two statistics. Specifically, the limiting distribution for the mean-based statistic is described as follows: for any $m \in \mathcal{R}$, Prob$(T_{n,\mathtt{M}}(\boldsymbol{\beta}_0)$ $< m) -$ Prob$(T^0_{\infty,M} < m) \to 0$, where $T^0_{\infty,M} := \int_{\mathbb{S}^{pd}}\int_{\mathcal{R}}(M^0_{\infty,M}(\boldsymbol{\alpha},u))^2F_{\langle\boldsymbol{X},\boldsymbol{\alpha}\rangle_m}$ $(du)d\boldsymbol{\alpha}$, and $\mathbb{E}[(M^0_{\infty,M}(\boldsymbol{\alpha},u))^2] = \mathbb{E}(\epsilon_i^2)F_{\langle\boldsymbol{X},\boldsymbol{\alpha}\rangle_m}(u)$. From Theorem 1, we have $\mathbb{E}[\int_{\mathcal{R}}(M^0_{\infty,F}(\boldsymbol{\alpha},u,v))^2F_\epsilon(dv)] = \int_{\mathcal{R}}\mathbb{E}[(M^0_{\infty,F}(\boldsymbol{\alpha},u,v))^2]F_\epsilon(dv) = \int_{\mathcal{R}}F_\varepsilon(v)(1 - F_\epsilon(v))F_\epsilon(dv)F_{\langle\boldsymbol{X},\boldsymbol{\alpha}\rangle_m}(u)(1 - F_{\langle\boldsymbol{X},\boldsymbol{\alpha}\rangle_m}(u)) \leq 0.25F_{\langle\boldsymbol{X},\boldsymbol{\alpha}\rangle_m}(u)$. When $\mathbb{E}(\epsilon_i^2) \geq 0.25$, it is easy to determine that $\mathbb{E}(T^0_{\infty,F}) \leq \mathbb{E}(T^0_{\infty,M})$, which indicates that the distribution-based statistic tends to generate smaller values and, as a result, is less powerful than the mean-based statistic. Next, we establish the asymptotic distribution of test statistics with estimated parameters.

Theorem 2. *Under conditions* (C1)–(C5) *and under the null hypothesis* (1.3), *for any* $m \in \mathcal{R}$, *Prob*$(T_{n,\mathtt{F}}(\hat{\boldsymbol{\beta}}) < m) -$ *Prob*$(T^1_{\infty,F} < m) \to 0$, *where* $T^1_{\infty,F} :=$

$\int_{\mathbb{S}^{pd}} \iint_{\mathcal{R}^2} (M^1_{\infty,F}(\boldsymbol{\alpha}, u, v))^2 \times F_\epsilon(dv) F_{\langle \boldsymbol{X}, \boldsymbol{\alpha} \rangle_m}(du) d\boldsymbol{\alpha}$, *and* $M^1_{\infty,F} \equiv M^0_{\infty,F} + M^e_{\infty,F}$, $M^e_{\infty,F}(\cdot,\cdot,\cdot)$ *is a Gaussian process with mean* 0 *and covariance function* $K_1((\boldsymbol{\alpha}_{1,\bullet}, u_1, v_1), (\boldsymbol{\alpha}_{2,\bullet}, u_2, v_2)) = \sigma^2_F(\boldsymbol{\alpha}_{1,\bullet}, u_1, \boldsymbol{\alpha}_{2,\bullet}, u_2) \times f_\epsilon(v_1) f_\epsilon(v_2)$, *and* $cov(M^0_{\infty,F}(\boldsymbol{\alpha}_{1,\bullet}, u_1, v_1), M^e_{\infty,F}(\boldsymbol{\alpha}_{2,\bullet}, u_2, v_2)) = \sigma_{c,F}(\boldsymbol{\alpha}_{1,\bullet}, u_1, v_1, \boldsymbol{\alpha}_{2,\bullet}, u_2) f_\epsilon(v_2)$, *where* $\sigma^2_F(\boldsymbol{\alpha}_{1,\bullet}, u_1, \boldsymbol{\alpha}_{2,\bullet}, u_2)$ *and* $\sigma_{c,F}(\boldsymbol{\alpha}_{1,\bullet}, u_1, v_1, \boldsymbol{\alpha}_{2,\bullet}, u_2)$ *are defined in the Supplementary Material.*

Corollary S.3 in the Supplementary Material presents the asymptotic distributions of $T_{n,\mathtt{M}}(\hat{\boldsymbol{\beta}})$. Comparing Corollary S.3 with Theorem 2, under certain situations, $T_{n,\mathtt{F}}(\hat{\boldsymbol{\beta}})$ has smaller asymptotic mean values, and hence is less powerful than $T_{n,\mathtt{M}}(\hat{\boldsymbol{\beta}})$. In addition, although substituting $\boldsymbol{\beta}_0$ with $\hat{\boldsymbol{\beta}}$ does not change the convergence property of the proposed test statistics, it increases the variance of the test statistics because of the additional terms $M^e_{\infty,F}$ and $M^e_{\infty,m}$, as stated in Theorem 2 and Corollary S.3.

Now, we analyze the asymptotic distribution of $T_{n,\mathtt{F}}$ using a sequence of local alternatives converging to null at a parametric rate $n^{-1/2}$. In particular, we consider the local alternative

$$H_{A,n} : Y_{i,n} = g_l\left(a, \langle X_{i,1}, \beta_1 \rangle, \ldots, \langle X_{i,d}, \beta_d \rangle, \epsilon_i, n^{-1/2} \mathcal{F}(\boldsymbol{X}_i)\right), \tag{3.1}$$

where $g_l\left(a, \langle X_{i,1}, \beta_1 \rangle, \ldots, \langle X_{i,d}, \beta_d \rangle, \epsilon_i, 0\right) := g\left(a, \langle X_{i,1}, \beta_1 \rangle, \ldots, \langle X_{i,d}, \beta_d \rangle, \epsilon_i\right)$, and $\mathcal{F}(\boldsymbol{X}_i)$ is a measurable function of $\{X_{i,j}(t)\}_{j=1}^d$.

Theorem 3. *Under conditions of* (C1)–(C6) *and local alternative* (3.1)*, for any* $m \in \mathcal{R}$, $Prob(T_{n,\mathtt{F}}(\check{\boldsymbol{\beta}}) < m) - Prob(T^a_{\infty,F} < m) \to 0$, *where* $T^a_{\infty,F} := \int_{\mathbb{S}^{pd}} \iint_{\mathcal{R}^2} (M^a_{\infty,F}(\boldsymbol{\alpha}, u, v))^2 \times F_\epsilon(dv) F_{\langle \boldsymbol{X}, \boldsymbol{\alpha} \rangle_m}(du) d\boldsymbol{\alpha}$, *and* $\check{\boldsymbol{\beta}}$ *is the estimate obtained from model* (2.1) *using data* $\{Y_{i,n}, \boldsymbol{X}_i\}_{i=1}^n$, $M^a_{\infty,F}(\boldsymbol{\alpha}, u, v) \equiv M^1_{\infty,F}(\boldsymbol{\alpha}, u, v) - D^a_F(\boldsymbol{\alpha}, u, v)$, *and*

$$\begin{aligned} &D^a_F(\boldsymbol{\alpha}, u, v) \\ &= \mathbb{E}\left\{ \Delta(\boldsymbol{X}_i; \boldsymbol{\alpha}, u) \left(m_y(Y_i, \boldsymbol{\eta}_i) \dot{g}_l(\boldsymbol{\eta}_i, \epsilon_i, 0) \mathcal{F}(\boldsymbol{X}_i) - \sum_{j=1}^q D_{i,j} \upsilon_j \right) \right\} f_\varepsilon(v), \end{aligned}$$

where $m_y(\cdot,\cdot), \dot{g}_l, \upsilon_j$ *are defined in the Supplementary Material.*

Theorem 3 implies that the proposed test achieves the parametric order $O(n^{-1/2})$ for the local power. This order is not attainable for tests based on the local approach, such as that of Patilea, Sánchez-Sellero and Saumard (2016), which is based on the conditional mean, and hence has order $O((nh^{1/2})^{-1/2})$ where h is the bandwidth.

Remark 3. The fast parametric order $O(n^{-1/2})$ for the local power in Theorem 3 is attributed to two aspects. First, most mean-based test methods, such as that of Patilea, Sánchez-Sellero and Saumard (2016), require calculating a conditional expectation, leading to the order $O((nh^{1/2})^{-1/2})$, because only local

data are involved. Instead, because $\mathbb{E}(U \mid \langle X, \boldsymbol{\alpha}\rangle_m) = \mathbb{E}(U)$ holds if and only if $\mathbb{E}(UI(\langle X, \boldsymbol{\alpha}\rangle_m \le u)) = \mathbb{E}(U)F_{\langle X,\boldsymbol{\alpha}\rangle_m}(u)$ holds, for any u and $\forall \boldsymbol{\alpha} \in \mathbb{S}^p$, our constructed PD test induces the indicator function $I(\langle X, \boldsymbol{\alpha}\rangle_m \le u)$ as a weight function. Consequently, the proposed PD test calculates the unconditional expectation $\mathbb{E}\big(UI(\langle X, \boldsymbol{\alpha}\rangle_m \le u)\big)$, which is estimated based on nonlocal data. Second, we integrate over all $\boldsymbol{\alpha}$ to avoid any subjective choice on $\boldsymbol{\alpha}$. This integration could improve the order to $O(n^{-1/2})$, even if an integrand is estimated at a nonparametric rate. In particular, we use a larger $K \asymp n^{1/(\alpha+2\kappa-1)}$ than the optimal $K_{opt} = O(n^{1/(\alpha+2\kappa)})$ to control the bias of $T_{n,F}(\hat{\boldsymbol{\beta}})$, and reduce its variance using integration. As a result, the parametric order of the local power of $T_{n,F}(\hat{\boldsymbol{\beta}})$ is maintained. Similar conclusions are established in the literature. For example, the convergence rate of the integration in Cai and Hall (2006) is $\mathbb{E}(\langle x, \hat{\beta}\rangle - \langle x, \beta_0\rangle)^2 = O(n^{-(2\kappa+\alpha-2)/(\alpha+2\kappa-1)})$, which is faster than the estimation rate $\int_0^1(\hat{\beta}(t) - \beta(t))^2dt = O_p(n^{-(2\kappa-2)/(\alpha+2\kappa-1)})$ established in Dou, Pollard and Zhou (2012) under the condition that $\kappa \ge \alpha + 2$, $\alpha > 1$, and $K \asymp n^{1/(\alpha+2\kappa-1)}$. Notably, the necessity of undersmoothing the nonparametric function to obtain a root-n-consistent estimation for the parameters using integration is standard in nonparametric regression; see, Carroll et al. (1997) and Hastie (2017).

Corollary S.4 in the Supplementary Material gives the asymptotic distributions of $T_{n,\mathtt{M}}(\check{\boldsymbol{\beta}})$, from which we determine that $(D_F^a(\boldsymbol{\alpha}, u, v))^2 \le (D_M^a(\boldsymbol{\alpha}, u))^2 f_\epsilon^2(v)$. According to Theorem 3, we have $\int_{\mathcal{R}}(D_F^a(\boldsymbol{\alpha}, u, v))^2F_\epsilon(dv) \le (D_M^a(\boldsymbol{\alpha}, u))^2 \int_{\mathcal{R}} f_\epsilon^2(v)F_\epsilon(dv)$. When the term $\int_{\mathcal{R}} f_\epsilon^2(v)F_\epsilon(dv) \le 1$, $T_{\infty,F}^a$ has a smaller asymptotic mean than $T_{\infty,M}^a$; that is, the mean test is more powerful than the distribution test.

The following local alternative hypothetical models are considered:

$$H_{A,n} : Y_{i,n} = g_l\left(a, \langle X_{i,1}, \beta_1\rangle, \ldots, \langle X_{i,d}, \beta_d\rangle, \epsilon_i, n^\nu \mathcal{F}(\boldsymbol{X}_i)\right). \tag{3.2}$$

Corollary 1. *Under the conditions of Theorem* 3 *and the alternatives* (3.2) *with* $-1/2 < \nu \le 0$*, we have Prob*$(T_{n,\mathtt{F}} > \eta) \to 1$*, as* $n \to \infty$*, for any* $\eta > 0$*.*

Corollary 1 shows that the statistics $T_{n,\mathtt{F}}$ diverge to infinity under the local alternatives (3.2) for $-1/2 < \nu < 0$ and the global alternative hypothesis for $\nu = 0$, indicating that the statistics have asymptotic power one. Theorem 3 and Corollary 1 imply that our proposed tests $T_{n,\mathtt{F}}$ can detect the alternative models converging to the null model with rate n^ν, for $-1/2 \le \nu \le 0$.

4. Implementation

In this section, we describe how to calculate the test statistics. We use $A_{ijl} = \int_{\mathbb{S}^{pd}} I(\langle \boldsymbol{X}_i, \boldsymbol{\alpha}\rangle_m \le \langle \boldsymbol{X}_l, \boldsymbol{\alpha}\rangle_m) I\left(\langle \boldsymbol{X}_j, \boldsymbol{\alpha}\rangle_m \le \langle \boldsymbol{X}_l, \boldsymbol{\alpha}\rangle_m\right) d\boldsymbol{\alpha}$, and $C_{ijl}(\boldsymbol{\beta}) = I(\varepsilon(\boldsymbol{X}_i; \boldsymbol{\beta}) \le \varepsilon(\boldsymbol{X}_l; \boldsymbol{\beta}))I(\varepsilon(\boldsymbol{X}_j; \boldsymbol{\beta}) \le \varepsilon(\boldsymbol{X}_l; \boldsymbol{\beta}))$. After calculations, we obtain that $T_{n,\mathtt{F}}(\boldsymbol{\beta}) = n^{-3}\sum_{i,j,k,l} A_{ijk}C_{ijl}(\boldsymbol{\beta}) - 2n^{-4}\sum_{i,j,k,l}\sum_s A_{ijk}C_{isl}(\boldsymbol{\beta}) + n^{-5}$

$\sum_{i,j,k,l}\sum_{s_1,s_2} A_{ijk}C_{s_1s_2l}(\boldsymbol{\beta})$, and $T_{n,\mathtt{M}}(\boldsymbol{\beta}) = n^{-2}\sum_{i,j,l}\varepsilon(\boldsymbol{X}_i;\boldsymbol{\beta})\varepsilon(\boldsymbol{X}_j;\boldsymbol{\beta})A_{ijl}$. Note that A_{ijk} involves pd-dimensional integrals, which require intensive computation. Calculating A_{ijl} follows a volume calculation in a pd-ball. Following simple algebra (Theorem 4.41 on Page 183 of Folland (2002)), the integral A_{ijl} is proportional to the volume of a spherical wedge and

$$\begin{aligned}
A_{ijl} &:= \int_{\mathbb{S}^{pd}} I\big(\langle \boldsymbol{X}_i,\boldsymbol{\alpha}\rangle_m \le \langle \boldsymbol{X}_l,\boldsymbol{\alpha}\rangle_m\big) I\big(\langle \boldsymbol{X}_j,\boldsymbol{\alpha}\rangle_m \le \langle \boldsymbol{X}_l,\boldsymbol{\alpha}\rangle_m\big) d\boldsymbol{\alpha} \\
&= \int_{\mathbb{S}^{pd}} I\left(\sum_{q=1}^{d}\langle X_{i,q},\boldsymbol{\alpha}_q\rangle_m \le \sum_{q=1}^{d}\langle X_{l,q},\boldsymbol{\alpha}_q\rangle_m\right) \\
&\quad I\left(\sum_{q=1}^{d}\langle \boldsymbol{X}_{j,q},\boldsymbol{\alpha}_q\rangle_m \le I(\sum_{q=1}^{d}\langle X_{l,q},\boldsymbol{\alpha}_q\rangle_m\right) d\boldsymbol{\alpha} \\
&= \int_{\mathbb{S}^{pd}} I\left(\sum_{q=1}^{d}\sum_{k=1}^{p}\xi_{iq,k}\alpha_{q,k} \le \sum_{q=1}^{d}\sum_{k=1}^{p}\xi_{lq,k}\alpha_{q,k}\right) \\
&\quad I\left(\sum_{q=1}^{d}\sum_{k=1}^{p}\xi_{jq,k}\alpha_{q,k} \le \sum_{q=1}^{d}\sum_{k=1}^{p}\xi_{lq,k}\alpha_{q,k}\right) d\boldsymbol{\alpha} \\
&= \frac{A_{ijl}^{(0)}\pi^{pd/2-1}}{\Gamma(pd/2+1)},
\end{aligned}$$

where $A_{ijl}^{(0)}$ is the complementary angle between the vectors $(\boldsymbol{\xi}_i-\boldsymbol{\xi}_l)$ and $(\boldsymbol{\xi}_j-\boldsymbol{\xi}_k)$, with $\boldsymbol{\xi}_i = (\boldsymbol{\xi}_{i,1}^\top,\ldots,\boldsymbol{\xi}_{i,d}^\top)^\top$, $A_{ijl}^{(0)} = \big|\pi - \arccos\{\sum_{k=1}^{d}\langle X_{i,k}-X_{l,k}, X_{j,k}-X_{l,k}\rangle_p / (\sqrt{\sum_{k=1}^{d}\|X_{i,k}-X_{l,k}\|^2}\sqrt{\sum_{k=1}^{d}\|X_{j,k}-X_{l,k}\|^2})\}\big|$, $\Gamma(\cdot)$ is the gamma function, $\|X_{i,k}\| = \sqrt{\sum_{s=1}^{p}\xi_{ik,s}^2}$, and $\langle X_{i,k}, X_{j,k}\rangle_p = \sum_{s=1}^{p}\xi_{ik,s}\xi_{jk,s}$. Hence, the computation of these integrals is simple, regardless of the dimension p.

Because there is no explicit asymptotic null distribution for $T_{n,\mathtt{F}}(\boldsymbol{\beta})$, we implement the test using a bootstrap procedure. We approximate the asymptotic null distribution of $M_{n,\mathtt{F}}$ by that of $M_{n,F}^* = n^{-1/2}\sum_{i=1}^{n}[I(\varepsilon^*(\boldsymbol{X}_i;\hat{\boldsymbol{\beta}}^*) \le v) - F_{n,\varepsilon}^*(v)] \times I(\langle \boldsymbol{X}_i,\boldsymbol{\alpha}\rangle_m \le u)$, for $v,u \in \mathcal{R}, \boldsymbol{\alpha}_j \in \mathbb{S}^p, j = 1,\ldots,d$. Here, $F_{n,\varepsilon}^*(v) = n^{-1}\sum_{i=1}^{n} I(\varepsilon^*(\boldsymbol{X}_i;\hat{\boldsymbol{\beta}}^*) \le v)$, and the sequence $\{\varepsilon^*(\boldsymbol{X}_i;\hat{\boldsymbol{\beta}}^*)\}_{i=1}^n$ includes the residuals computed from $\varepsilon^*(\boldsymbol{X}_i;\hat{\boldsymbol{\beta}}^*) = m(Y_i^*,\hat{a}^*,\langle X_{i,1},\hat{\beta}_1^*\rangle,\ldots,\langle X_{i,d},\hat{\beta}_d^*\rangle)$, where $Y_i^* = g(Y_i,\hat{a},\langle X_{i,1},\hat{\beta}_1\rangle,\ldots,\langle X_{i,d},\hat{\beta}_d\rangle,\varepsilon(\boldsymbol{X}_i;\hat{\boldsymbol{\beta}})\varrho_i)$, $\{\hat{a}^*,\{\hat{\beta}_j^*\}_{j=1}^d\}$ is the bootstrap estimator calculated from the data $\{(Y_i^*,\boldsymbol{X}_i)\}_{i=1}^n$, and $\{\varrho_i\}_{i=1}^n$ satisfies condition (C7). For example, ϱ_i uses values of -1 or 1 with a probability of 0.5, for $i = 1,\ldots,n$.

For the bootstrap test statistic $T_{n,F}^*$, we have the following result.

Theorem 4. *Under the null hypothesis* (2.1) *or the alternative hypothesis* (3.2) *with* $\nu \le 0$, *if conditions* (C1)–(C7) *are satisfied, then the conditional distribution of* $T_{n,F}^*$ *converges in distribution to the limiting null distribution of* $T_{n,\mathtt{F}}$, *giving*

$\{Y_i, \boldsymbol{X}_i\}_{i=1}^n$.

Theorem 4 shows that the bootstrap distribution of the test statistic is equivalent to the asymptotic distribution of the proposed test. The critical value determined using this method approximates the theoretical value, regardless of whether the data are derived from the null hypothetical model (2.1) or the alternative hypothetical model (3.2). Corollary S.5 in the Supplementary Material shows similar results for the bootstrap mean test statistic $T^*_{n,M}$. The proposed bootstrap procedure also works for the variance-based test statistic and other higher-order moment test statistics under condition (C7), using derivations similar to those in Corollary S.5. However, note that the finite-sample performance is poor due to the instability caused by estimating the variance or other higher-order moments. Thus, we suggest using the proposed distribution-based test and mean-based test in practice, rather than using the higher-order moment test.

The entire procedure involves two tuning parameters, p and K, which denote the dimension of the projection parameter $\boldsymbol{\alpha}$, and the number of principal component functions, respectively. Because we use both projection parameters and principal components to capture information from the covariates $\boldsymbol{X}$, we set $K = p$, for simplicity. Larger p and K indicate that more information is captured from the covariates $\boldsymbol{X}$, but with a larger variance and heavier computational burden. We choose K to be the number of principal components such that at least 95% of the variability of $\boldsymbol{X}$ is captured, which performs well in our numerical studies.

5. Numerical Studies

In this section, we compare the performance of the proposed PD test statistics, namely, the distribution-based statistic $T^f_{n,\mathtt{F}}$ and the mean-based statistic $T^f_{n,\mathtt{M}}$, to that of state-of-the-art tests, including the FMDD of Lee, Zhang and Shao (2020) and the fdapss proposed by Patilea, Sánchez-Sellero and Saumard (2016), both of which are based on the conditional mean.

Example 1 (FLMfR). We consider a model in which the response, Y_i, is a functional response, and the predictor, $X_i(t)$, is a univariate functional predictor. The functional linear model is expressed as follows:

$$Y_i(t) = \int_0^1 c_1 \cdot \beta(s,t) X_i(s) ds + \int_0^1 c_2 \cdot \beta(s,t) X_i^2(s) ds + \{X_i(t)\}^{c_3} \epsilon_i(t), \quad 1 \le i \le n, \tag{5.1}$$

where $\{X_i(t)\}_{i=1}^n$ are generated independently from Brownian bridges, $\{\epsilon_i(t)\}_{i=1}^n$ follows $N(0, 0.1^2)$, $\beta(s,t) = \exp(s^2+t^2)/2$, and $c_1 = 0.25$. Setting a homogeneous scenario ($c_3 = 0$), we consider $c_2 = 0$ for the null hypothesis, and $c_2 = 0.05, n^{-1/2}, n^{-2/5}$, and 1 for the alternatives.

Table 1 shows the empirical sizes and power of our proposed test compared with those of the FMDD and fdapss based on 500 repetitions. For each repetition, we use 500 bootstrap samples of the original sample to compute the critical value. Because the FMDD and fdapss are both based on the true value of the coefficient function $\beta(\cdot)$, for comparison, we demonstrate our proposed test using the true value $\beta_0(s,t)$ and the estimated value $\hat{\beta}(s,t)$. The number of components K for each sample is chosen so that the percentage of explained variance is larger than 95%, and $p = K$.

The upper block of Table 1 presents the percentages of rejections for nominal levels at 10% and 5% when the sample size is $n = 40, 100$, which suggests that the empirical size of $T^f_{n,\mathtt{F}}$ is slightly larger than the nominal level, because $\beta_0(s,t)$ is replaced with its estimator. This result may be attributed to the small sample size. In addition, our proposed distribution-based statistic $T^f_{n,\mathtt{F}}$ and fdapss are slightly conservative for small samples under the null hypothesis, whereas $T^f_{n,\mathtt{M}}$ and FMDD are the opposite. For the power under the alternative with $c_3 = 0$, $T^f_{n,\mathtt{M}}$ is more powerful than $T^f_{n,\mathtt{F}}$, which outperforms fdapss when the null hypothesis (1.1) does not hold. These findings are consistent with the conclusion stated in Theorem 3, and occur because the mean-based test is able to detect the relationship between the functional covariate and the response when $c_3 = 0$. In addition, when the alternative part becomes more significant as c_2 increases, $T^f_{n,\mathtt{F}}$ performs much better than FMDD in terms of test power.

Furthermore, to consider the effect of the heterogeneous variance, we generate data with $c_2 = 0, c_3 = 0$ for the null, and $c_2 = 0, c_3 = 2$ for the alternative. The bottom-right block of Table 1 presents the results of the test statistics $T^f_{n,\mathtt{F}}$ and $T^f_{n,\mathtt{M}}$ under heterogeneity compared with FMDD and fdapss. The results show that under model (5.1), only the distribution-based test $T^f_{n,\mathtt{F}}$ detects the heterogeneity from the variance; the mean-based tests $T^f_{n,\mathtt{M}}$, FMDD, and fdapss, fail to achieve this detection.

We also conducted simulations on the same functional response setting, but with $\{\epsilon_i(t)\}_{i=1}^n$ following a non-Gaussian distribution, such as Pareto noise with a finite second moment and Brownian bridges. Results similar to those in Table 1 are obtained, and are relegated to the Supplementary Material. We also list the computation times of the different methods in Table 2 of the Supplementary Material, which shows that the mean-based tests are much faster than the distribution-based test, and for the mean-based tests, the proposed test $T^f_{n,\mathtt{M}}$ is faster than FMDD, but slower than fdapss.

We also conduct a simulation for the scalar response in the Supplementary Material, Example 1.2, with conclusions similar to those for the obtained functional response.

Table 1. Simulation results for Example 1 based on the proposed test, FMDD, and fdapss under model (5.1). The rows of β_0 and $\hat{\beta}$ show the results based on using the true $\beta_0(s,t)$ and the estimated value $\hat{\beta}(s,t)$, respectively.

		Level=10%		Level=5%		Level=10%		Level=5%	
test	β	$n=40$	$n=100$	$n=40$	$n=100$	$n=40$	$n=100$	$n=40$	$n=100$
		$c_2=0, c_3=0$				$c_2=n^{-2/5}, c_3=0$			
$T^f_{n,F}$	β_0	0.076	0.083	0.022	0.030	0.808	1.000	0.542	1.000
	$\hat{\beta}$	0.106	0.086	0.060	0.049	0.992	1.000	0.956	1.000
$T^f_{n,M}$	β_0	0.104	0.105	0.052	0.058	1.000	1.000	1.000	1.000
	$\hat{\beta}$	0.106	0.094	0.060	0.046	1.000	1.000	1.000	1.000
FMDD	β_0	0.118	0.138	0.052	0.066	0.832	1.000	0.530	0.996
fdapss	β_0	0.080	0.100	0.028	0.049	0.502	0.969	0.366	0.952
		$c_2=0.05, c_3=0$				$c_2=1, c_3=0$			
$T^f_{n,F}$	β_0	0.208	0.582	0.122	0.384	0.994	1.000	0.906	1.000
	$\hat{\beta}$	0.418	0.510	0.274	0.294	0.998	1.000	0.988	1.000
$T^f_{n,M}$	β_0	0.690	0.986	0.562	0.976	1.000	1.000	1.000	1.000
	$\hat{\beta}$	0.754	0.998	0.624	0.990	1.000	1.000	1.000	1.000
FMDD	β_0	0.148	0.118	0.072	0.052	0.888	1.000	0.554	1.000
fdapss	β_0	0.065	0.132	0.022	0.079	0.823	1.000	0.691	1.000
		$c_2=n^{-1/2}, c_3=0$				$c_2=0, c_3=2$			
$T^f_{n,F}$	β_0	0.638	0.970	0.386	0.920	0.988	1.000	0.948	1.000
	$\hat{\beta}$	0.956	0.984	0.898	0.938	0.970	1.000	0.920	1.000
$T^f_{n,M}$	β_0	1.000	1.000	1.000	1.000	0.028	0.044	0.008	0.014
	$\hat{\beta}$	1.000	1.000	1.000	1.000	0.000	0.000	0.000	0.000
FMDD	β_0	0.748	0.986	0.486	0.960	0.298	0.969	0.064	0.952
fdapss	β_0	0.212	0.511	0.134	0.450	0.080	0.069	0.000	0.029

Example 2. (MiFRM). In this simulation example, we consider a type of MiFRM model for $t \in [0,1]$, $Y_i(t) = g(c_1\beta(t)X_i(t)) + c_2 g(\beta(t)X_i^2(t)) + X_i^{c_3}(t)\epsilon_i(t)$, where $X_i(t)$ and $\epsilon_i(t)$ are generated as shown in Example 1, $\beta(t) = \exp(-4(t-0.3)^2)$, $g(t) = \exp(t)/(1+\exp(t))$, and $c_1 = 0.25$, $(c_2, c_3) = (0,0)$ for the null, and $(c_2, c_3) = (0,2), (1,0), (1,2)$ for the three alternative model scenarios. The results with $\beta(t)$ given and based on 500 simulations are presented in Table 2 for the test statistics $T^f_{n,\mathtt{F}}$, $T^f_{n,\mathtt{M}}$, and fdapss. Table 2 shows that the empirical sizes of $T^f_{n,\mathtt{F}}$ and $T^f_{n,\mathtt{M}}$ are closest to the nominal size, which is less true for fdapss. The power of $T^f_{n,\mathtt{M}}$ and FMDD is almost zero for $c_2 = 0$ and $c_3 = 2$, because the link function for the mean part takes a logistic form, the variation of which is weak. This problem is alleviated to some extent by fdapss by the standardization process. As long as the null hypothesis of the conditional mean does not hold, $T^f_{n,\mathtt{M}}$ performs best. When the null hypothesis of the conditional mean cannot be rejected, but the model contains heterogeneous variance (corresponding to $c_2 = 0$ and $c_3 = 2$), $T^f_{n,\mathtt{F}}$ and FMDD can detect the heterogeneity.

Table 2. Simulation results for Example 2 based on the proposed test, fdapss, and FMDD. The caption is the same as that of Table 1.

	Level=10%		Level=5%		Level=10%		Level=5%	
test	$n=40$	$n=100$	$n=40$	$n=100$	$n=40$	$n=100$	$n=40$	$n=100$
	$c_2=0,\ c_3=0$				$c_2=0,\ c_3=2$			
$T^f_{n,F}$	0.088	0.090	0.027	0.040	0.234	0.240	0.132	0.180
$T^f_{n,M}$	0.098	0.096	0.062	0.042	0.000	0.000	0.000	0.000
FMDD	0.131	0.138	0.076	0.066	0.006	0.002	0.004	0.002
fdapss	0.065	0.086	0.022	0.030	0.021	0.095	0.011	0.040
	$c_2=1,\ c_3=0$				$c_2=1,\ c_3=2$			
$T^f_{n,F}$	0.495	0.980	0.127	0.900	0.639	1.000	0.408	0.980
$T^f_{n,M}$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
FMDD	0.992	1.000	0.986	1.000	0.994	1.000	0.988	1.000
fdapss	0.541	1.000	0.405	1.000	0.843	1.000	0.746	1.000

6. Real-data examples

In this section, we apply the proposed PD test to check the goodness of fit of several models for two data sets: Canadian weather data, and Chinese air pollution data.

6.1. Analysis of Canadian weather data

The Canadian weather data are obtained from the R package **fda**. The data consist of the daily mean temperature and rainfall registered at 35 weather stations in Canada from 1960 to 1994. For detailed explanations of the data, refer to Ramsay and Silverman (2002). Specifically, in this data set, the stations are classified into four climatic zones, namely, Atlantic, Pacific, Continental, and Arctic, leading to functional ANOVA models. The aim of this analysis is to assess the validity of six models: FLMcR, FLMfR, FLMcR coupled with ANOVA (FLMcR + ANOVA), FLMfR coupled with ANOVA (FLMfR + ANOVA), FLMcR with heterogeneous variance (FLMcRw), and FLMfR with heterogeneous variance (FLMfRw). The first four types of models are also analyzed in Patilea, Sánchez-Sellero and Saumard (2016).

Table 3 contains the p-values for testing the goodness of fit of the models based on the proposed tests $T^f_{n,\mathtt{F}}$ and $T^f_{n,\mathtt{M}}$ and the conditional mean tests fdapss and FMDD, where the response $Y_{ij}(t)$ and the covariate $X_{ij}(t)$ represent the logarithm of the rainfall and temperature, respectively, at station i of climate zone j on day t. The results are based on 500 bootstrap replicates, and both the response $Y_{ij}(t)$ and the covariates $X_{ij}(t)$ are centralized so that no models include the intercept term. From Table 3, with the first four types of models, we draw the same conclusions presented in Patilea, Sánchez-Sellero and Saumard (2016). That is, there exists a varying correlation between rainfall and temperature, with

Table 3. Canadian weather data: the p-values for testing the goodness of fit of various models, and the results are based on using the estimated coefficient value.

Name of the model	Formula	p-value			
		$T^f_{n,\mathtt{M}}$	$T^f_{n,\mathtt{F}}$	fdapss	FMDD
No-effect	$Y_{ij}(t) = \epsilon_{ij}(t)$	0.009	0.000	0.000	0.000
Functional ANOVA	$Y_{ij}(t) = \alpha_j(t) + \epsilon_{ij}(t)$	0.307	0.173	0.226	0.000
No-effect+heterogeneity	$Y_{ij}(t) = X_{ij}(t)\epsilon_{ij}(t)$	0.455	0.869	0.407	0.409
FLMcR	$Y_{ij}(t) = X_{ij}(t)\beta(t) + \epsilon_{ij}(t)$	0.046	0.023	0.000	0.000
FLMcR + ANOVA	$Y_{ij}(t) = \alpha_j(t) + X_{ij}(t)\beta(t) + \epsilon_{ij}(t)$	0.174	0.367	0.323	0.222
FLMcRw	$Y_{ij}(t) = X_{ij}(t)\beta(t) + X_{ij}(t)\epsilon_{ij}(t)$	1.000	0.980	0.401	0.515
FLMfR	$Y_{ij}(t) = \int_0^1 \beta_1(s,t)X_{ij}(s)ds + \epsilon_{ij}(t)$	0.000	0.000	0.000	0.000
FLMfR+ANOVA	$Y_{ij}(t) = \alpha_j(t) + \int_0^1 \beta_1(s,t)X_{ij}(s)ds + \epsilon_{ij}(t)$	0.782	0.713	0.170	0.695
FLMfRw	$Y_{ij}(t) = \int_0^1 \beta_1(s,t)X_{ij}(s)ds + X_{ij}(t)\epsilon_{ij}(t)$	1.000	0.771	0.401	0.443

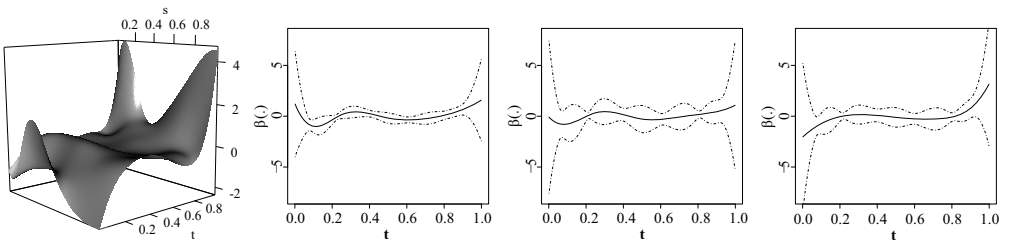


Figure 1. The estimated $\beta(s,t)$ in FLMfRw for Canada weather data: the estimated surface (left), and the functions with the second coordinate fixed at $t = 0.25, 0.5$, and 0.75, respectively.

the correlation varying across climatic zones. Compared with the conventional ANOVA models, one extra finding is that heterogeneous variance models in which the heterogeneity depends on temperature also work well. For model FLMfRw, Figure 1 shows the estimate of $\beta(s,t)$ and its pointwise confidence intervals. As suggested, $\beta(s,t)$ is not statistically significant, resulting in the FLMfRw degenerating to the no-effect heterogeneity model. In summary, the results of our tests on the models suggest that heterogeneity in rainfall exists among different climatic zones, and can be expressed using simple and explicit heterogeneity models or using an ANOVA, as in Patilea, Sánchez-Sellero and Saumard (2016).

6.2. Analysis of Chinese air pollution data

The data consist of the daily air quality index (AQI) and PM2.5 in Beijing, Chengdu, and Guangzhou from 2014 to 2019. Higher AQI values indicate worse air quality. The data are collected from the air quality monitoring website. Our purpose is to explore the relationship between AQI(Y) and PM2.5 (X), which are observed daily with a data size of 16. We consider three models: the no-effect model, FLMcR, and FLMcRw. Table 4 lists the p-values based on the proposed tests, $T^f_{n,\mathtt{F}}$ and $T^f_{n,\mathtt{M}}$, and the conditional mean tests, fdapss and FMDD. The

Table 4. China air pollution data: the p-values for testing the goodness of fit of various models, with the results based on the estimated coefficient value.

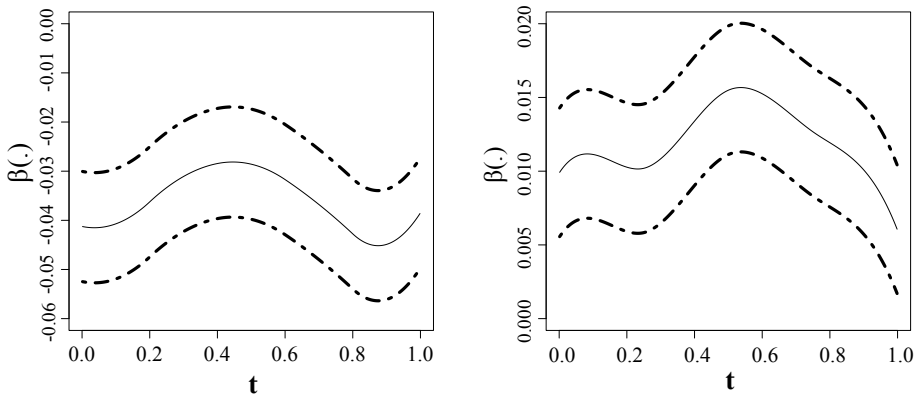
Name of the model	Formula	p-value			
		$T^f_{n,\mathtt{M}}$	$T^f_{n,\mathtt{F}}$	fdapss	FMDD
No-effect	$Y_i(t) = \epsilon_i(t)$	0.898	0.010	0.489	0.000
No-effect+heterogeneity	$Y_i(t) = f(X_i(t))\epsilon_i(t)$	0.341	1.000	0.814	0.631
FLMcR	$Y_i(t) = X_i(t)\beta(t) + \epsilon_i(t)$	1.000	0.076	0.142	1.000
FLMcRw	$Y_i(t) = X_i(t)\beta(t) + f(X_i(t))\epsilon_i(t)$	1.000	0.606	0.408	0.535

results are based on 500 bootstrap replicates. Note that the performance of the fdapss test depends highly on the selection of the bandwidth, which is rather sensitive in this example.

As shown in Table 4, the mean-based tests, $T^f_{n,\mathtt{M}}$, fdapss, and FMDD all fail to detect the heterogeneous variance expressed by the models of No-effect+ heterogeneity and FLMcRw. The null conditional mean zero assumption is not rejected, with p-values of 0.898 and 0.489 by $T^f_{n,\mathtt{M}}$ and fdapss, respectively, under the no-effect model, and with p-values of 1.00, 0.142, and 1.000 by $T^f_{n,\mathtt{M}}$, fdapss, and FMDD, respectively, under the model FLMcR. However, $T^f_{n,\mathtt{F}}$ rejects the null distributional independence assumption, with p-values of 0.010 and 0.076 for the models of no-effect and FLMcR, respectively. This leads to a no-effect + heterogeneity model and an FLMcRw with the heterogeneous variance taking the form of $f(X_i(t)) = X_i^2(t), t \in [0, 1]$. These results indicate that the heterogeneous variance of AQI can be explained by PM2.5. Furthermore, Figure 2(b) shows the estimate of $\beta(t)$ and its pointwise confidence intervals for FLMcRw. As suggested in Figure 2(b), we find that the covariate PM2.5 positively affects the daily air quality index, that is, larger values of PM2.5 tend to cause large values of the AQI, resulting in worse air quality.

7. Conclusion

We have developed a projection-based procedure for assessing the goodness of fit of generalized functional regression models. The procedure offers several features. First, it offers generality, because the proposed test can check the goodness of fit for a large number of FLMs, such as the FLMcR, FLMsR, FLMfR, generalized FLM, and functional index models. Secondly, it offers uniformity, because we provide a unified test framework for functional regression models. Remarkably, the proposed framework accommodates not only the distribution-based test, but also the mean-based and higher-order moment-based tests. Based on our theoretical and numerical results, as long as the null mean hypothesis (1.1) does not hold, the mean-based test is more powerful than the distribution-based test, which is attributed to the unified framework, under some mild conditions. By following this strategy, we obtain greater insight into the model



(a) (b)

Figure 2. (a): The estimated $\beta(\cdot)$ and associated 95% confidence bands in FLMcR+ANOVA for the Canadian weather data. (b): The estimated $\beta(\cdot)$ and associated 95% confidence bands in FLMcRw for the Chinese air pollution data.

structure and avoid calculating redundant test statistics, thus alleviating the computational burden. Third, it offers flexibility, because the proposed test is free of any distribution assumptions, and is constructed based on independence in distribution, which accounts for the mean-based independence considered in the literature and any order moment-based independence. Fourth, it provides the parameter rate of the local alternative. The proposed test has outstanding power performance under the alternatives, that is, $O(n^{-1/2})$, in contrast to the nonparametric order obtained in the literature. Fifth, it offers computational convenience. The proposed test is free of user-chosen parameters, which enhances computational expedience and avoids subjective selection.

There are several possible extensions of our method. First, we focus on generalized functional models with a known link function. Extending this to the generalized FLM with an unknown link function requires extra effort, and deserves further exploration. Second, our method requires that the covariates $\boldsymbol{X}$ are continuous functions. Because there is no KL expansion for discrete covariates, especially for binary covariates, accommodating discrete covariates is worthy of further investigation to address specific scientific questions. Third, the asymptotic distribution of the proposed statistics does not have an easily handled form. Thus, we use the bootstrap procedure, which generates extra computational costs. Therefore, finding an alternative method or developing more efficient algorithms is left for future work.

Supplementary Material

Supplementary Material contains additional notation, simulation results, and technique details, including proofs of the theorems.

Acknowledgments

The research was partially supported by National Key R&D Program of China (No.2022YFA1003702), National Natural Science Foundation of China (Nos.12271441 and 11931014), and New Cornerstone Science Foundation for Lin and Zhou. The authors thank the co-editor, associate editor, and two reviewers for their constructive comments and helpful suggestions. The first two authors contributed equally to this work.

References

Cai, T. T. and Hall, P. (2006). Prediction in functional linear regression. *Ann. Statist.* **34**, 2159–2179.

Cardot, H., Ferraty, F. and Sarda, P. (2003). Spline estimators for the functional linear model. *Statist. Sinica* **13**, 571–591.

Carroll, R. J., Fan, J., Gijbels, I. and Wand, M. P. (1997). Generalized partially linear single-index models. *J. Amer. Statist. Assoc.* **92**, 477–489.

Chen, D., Hall, P. and Muller, H.-G. (2011). Single and multiple index functional regression models with nonparametric link. *Ann. Statist.* **39**, 1720–1747.

Chiou, J.-M. and Müller, H.-G. (2007). Diagnostics for functional regression via residual processes. *Comput. Statist. Data Anal.* **51**, 4849–4863.

Cuesta-Albertos, J. A., García-Portugués, E., Febrero-Bande, M. and González-Manteiga, W. (2019). Goodness-of-fit tests for the functional linear model based on randomly projected empirical processes. *Ann. Statist.* **47**, 439–467.

Delsol, L., Ferraty, F. and Vieu, P. (2011). Structural test in regression on functional variables. *J. Multivariate Anal.* **102**, 422–447.

Dhar, S. S., Bergsma, W. and Dassios, A. (2018). Testing independence of covariates and errors in non-parametric regression. *Scand. J. Stat.* **45**, 421–443.

Ding, H., Liu, Y., Xu, W. and Zhang, R. (2017). A class of functional partially linear single-index models. *J. Multivariate Anal.* **161**, 68–82.

Dou, W. W., Pollard, D. and Zhou, H. H. (2012). Estimation in functional regression for general exponential families. *Ann. Statist.* **40**, 2421–2451.

Escanciano, J. C. (2006). A consistent diagnostic test for regression models using projections. *Econometric Theory* **22**, 1030–1051.

Fan, J. and Zhang, J.-T. (2000). Two-step estimation of functional linear models with applications to longitudinal data. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **62**, 303–322.

Feng, S. and Xue, L. (2016). Partially functional linear varying coefficient model. *Statistics* **50**, 717–732.

Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis: Theory and Practice*. Springer.

Folland, G. B. (2002). *Advanced Calculus*. Pearson Educacion.

García-Portugués, E., González-Manteiga, W. and Febrero-Bande, M. (2014). A goodness-of-fit test for the functional linear model with scalar response. *J. Comput. Graph. Statist.* **23**, 761–778.

Hall, P. and Horowitz, J. L. (2007). Methodology and convergence rates for functional linear regression. *Ann. Statist.* **35**, 70–91.

Hastie, T. J. (2017). Generalized additive models. In *Statistical Models in S*, 249–307. Routledge.

Horváth, L. and Kokoszka, P. (2012). *Inference for Functional Data with Applications*. Springer Science & Business Media.

Hsing, T. and Eubank, R. (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. John Wiley & Sons.

Kokoszka, P., Maslova, I., Sojka, J. and Zhu, L. (2008). Testing for lack of dependence in the functional linear model. *Canad. J. Statist.* **36**, 207–222.

Kokoszka, P. and Reimherr, M. (2017). *Introduction to Functional Data Analysis*. Chapman and Hall/CRC.

Lee, C., Zhang, X. and Shao, X. (2020). Testing conditional mean independence for functional data. *Biometrika* **107**, 331–346.

Lei, J. (2014). Adaptive global testing for functional linear models. *J. Amer. Statist. Assoc.* **109**, 624–634.

Li, J., Huang, C. and Zhu, H. (2017). A functional varying-coefficient single-index model for functional response data. *J. Amer. Statist. Assoc.* **112**, 1169–1181.

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

Li, Y., Wang, N. and Carroll, R. J. (2010). Generalized functional linear models with semiparametric single-index interactions. *J. Amer. Statist. Assoc.* **105**, 621–633.

Ma, S. (2016). Estimation and inference in functional single-index models. *Ann. Inst. Statist. Math.* **68**, 181–208.

McLean, M. W., Hooker, G. and Ruppert, D. (2015). Restricted likelihood ratio tests for linearity in scalar-on-function regression. *Stat. Comput.* **25**, 997–1008.

McLean, M. W., Hooker, G., Staicu, A.-M., Scheipl, F. and Ruppert, D. (2014). Functional generalized additive models. *J. Comput. Graph. Statist.* **23**, 249–269.

Muller, H.-G. and Stadtmuller, U. (2005). Generalized functional linear models. *Ann. Statist.* **33**, 774–805.

Neumeyer, N. (2009). Testing independence in nonparametric regression. *J. Multivariate Anal.* **100**, 1551–1566.

Patilea, V., Sánchez-Sellero, C. and Saumard, M. (2016). Testing the predictor effect on a functional response. *J. Amer. Statist. Assoc.* **111**, 1684–1695.

Ramsay, J. O. (1982). When the data are functions. *Psychometrika* **47**, 379–396.

Ramsay, J. O. and Silverman, B. W. (2002). *Applied Functional Data Analysis*. Springer Series in Statistics. Springer-Verlag.

Tang, Q., Kong, L., Ruppert, D. and Karunamuni, R. J. (2021). Partial functional partially linear single-index models. *Statist. Sinica* **31**, 107–133.

Wang, J.-L., Chiou, J.-M. and Müller, H.-G. (2016). Functional data analysis. *Annu. Rev. Stat. Appl.* **3**, 257–295.

Yao, F., Müller, H.-G. and Wang, J.-L. (2005). Functional data analysis for sparse longitudinal data. *J. Amer. Statist. Assoc.* **100**, 577–590.

Guizhen Li

School of Statistics, Southwestern University of Finance and Economics, Chengdu 611130, China.

E-mail: ligz@smail.swufe.edu.cn

Mengying You

School of Statistics and Information, Shanghai University of International Business and Economics, Shanghai 201620, China.

E-mail: mengyy@suibe.edu.cn

Ling Zhou

School of Statistics, Southwestern University of Finance and Economics, Chengdu 611130, China.

E-mail: zhouling@swufe.edu.cn

Hua Liang

Department of Statistics, The George Washington University, Washington, DC 20052, USA.

E-mail: hliang@email.gwu.edu

Huazhen Lin

School of Statistics, Southwestern University of Finance and Economics, Chengdu 611130, China.

E-mail: linhz@swufe.edu.cn

(Received March 2022; accepted January 2023)